Combining Incompatible Spatial Data

2002 paper by Carol A. Gotway and Linda J. Young

John Paige

Statistics Department
University of Washington

March 2, 2017

Change of Support Problems (COSPs): Types

Table 1. Examples of COSPs

We observe or analyze	*But the nature of the process is*	*Examples*
Point	Point	Point kriging; prediction of undersampled variables
Area	Point	Ecological inference; quadrat counts
Point	Line	Contouring
Point	Area	Use of areal centroids; spatial smoothing; block kriging
Area	Area	The MAUP; areal interpolation; incompatible/misaligned zones
Point	Surface	Trend surface analysis; environmental monitoring; exposure assessment
Area	Surface	Remote sensing; multiresolution images; image analysis

Change of Support Problems (COSPs): What can go wrong?

How are areally aggregated values related to point observations?

- Scale/aggregation effect
  - Larger area $\Rightarrow$ more averaging/smoothing $\Rightarrow$ less knowledge of finer resolution behavior
- Zoning effect
  - Overlapping areas?
  - Combined areas contiguous?
- Ecological bias
  - Aggregation bias
  - Specification bias

Examples

- Coarse resolution gridded satellite data $\Rightarrow$ fine resolution gridded data
  - Removing ecological bias may be impossible
- Have county data on air quality, want to infer individual or city mortality/cancer incidence
  - Even if we had pointwise air quality observations and individual's exact address, not clear how to best use this information to get individual cancer incidence

Modeling COSPs

Let $Z(s)$ for $s \in \mathcal{R}^d$ be our spatial process at point location s. Assume $EZ(s) = \mu(s)$ and $\text{Cov}\left(Z(u), Z(v)\right) = C(u, v)$. Let $B_1, ..., B_n$ be regions in the domain of $Z(\cdot)$, and define:

$$Z(B_i) \equiv \frac{1}{|B_i|} \int_{B_i} Z(s)\ ds$$

where $|B_i|$ is the volume of B_i.

Modeling COSPs

Assume $EZ(s) = \mathbf{x}(s)'\boldsymbol{\beta}$. Then

$$EZ(B) = \mathbf{x}(B)'\boldsymbol{\beta}$$

for

$$x_j(B) \equiv \frac{1}{|B|} \int_B x_j(s)\ ds$$

Similarly,

$$\mathrm{Cov}\left(Z(B_i), Z(B_j)\right) = \frac{1}{|B_i||B_j|} \int_{B_i} \int_{B_j} C(u, v)\ du\ dv$$

$$\mathrm{Cov}\left(Z(B), Z(s)\right) = \frac{1}{|B|} \int_B C(u, s)\ du$$

Note: we can approximate these integrals with sums

Modeling COSPs

- Without point data, estimating $C(u, v)$ using likelihood can be intractable. Using variograms is an option

Nonlinear COSPs

What happens if our observations are nonlinear transformations of normal random variables?

- Multi-Gaussian approach
- Binary data
- Isofactorial models

Nonlinear COSPs: Multi-Gaussian approach

Assume our process at a point s can be written $Z(s) = \phi(Y(s))$ for $Y(s)$ gaussian and $\phi(\cdot)$ some transformation. If we can fit a model for $Y(s)$, then we use simulations to discretize B into points $u'_1, ..., u'_N$ to estimate:

$$
\begin{aligned}
P(Z(B) < z) &\approx P\left(\frac{1}{N}\sum_{j=1}^{N} Z(u'_j)|Z(s_1), ..., Z(s_n)\right) \\
&= P\left(\frac{1}{N}\sum_{j=1}^{N} \phi(Y(u'_j))|Y(s_1), ..., Y(s_n)\right)
\end{aligned}
$$

for $s_1, ..., s_n$ observation locations

Nonlinear COSPs: Binary data

- Quantity of interest is the indicator $I(Z(s) \leq z)$
- Note that for point to block predictions, we might not want:

$$I^*(B) = \frac{1}{|B|} \int_B I(Z(d) \leq z) \; ds$$

instead we might want:

$$I(B) = I(Z(B) \leq z)$$

because then the conditional expectation in kriging prediction yields $P(Z(B) \leq z)$

Nonlinear COSPs: Binary data

- Quantity of interest is the indicator $I(Z(s) \leq z)$
- Note that for point to block predictions, we might not want:

$$I^*(B) = \frac{1}{|B|} \int_B I(Z(d) \leq z) \, ds$$

  instead we might want:

$$I(B) = I(Z(B) \leq z)$$

  because then the conditional expectation in kriging prediction yields $P(Z(B) \leq z)$
- For nonlinear block prediction using simulation, we can estimate the empirical cdf of $Z(B)$ using multiple indicator kriging

Nonlinear COSPs: Isofactorial models

$$G_{ij}(dz_i, dz_j) = \sum_{m=0}^{\infty} T_m(i,j)\chi_m(z_i)\chi_m(z_j)G(dz_i)G(dz_j)$$

- $\chi_m(z)$ are orthonormal, so:

$$\begin{aligned}
\chi_0(z) &= 1 \\
E(\chi_m(Z_i)) &= 0 \\
\mathrm{Var}\,(\chi_m(Z_i)) &= 1 \\
\mathrm{Cov}\,(\chi_m(Z_i), \chi_p(Z_j)) &= 0 \\
\mathrm{Cov}\,(\chi_m(Z_i), \chi_m(Z_j)) &= T_m(i,j)
\end{aligned}$$

Nonlinear COSPs: Isofactorial models

$$G_{ij}(dz_i, dz_j) = \sum_{m=0}^{\infty} T_m(i,j)\chi_m(z_i)\chi_m(z_j)G(dz_i)G(dz_j)$$

- Form of polynomials $\chi_m(z)$ determined by marginal distribution, $G(dz)$. If $G(dz)$ is Gaussian, then authors suggest using Hermite polynomials
- $T_m(i,j)$ determined from joint distributions $(Z(s_i), Z(s_j))$. If bivariate normal with correlation function $\rho(\|i-j\|)$, then $T_m(i,j) = [\rho(\|i-j\|)]^m$

Nonlinear COSPs: Isofactorial models

- For a Gaussian process, we have that

$$I(B) = G(z) + \sum_{m=1}^{\infty} \frac{1}{\sqrt{m}} H_{m-1}(z) g(z) H_m(Z(B))$$

for G and g Gaussian cdf and pdf respectively.

- The prediction is then obtained by replacing each $H_m(Z(B))$ with the predictor obtained via kriging. In practice this is very difficult, since we must do this while estimating the model parameters

Multiscale Tree Model

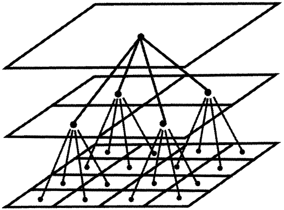

Figure 1. A Tree Structure for Multiscale Processes.

$$\mathbf{Z}(s) = \mathbf{K}(s)\mathbf{X}(s) + \varepsilon(s)$$

- $\mathbf{Z}(s)$ is an $n \times 1$ vector of measurements at location s
- $\mathbf{X}(s)$ is the true state with covariance $\mathbf{R}(s)$
- $\mathbf{K}(s)$ is an $n \times m$ selection matrix specifying which elements of state vector are measured

Multiscale Tree Model

We use both downtree and uptree models:

$$\mathbf{X}(s) = \mathbf{\Phi}(s)\mathbf{X}(p_s) + \boldsymbol{\eta}(s)$$
$$\mathbf{X}(p_s) = \underbrace{F(s)\mathbf{X}(s)}_{=E[\mathbf{X}(p_s)|\mathbf{X}(s)]} + \boldsymbol{\omega}(s)$$

for

$$F(s) = P_{p_s}\mathbf{\Phi}'(s)P_s^{-1}$$
$$\boldsymbol{\omega}(s) = \mathbf{X}(p_s) - P_{p_s}\mathbf{\Phi}'(s)P_s^{-1}\mathbf{X}(s)$$
$$W(s) \equiv E(\boldsymbol{\omega}(s)\boldsymbol{\omega}'(s)) = P_{p_s}(I - \mathbf{\Phi}'(s)P_s^{-1}\mathbf{\Phi}(s)P_{p_s})$$

- $\boldsymbol{\eta}(s)$ is independent white noise process with covariance $Q(s)$
- P_s is the covariance matrix of $\mathbf{X}(s)$, calculated with $P_s = \mathbf{\Phi}(s)P_{p_s}\mathbf{\Phi}'(s) + Q(s)$

Multiscale Tree Model

Uses 2 step fitting process:

1. leaves $\rightarrow$ root
2. root $\rightarrow$ leaves

- Computationally efficient since each node only requires local observations (model assumes that conditional on any node, its subtrees are assumed to be independent.
- The question then becomes how to model differences in variance at different scales. You could assume variance is the same no matter the scale or 'heterogeneous' variances.

Rest of paper and Conclusions

- The authors recommend multiscale modeling as a possible solution to the problem of COSP, since it more deliberately breaks down the relationships between scales
- Conditional modeling approach with tree model lends itself well to Bayesian hierarchical modeling
- Conditional on parent node, do we really think subtrees are independent?