Spatial change-of-support and misalignment problems

Peter F. Craigmile

http://www.stat.osu.edu/~pfc/

Pan-American Advanced Study Institute on Spatio-Temporal Statistics
Búzios, RJ, Brazil
Thursday, June 26 2014
Spatial statistics

• Remember, a spatial statistical analysis involves studying and modeling the dependence of data, collected in space.

  – Depending on the form of the spatial process, the spatial data are typically indexed by locations/points or regions.

  – The locations themselves may be random.

• We imagine the spatial data to be a realisation of a stochastic process, a family of random variables,

\[ \{Z(s) : s \in D\}, \]

indexed by \( s \in D \), defined on a probability space (’s’ could be a region).

• Clearly the choice of \( D \) determines the choice of spatial scale.
The choice of spatial scale – some questions

1. Which spatial scale is correct?

2. What about if there is spatial misalignment – we collected the data on one scale, but need to make inferences on a different scale.

3. How do we change from one spatial scale to another?

4. What if we have different spatial datasets that come to us on different spatial scales? How do we combine data sources?

We need to be careful not to be misled in our inferences.
Some useful references

• Gotway and Young [2002]

• Gelfand et al. [2001] and Banerjee et al. [2004, Chapter 6]
The modifiable areal unit problem (MAUP)

[The term is due to Openshaw and Taylor, 1979, but the problem is much older]

• Some spatial processes are interpretable only on areal scales.

• Famous example: crop yield, which is defined in terms of the amount of crop grown (e.g., weight) in a specific area.

  – The choice of the area is user-defined or modifiable.

  – An open question: how do we select the areal unit?
Two problems with MAUP

1. Scale effect or aggregation effect: we obtain potentially different inferences as we aggregate to larger regions.

2. Grouping effect or the zoning effect: inferences can change depend on how we choose to aggregate (for the same size of area).

These issues have some relationship to problems found in ecological inference [e.g., Cressie, 1996].
Ecological inference

• Robinson [1950] pointed out that often inference about individuals are made based on group-level data – an ecological inference.

• Using data from the 1930 US Census he related the illiteracy rate and the proportion of the population born outside the US.

  – At the state (regional) level: Corr = −0.53
  – At individual level: Corr = 0.12.

• The contradiction is due to a strong spatial effect.

  (Immigrants tended to settle in states where the native population was more literate.)
The ecological fallacy

- We obtain an ecological fallacy when analyses based on group data lead to different conclusions than we would obtain from the individual level data (also called an ecological bias).

- Two effects [Morgenstern, 1982]:
  1. Aggregation bias: the effect of the spatial aggregation itself.
  2. Specification bias: the distribution of confounding variables is different under aggregation.

- See, e.g., Wakefield and Lyons [2010].
Example: Pediatric lymphoma in Brazil

Figure 8 - Overall distribution of the most frequent pediatric lymphomas according to main geographic regions in Brazil.  

[Gualco et al., 2010]
The union of twenty-seven federal units in Brazil

(http://en.wikipedia.org/wiki/States_of_Brazil)
A solution

- The solution is to **try** to build our statistical model for the process of interest on the finest spatial scale possible.

  - We can then aggregate to produce inferences over coarser scales.

- Problems:

  - The spatial data for the process of interest may not always be available at the finest scale.

  - Important covariates may collected at different scales.
The change of support problem (COSP)

- The **spatial support** is the volume, shape, size, and orientation associated with each spatial measurement.

- A change of support is an example of data transformation.

- Thus, the change of support problem (COSP) involves studying the statistical properties of a spatial process as we change the spatial support.
**Example changes of support**

[Adapted from Gotway and Young, 2002]

<table>
<thead>
<tr>
<th>Observed at</th>
<th>Inference at</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Point</td>
<td>Point</td>
<td>Kriging</td>
</tr>
<tr>
<td>Point</td>
<td>Line</td>
<td>Contouring (upscaling)</td>
</tr>
<tr>
<td>Point</td>
<td>Area</td>
<td>Block kriging (upscaling)</td>
</tr>
<tr>
<td>Point</td>
<td>Surface</td>
<td>Environmental monitoring</td>
</tr>
<tr>
<td>Area</td>
<td>Point</td>
<td>Ecological inference (downscaling)</td>
</tr>
<tr>
<td>Area</td>
<td>Area</td>
<td>MAUP, misaligned regions (up- or downscaling)</td>
</tr>
</tbody>
</table>
An example

• Consider a Gaussian geostatistical process

\[ Z = \{ Z(s) : s \in D \} \]

defined on the domain \( D \subset \mathbb{R}^2 \) (easily generalizes to other domains).

• Suppose that the process has mean function \( E(Z(s)) = \mu(s) \) and stationary covariance function \( \text{cov}(Z(s), Z(s')) = C(s - s') \).

• Now let \( B \) denote a block defined in \( D \), and let \( Z(B) \) denote the block average of the process \( Z \) in \( B \):

\[ Z(B) = |B|^{-1} \int_B Z(s)ds, \]

where \( \int_B \) is shorthand for \( \int_{\{s \in B\}} \) and \( |B| = \int_B 1ds \) is the area of \( B \).
Properties of the block average

- The process $Z(B)$ is a random integral.

- We have that the block mean is
  $$\mu(B) = E(Z(B)) = |B|^{-1} \int_B \mu(s) ds.$$

- For two blocks $B$ and $B'$
  $$C(B, B') = \text{cov}(Z(B), Z(B')) = |B|^{-1}|B'|^{-1} \int_B \int_{B'} \text{cov}(Z(s), Z(s')) ds ds'$$
  $$= |B|^{-1}|B'|^{-1} \int_B \int_{B'} C(s - s') ds ds'.$$

- We also have
  $$C(s, B) = \text{cov}(Z(s), Z(B)) = |B|^{-1} \int_B \text{cov}(Z(s), Z(s')) ds'$$
  $$= |B|^{-1} \int_B C(s - s') ds'.$$
Exercises

1. If the mean function is constant over a block what is the block mean?

2. Suppose the mean function depends on a spatially-varying covariate \{x(s) : s \in D\} through the linear function \(\mu(s) = \beta_0 + \beta_1 x(s)\).

   What happens to the mean function over blocks?

   (This should tell you something about a grouping effect).

3. Let \(A_1, \ldots, A_m\) denote a partition of \(D\); i.e., \(\bigcup_{k=1}^m A_k = D\) and \(A_i \cap A_j = \emptyset\) for any \(i\) and \(j\). Suppose that for any points \(s_i \in A_i\) and \(s_j \in A_j\) \((i \neq j)\) that \(\text{cov}(Z(s_i), Z(s_j)) = 0\).

   Describe the statistical properties of \(\{Z(A_i) : i = 1, \ldots, m\}\).
Approximating the integrals

- In practice the integrals may not be available in a closed-form.

- Gelfand et al. [2001] shows that we can approximate the integrals using points from \( B \).

- For example letting \( s_1, \ldots, s_{LB} \) denote \( LB \) points sampled uniformly from \( B \),

\[
\mu(B) \approx L_B^{-1} \sum_{k=1}^{LB} \mu(s_k).
\]

- The sum converges in probability to the LHS as long as the spatial process \( Z \) is mean squared continuous [e.g. Stein, 1999]; i.e., \( \lim_{h \to 0} E[Z(s + h) - Z(s)]^2 = 0 \).

If \( Z \) is stationary, only need the covariance function to be continuous at \( 0 \).
Standard kriging

- Standard kriging is an example of a COSP.

  - Based on \( Z = (Z(s_1), \ldots, Z(s_n))^T \) we predict the geostatistical process \( Z \) at a new location \( s^* \).

- The best linear predictor of \( Z(s^*) \) given \( Z \) is

  \[
  E(Z(s^*)|Z) = \mu(s^*) + c_s(s^*)^T \Sigma_s^{-1}(Z - \mu_s),
  \]

  where \( \mu_s \) is the mean vector of \( Z \), \( \Sigma_s \) is the \( n \times n \) covariance matrix for \( Z \), and \( c_{s^*}(s) \) is a \( n \)-vector with \( i \)th element \( \text{cov}(Z(s^*), Z(s_i)) \).

- The kriging variance is

  \[
  \text{var}(Z(s^*)|Z) = C(0) - c_s(s^*)^T \Sigma_s^{-1} c_s(s^*).
  \]
Predicting blocks

• Now based on \( \mathbf{Z} = (Z(s_1), \ldots, Z(s_n))^T \) we predict \( Z(B) \) at block \( B \).

• The best linear predictor of \( Z(B) \) given \( \mathbf{Z} \) is

\[
E(Z(B)|\mathbf{Z}) = \mu(B) + \mathbf{c}_{s,b}(B)^T \Sigma_s^{-1}(\mathbf{Z} - \mu_Z),
\]

where \( \mu_Z \) is the mean vector of \( \mathbf{Z} \), \( \Sigma_s \) is the \( n \times n \) covariance matrix for \( \mathbf{Z} \), \( \mu(B) \) is the mean of \( Z(B) \), and \( \mathbf{c}_{s,b}(B) \) is a \( n \)-vector with \( i \)th element \( \text{cov}(Z(s_i), Z(B)) \).

• The kriging variance is

\[
\text{var}(Z(B)|\mathbf{Z}) = \text{var}[Z(B)] - \mathbf{c}_{s,b}(B)^T \Sigma_s^{-1} \mathbf{c}_{s,b}(B).
\]
Exercise: Block kriging

(This is the solution to the MAUP.)

• Now suppose we observe $Z(B_1), \ldots Z(B_n)$ and wish to predict $Z(B^*)$ at a new block $B^*$ different from $B_i$.

1. Derive the kriging mean and variance for this predictor.

2. Show that if $\mu(s) = \mu_i$ for each $s \in B_i$ then

$$
\mu(B^*) \approx \frac{\sum_{i=1}^{n} |B_i \cap B^*| Y(B_i)}{|B^*|},
$$

the areally-weighted average.

• The solution to downsampling (predicting $Z(s^*)$, say, based on block averages) follows in a similar fashion.
Practical remarks

- In practice we need to estimate the mean and covariance function of the Gaussian process $Z$, and plug the estimates into the relevant kriging equation.

  (As discussed previously we may also need to approximate the integrals.)

- Common estimation approaches include [see, e.g., Banerjee et al., 2004, Cressie and Wikle, 2011]:
  
  1. Naive approaches (e.g., weighted least squares).
  2. Maximum (reduced) likelihood.
  3. Bayesian methods.
Spatial misalignment and the use of hierarchical modeling

- Often variables come to us at different spatial scales.

- This makes the statistical inference problem harder.

- Some advice:
  
  – Which **scale** do you want to carry out inference on?
  
  – It helps to write down **hierarchical models** to relate the different variables – if possible, try to relate them on a **common** spatial scale.

  – But, remember the problem of the **ecological fallacy**. It helps to build models on the finest scale possible.
Pediatric lymphoma revisited

Suppose that we wish to predict the rate of pediatric Hodgkins lymphoma (HL) in the federal units level in Brazil, based on the following information.

<table>
<thead>
<tr>
<th>Region</th>
<th>N</th>
<th>NE</th>
<th>CW</th>
<th>S</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>HL</td>
<td>57</td>
<td>161</td>
<td>43</td>
<td>71</td>
<td>214</td>
</tr>
<tr>
<td>Total</td>
<td>114</td>
<td>373</td>
<td>105</td>
<td>209</td>
<td>500</td>
</tr>
</tbody>
</table>

What is a naive estimate of the rate of HL in each federal unit?
The map (again)
Building a hierarchical model


Consider the following simplified example from Mugglin and Carlin [1998]:

In the right hand figure, $m_1$ and $m_2$ are two different means.
Multiscale models

- There is considerable interest in building statistical models over different spatial scales.

- For early climatology examples see Berliner et al. [1999], Wikle et al. [2001], Nychka et al. [2002].


- For tree models see, e.g., Basseville et al. [1992], Chou et al. [1994], Huang and Cressie [2000] – Johannesson et al. [2007] has a space-time extension.
Spatio-temporal extensions

- Spatio-temporal extensions follow naturally.

- See Wei [2005] for formal aggregation results for time series processes.

- For methods for spatio-temporal processes see Gelfand et al. [2001].

- Read Dominici et al. [2010] for the challenges of modeling multiple pollutants collected at different spatial resolutions. Also see Choi et al. [2009].

- In context of modeling temperature, Peter Guttorp will discuss Craigmile and Guttorp [2011] this morning.
References


