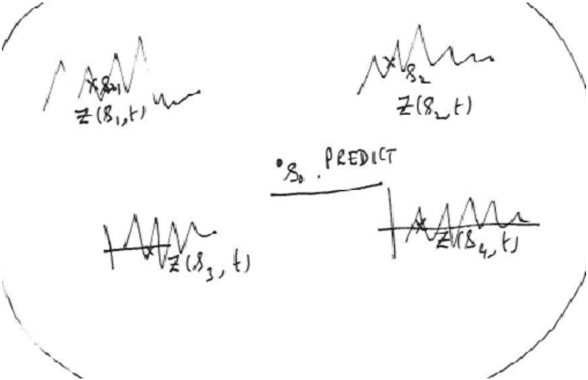
A new covariance function for spatio-temporal data analysis with application to atmospheric pollution and sensor networking

György Terdik and Subba Rao Tata

UofD, HU & UofM, UK

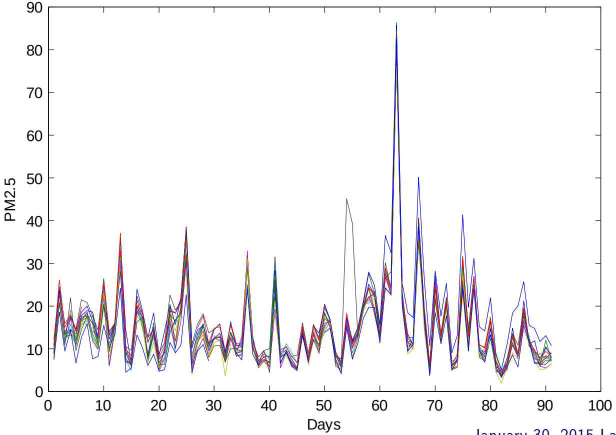
January 30, 2015
Laboratoire des Signaux et Systemes, Supelec, Paris, Gif-sur-Yvette

Spatio-temporal observations

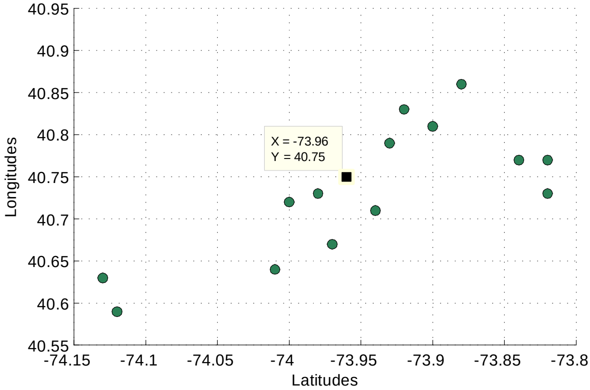


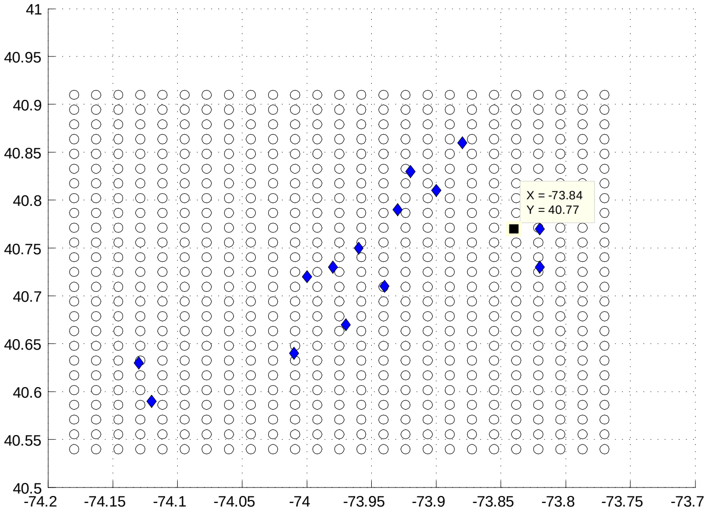
New York City air pollution data, PM2.5 measure is one of six primary air pollutants and is a mixture of fine particles and gaseous compounds such as sulphur dioxide (SO2) and nitrogen oxides (NOx)

In 2002, every 3 days and during the first 9 months, 91 equally spaced days, observed at 15 monitoring stations, $X(\mathbf{s}, t)$, see [SM05],
$\{Z(\mathbf{s}, t) = \Delta_t X(\mathbf{s}, t) : (\mathbf{s}, t) : \mathbf{s} \in \mathbb{R}^2,\ t \in \mathbb{Z}\}$



Missing data: mean by locations and fixed in time, but 3/4 of data at Location #11 is missing, namely from days 24th to 91st, location $\mathbf{s}_0$, $\{Z(\mathbf{s}_0, t);\ t = 1, 2, 3...n\}$. Sample $\{Z\left(\mathbf{s}_i, t\right);\ \mathbf{s}_i = 1, 2, ...m;\ t = 1, \ldots, n\}$. Bronx, Brooklyn, Manhattan, Queens, Staten Island




41
40.95
40.9
40.85
40.8
40.75
40.7
40.65
40.6
40.55
40.5
-74.2
-74.15
-74.1
-74.05
-74
-73.95
-73.9
-73.85
-73.8
-73.75
-73.7
X = -73.84
Y = 40.77

We assume that the random process is spatially and temporally second order stationary, homogeneous and isotropic, i.e. $E\left[Z\left(\mathbf{s},t\right)\right]=\mu$, $Var\left[Z\left(\mathbf{s},t\right)\right]=\sigma_Z^2<\infty$

$$Cov\left[Z\left(\mathbf{s},t\right),Z\left(\mathbf{s}+\mathbf{h},t+u\right)\right]=c\left(\mathbf{h},u\right)=c\left(\left\|\mathbf{h}\right\|,u\right),$$

We note that $c\left(\mathbf{h},0\right)$ and $c\left(\mathbf{0},u\right)$ correspond to the purely spatial and purely temporal covariances.
Spatio-temporal **variogram** for $\left\{Z\left(\mathbf{h},t\right)\right\}$

$$2\gamma\left(\mathbf{h},u\right)=Var\left[Z\left(\mathbf{s}+\mathbf{h},t+u\right)-Z\left(\mathbf{s},t\right)\right].$$

$$2\gamma\left(\mathbf{h},u\right)=2\left[c\left(\mathbf{0},0\right)-c\left(\mathbf{h},u\right)\right],$$

for an isotropic process, $\gamma\left(\mathbf{h},u\right)=\gamma\left(\left\|\mathbf{h}\right\|,u\right)$.

DFT

Frequency domain in time and Spatial domain in space, $\{Z(\mathbf{s}_i, t);\ t = 1, \ldots, n\}$
DFT at $\omega_k = \frac{2\pi k}{n}$, $k = 0, \pm 1, \ldots, \pm \left[\frac{n}{2}\right]$.

$$J_{\mathbf{s}_i}(\omega_k) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^{n} Z(\mathbf{s}_i, t)\, e^{-it\omega_k},$$

periodogram

$$I_{\mathbf{s}_i}(\omega_k) = \left|J_{\mathbf{s}_i}(\omega_k)\right|^2.$$

$J_{\mathbf{s}_i}(\omega_k)$ asymptotically independent and Gaussian

$$\begin{aligned} E\left(J_{\mathbf{s}_i}(\omega_k)\right) &= 0 \\ Var\left(J_{\mathbf{s}_i}(\omega_k)\right) &= E\left(I_{\mathbf{s}_i}(\omega_k)\right) \simeq g_{\mathbf{s}_i}(\omega_k). \end{aligned}$$

$$Cov\left(J_{\mathbf{s}_i}(\omega_k), J_{\mathbf{s}_i}(\omega_{k'})\right) \simeq 0, \quad k \neq k'.$$

Isotropy

$$g_{\mathbf{s}_i}(\omega_k) = g(\omega_k), \text{ for all } i$$

Fourier T and Spectral Repr

Introduce white noise

$$J_{\mathbf{s},e}(\omega) = \int e^{i\mathbf{s}\cdot\underline{\lambda}} \left[\sqrt{\frac{n}{2\pi}}\right] dZ_e(\underline{\lambda}, \omega).$$

Laplacian

$$\left[\frac{\partial^2}{\partial s_1^2} + \frac{\partial^2}{\partial s_2^2} - |c(\omega)|^2\right]^{\nu} J_s(\omega) = J_{s,e}(\omega).$$

$$\left(-\lambda_1^2 - \lambda_2^2 - |c(\omega)|^2\right)^{\nu} dZ_z(\underline{\lambda}, \omega) = dZ_e(\underline{\lambda}, \omega),$$

Spectral density

$$f_z(\underline{\lambda}, \omega) = \frac{\sigma_e^2}{(2\pi)^2 \left(\lambda_1^2 + \lambda_2^2 + |c(\omega)|^2\right)^{2\nu}}$$

Covariance

$$\begin{aligned} & Cov\left(J_{\mathbf{s}}(\omega), J_{\mathbf{s+h}}(\omega)\right) \\ = \; & \frac{\sigma_e^2}{2\pi} \left(\frac{\|\mathbf{h}\|}{2|c(\omega)|}\right)^{2\nu-1} \frac{K_{2\nu-1}\left(|c(\omega)| \, \|\mathbf{h}\|\right)}{\Gamma(2\nu)} \end{aligned}$$

$K_{2\nu-1}$: modified Bessel function of the second kind of order $2\nu - 1$.

General, $\mathbb{R}^d$, $\nu \geq 2$ integer, correlation function is given by

$$\rho\left(\|\mathbf{h}\|,\omega\right)=\frac{\left(\|\mathbf{h}\|\left|c\left(\omega\right)\right|\right)^{2\nu-\frac{d}{2}}}{2^{2\nu-\frac{d}{2}-1}\Gamma\left(2\nu-\frac{d}{2}\right)}K_{2\nu-\frac{d}{2}}\left(\|\mathbf{h}\|\left|c\left(\omega\right)\right|\right),$$

and

$$C\left(0,\omega\right)=\frac{\sigma_e^2}{\left(2\pi\right)^{\frac{d}{2}}2^{\frac{d}{2}}\left(\left|c\left(\omega\right)\right|^2\right)^{2\nu-\frac{d}{2}}}\frac{\Gamma\left(2\nu-\frac{d}{2}\right)}{\Gamma\left(2\nu\right)}=g\left(\omega\right).$$

Parameters

$$C(0,\omega)=\frac{\sigma_e^2}{2(2\nu-1)\left(|c(\omega)|^2\right)^{2\nu-1}}=g(\omega).$$

$|c(\omega)|^2$ and common spectral density $g(\omega)$, ARMA, FARMA etc..
Estimation of parameters

Correlation

$$\rho\left(\|\mathbf{h}\|,\omega\right) = \frac{C\left(\|\mathbf{h}\|,\omega\right)}{C\left(0,\omega\right)} = \frac{1}{2^{2\nu-2}\Gamma\left(2\nu-1\right)} \times \left(\|\mathbf{h}\|\left|c\left(\omega\right)\right|\right)^{2\nu-1} K_{2\nu-1}\left(\left|c\left(\omega\right)\right|\|\mathbf{h}\|\right).$$

Special case $\nu = 1$,

$$\rho\left(\|\mathbf{h}\|,\omega\right) = \|\mathbf{h}\|\left|c\left(\omega\right)\right| K_1\left(\|\mathbf{h}\|\left|c\left(\omega\right)\right|\right).$$

Spatio-temporal prediction

$$J_{\mathbf{s}_0}(\omega) = \frac{1}{\sqrt{(2\pi n)}} \sum_{t=1}^{n} Z(\mathbf{s}_0, t)\, e^{-it\omega},$$

and by inversion, we have

$$Z(\mathbf{s}_0, t) = \sqrt{\frac{n}{2\pi}} \int_{-\pi}^{\pi} J_{\mathbf{s}_0}(\omega)\, e^{it\omega} d\omega.$$

In other words given $\{J_{\mathbf{s}_0}(\omega),\ \text{for all } -\pi \leq \omega \leq \pi\}$, we can uniquely recover the sequence $\{Z(\mathbf{s}_0, t);\ t = 1, \ldots, n\}$.

Given

$$\underline{J}'_m(\omega) = \left[J_{\mathbf{s}_1}(\omega), J_{\mathbf{s}_2}(\omega), \ldots, J_{\mathbf{s}_m}(\omega)\right].$$

We note

$$\begin{aligned} E\left[\underline{J}_m(\omega)\right] &= 0 \\ E\left[\underline{J}_m(\omega)\,\underline{J}^*_m(\omega)\right] &= \underline{F}_m(\omega), \end{aligned}$$

$F_m(\omega) = \left(C\left(\|\mathbf{s}_i - \mathbf{s}_j\|, \omega\right);\ i,j = 1,2,\ldots,m\right)$, and each element $C\left(\|\mathbf{s}_i - \mathbf{s}_j\|, \omega\right)$ is given above.

$$\underline{J}'_{m+1}(\omega) = \left[J_0(\omega), \underline{J}'_m(\omega)\right],$$

which has zero mean, and variance covariance matrix

$$\begin{aligned}
&E\left[\underline{J}_{m+1}(\omega)\,\underline{J}^*_{m+1}(\omega)\right] \\
&= \begin{bmatrix} C_0(0,\omega) & E\left(J_0(\omega)\,\underline{J}^{*\prime}_m(\omega)\right) \\ E\left(\underline{J}_m(\omega)\,J^*_0(\omega)\right) & E\left(\underline{J}_m(\omega)\,\underline{J}^*_m(\omega)\right) \end{bmatrix} \\
&= \begin{bmatrix} C_0(0,\omega) & \underline{G}'_0(\omega) \\ \underline{G}_0(\omega) & \underline{F}_m(\omega) \end{bmatrix},
\end{aligned}$$

$$C_0(0,\omega) = E\left(J_0(\omega) J_0^*(\omega)\right) = C(0,\omega),$$

$$\begin{aligned} \underline{G}_0'(\omega) &= E\left[J_0(\omega) J_m^{*\prime}(\omega)\right] \\ &= \left[C\left(\left\|\mathbf{s}_0 - \mathbf{s}_1\right\|, \omega\right), \ldots, C\left(\left\|\mathbf{s}_0 - \mathbf{s}_m\right\|, \omega\right)\right] \end{aligned}$$

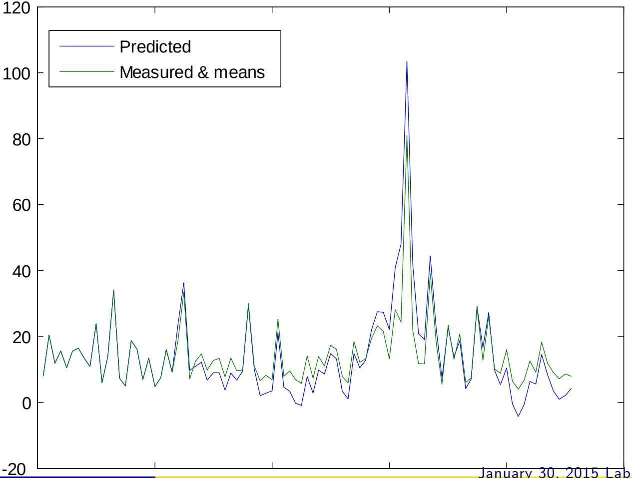
$$E\left[J_0(\omega) \,|\, \underline{J}_m(\omega)\right] = \underline{G}_0'(\omega) \underline{F}_m^{-1}(\omega) \underline{J}_m(\omega)$$

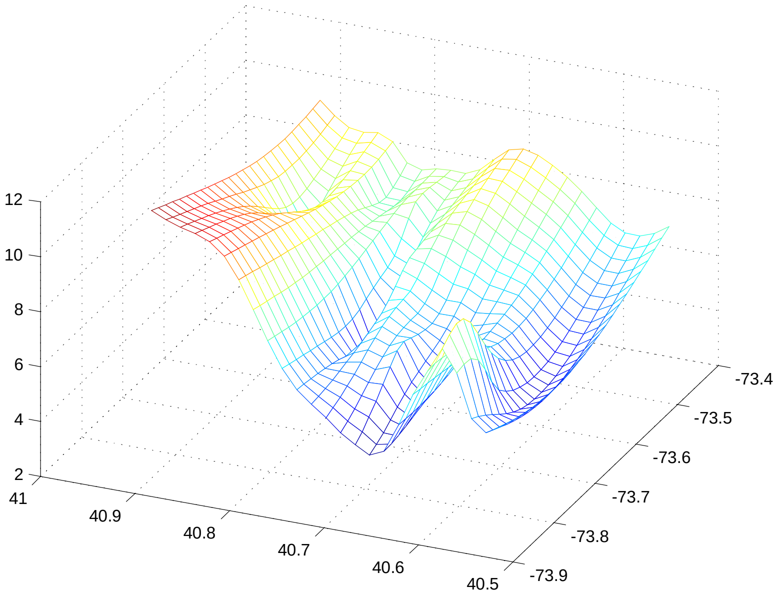
the minimum mean square error

$$\sigma_m^2(\omega) = C(0,\omega) - \underline{G}_0'(\omega) \underline{F}_m^{-1}(\omega) \underline{G}_0(\omega).$$

$$\hat{J}_0(\omega) = \underline{\hat{G}}_0'(\omega) \underline{\hat{F}}_m^{-1}(\omega) J_m(\omega).$$

Air pollution data, common spectral density $g(\omega)$, ARMA(1,1), parameters estimation by Whittle method, see [ST13] for details




12
10
8
6
4
2
41
40.9
40.8
40.7
40.6
40.5
-73.4
-73.5
-73.6
-73.7
-73.8
-73.9

S. K. Sahu and K. V. Mardia.
A bayesian kriged kalman model for short-term forecasting of air pollution levels.
Journal of the Royal Statistical Society: Series C (Applied Statistics) **54**(1), 223–244 (2005).

T. Subba Rao and Gy. Terdik.
A space-time covariance function for spatio-temporal random processes and spatio-temporal prediction (kriging).
ArXiv e-prints (November 2013).