Supplemental Material

Comparison between Geostatistical Interpolation and Remote Sensing Techniques for the Estimation of Long-Term Exposure to Ambient $PM_{2.5}$ Concentrations across the Continental U.S.

Seung-Jae Lee, Marc L. Serre, Aaron van Donkelaar, Randall V. Martin, Richard T. Burnett, Michael Jerrett

Geostatistical approaches may also suffer from the misuse of smoothing filters especially when a study domain spreads throughout broad spatial scales (i.e., U.S. continent). For example, there may exist a problem fitting the different periodicity in time on the west and east portion of the U.S. because $PM_{2.5}$ levels vary inversely by season with the highest levels being observed in the east during the summer months and the highest in the west during winter months (Bell et al. 2007 and also Figure 2a).

Figure 2(a) is a map of the U.S. with colored circles that represents months of the year when $PM_{2.5}$ levels reach their maximum, and confirms that there is an inverse seasonal variation in $PM_{2.5}$ concentrations from the east to west of the U.S. (i.e., highest levels in the west during winter and highest in the east during summer). We select two monitoring stations (their locations are marked as black unfilled circles in Figure 2a). Figure 2(b) is showed as an example of western portion of the U.S. whereas Figure 2(c) as an eastern part. Figures 2(b) and 2(c) depict mean trend values from the CSTM (red solid lines) and SSTM (black dotted lines) together with corresponding measurements (blue sold lines). The periodicities found from Figure 2(b) and Figure 2(c) look similar but are shifted in time because they are at different locations. The CSTM solves the problems fitting the different periodicity in time on the west and east portions of the U.S. because it successfully fits the temporal variation of $PM_{2.5}$ data at stations across the U.S.

To show Kriging with the CSTM (KC) is more accurate than Kriging with SSTM (KS) we implemented cross-validation procedure: With the measurements at the space/time points $\boldsymbol{p}_{\mathrm{cv}}=[\boldsymbol{s}_{\mathrm{cv}},t_{\mathrm{cv}}]$ where $\boldsymbol{s}_{\mathrm{cv}}$ =randomly selected 25 sites across the U.S. (out of the locations in Figure 2a) and t_{cv}=each month in between 2001 and 2006, 1) we remove one measurement at a time, 2) re-estimate it using only nearby measurements, 3) iterate this kind of estimation procedure for all of the measurements at the 1800 $\boldsymbol{p}_{\mathrm{cv}}$ (25 sites×72 months), and 4) compute estimation errors (difference between estimates and measurements left out of the cross-validation procedure). In the end we calculate Mean Square Error (MSE) (average of the squares of the estimation errors) as an indication of mapping accuracy for KS and KC. We test whether KC is more accurate than KS (equivalently the CSTM works better than the SSTM), as demonstrated below.

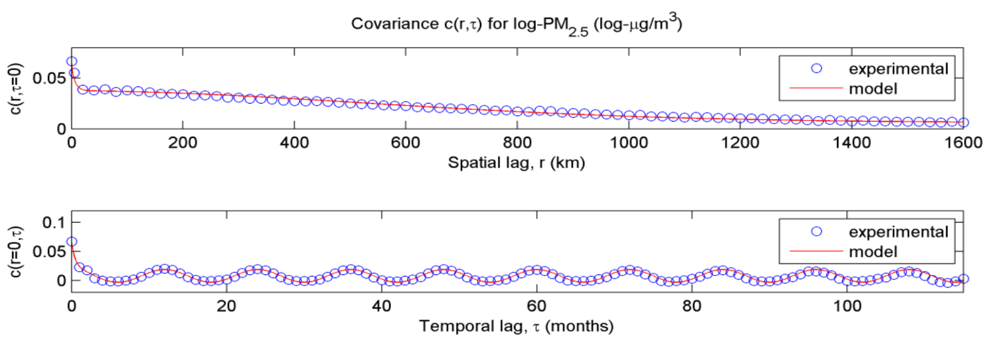
KC uses the covariance information (Supplemental Material, Equation and Supplemental Material, Figure) to obtain the kriging weight in Eq. [2] and calculates the mean trend values $\boldsymbol{m}_X(\boldsymbol{p}_{\mathrm{d}})$ at data points and $\boldsymbol{m}_X(\boldsymbol{p}_{\mathrm{cv}})$ at the cross-validation points (1800 spatiotemporal points). The MSE of KC is only 0.0561 (log-μg/m^3)^2 whereas that of KS (its covariance was not shown here) is 0.0635 (log-μg/m^3)^2. The MSE reduction from the latter to the former -11.65% indicating the former is more accurate than the latter by 11.65%.

Overall KC is proven more accurate and adequate for the entire U.S. than KS. KC hardly, however, outperforms KS at certain cross-validation points over space and time (thus only the 11.65% improvement overall) because KC relies on measurements at hand and therefore limits its ability to have more dependable mean trend values at the points around which there are no available measurements spatiotemporally. KC is still more accurate than KS and it is the interpolation method to contrast with the $PM_{2.5}$ estimates using remote sensing (i.e., referred to as RS in the main text).

With CSTM-induced residuals ($PM_{2.5}$ measurements − CSTM) we may estimate space/time variability (experimental covariance, also see the circles in Supplemental Material, Figure) for a given spatial lag r and temporal lag τ. The covariance may be parameterized by sill ($v_{01} - v_{04}$) and range ($a_{r2} - a_{r4}$, $a_{t2} - a_{t4}$) in a covariance model (red curve in Supplemental Material, Figure) that fits the experimental covariance, i.e.:

$$c_{X_R}(r,\tau) = v_{01}\delta(r)\delta(\tau) + v_{02}\exp\left(\frac{-3r}{a_{r2}}\right)\exp\left(\frac{-3\tau}{a_{t2}}\right) + v_{03}\exp\left(\frac{-3r^2}{a_{r3}}\right)\exp\left(\frac{-3\tau}{a_{t3}}\right) + v_{04}\exp\left(\frac{-3r}{a_{r4}}\right)\cos\left(\frac{2\pi\tau}{a_{t4}}\right)$$ [Supplemental Material, Equation]

where v_{01}=0.62 (log-μg/m^3)^2, v_{02}=0.023 (log-μg/m^3)^2, v_{03}=0.0267 (log-μg/m^3)^2, v_{04}=0.0109 (log-μg/m^3)^2, $\delta(r)$=1 if r=0, $\delta(r)$=0 if r>0, $\delta(\tau)$=1 if τ =0, $\delta(\tau)$=1.3 if τ>0, a_{r2}=20 km, a_{t2}=1 month, a_{r3}=1300 km, a_{t3}=2 months, a_{r4}=9000 km, and a_{t4}=12 months. The covariance model (Supplemental Material, Equation) is used for obtaining $\boldsymbol{c}_{k,d}$ and $\boldsymbol{c}_{k,d}$ in Eq. [2] which is an input for the KC estimator in Eq. [1]. The top and bottom plots in Supplemental Material, Figure denote purely spatial (when temporal lag τ=0) and purely temporal (when spatial lag r=0) covariances, respectively. The spatial piece (top plot) is a linear combination of the nugget effect, exponential, and gaussian functions in *BMElib*, whereas the temporal portion (bottom plot) includes a linear combination of nugget effect, vertical shift, exponential function together with cosinusoidal functions associated with the seasonal effects of the $PM_{2.5}$ attribute.



Supplemental Material, Figure: Space/time experimental covariance values and their covariance models using the CSTM-induced residuals