What’s for today

- Change of support problem and block kriging
- Examples of Autoregressive models in Spatial context
Change of support problem can happen in many situations.

- It refers to the problem when the “support” of the data do not match.
- For example, most of the methods that we have learned so far assume that the process is defined on a point.
- However, if you want to apply the above methods to block averaged or areal averaged data, then your support from the method is not the right support to consider.

- Another situation where change of support problem arises is when you compare two different data set that have different supports.
- For example, if you consider the block averaged satellite data with observations on a point, you cannot directly compare the two.
**Block Kriging**

- Consider the process $Z(s)$ for $s \in D$ and $Z(s)$ has mean $\mu$ and covariance function $\text{Cov}\{Z(u), Z(v)\} = K(u, v)$

- Suppose we are interested in predicting the average over a region $B$ (block) of volume $|B|$:

$$Z(B) = \frac{1}{|B|} \int_B Z(s) ds$$

- The same idea of ordinary Kriging can be applied here except that we need to use the fact that:

$$\text{Cov}\{Z(B), Z(s)\} = \frac{1}{|B|} \int_B K(u, s) du$$

- To calculate the MSE, you may need the fact that:

$$\text{Cov}\{Z(B), Z(B)\} = \frac{1}{|B|^2} \int_B \int_B K(u, v) du dv$$
In practice, you may approximate the above integrals in the following way

\[
\text{Cov}\{Z(B), Z(s)\} \approx \frac{1}{N} \sum_{i=1}^{N} K(u_i, s)\]

if you discretize \(B\) into points \(u_i\) for \(i = 1, \ldots, N\)

In the same way, you may have

\[
\text{Cov}\{Z(B), Z(B)\} \approx \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} K(u_i, u_j)
\]

How would you get \(\text{Cov}(Z(A), Z(B))\) for two different blocks \(A\) and \(B\)?
If we are comparing point measured observations to block averaged data, it is like you are comparing apples to oranges.

However, using the fact that if $Z(s)$ is Gaussian, its block average, $\int_B Z(u)du$ is also Gaussian and its covariances can be calculated as given in the previous slide, you can “solve” the problem.

Example: evaluation of numerical models (weather, air pollution, climate) to observations from stations.
Comparison of CMAQ output to observations

Figure 1. (A) Weekly average of SO$_2$ concentrations (parts per billion; ppb) at the Clean Air Status and Trends Network (CASTNet) sites, for the week of July 11, 1995. (B) Output of Models-3, weekly average of SO$_2$ concentrations (ppb), for the week of July 11, 1995.
In time series, autoregressive models express the data at time $t$ as a linear combination of the values in the past.

For example, if $Z(t)$ is a time series of interest, AR($p$) model is in the form

$$Z(t) = c + \sum_{i=1}^{p} \phi_i Z(t - i) + \epsilon(t)$$

We can do similar things with spatial data.

We will see SAR models and CAR models today.
Simultaneous Autoregressive (SAR) Model

- Consider a spatial regression problem with Gaussian data.
- If $B$ is a matrix of spatial dependence parameters with $b_{ii} = 0$,

$$Z(s) = m(s)\beta + e(s)$$

$$e(s) = B e(s) + \nu$$

- The residuals $v_i$, $i = 1, \cdots, n$, have mean zero and a diagonal covariance matrix.
Simultaneous Autoregressive (SAR) Model

- We can also express this as

\[(I - B)(Z(s) - m(s)\beta) = \nu\]

- Then the covariance matrix of \(Z(s)\),

\[\Sigma_{SAR} = (I - B)^{-1}\Sigma_v(I - B^T)^{-1}\]

if \(\Sigma_v\) is the diagonal covariance matrix of \(\nu\)

- The covariance structure of \(Z(s)\) is completely determined by \(B\) and \(\Sigma_v\)
Simultaneous Autoregressive (SAR) Model

- In practice, we may need to model $B$ using a parametric model for $b_{ij}$
- We may let $B = \rho W$ where $W$ is a proximity matrix that consists of 0’s and 1’s
- Under this setting, we can easily see how SAR model reduces to a spatial model with uncorrelated errors
- In time series, if the conditional distribution of $Z(t + 1)$ given $Z(s)$, $s = 1, \cdots, t$ is the same as that of $Z(t + 1)$ given $Z(t)$, we say the process has Markov property.

- We can extend this to spatial data.

- We may say a spatial random field $Z(s)$ is a Markov random field if $Z(s_i)$ only depends on its neighbors $N_i$. 
Conditional Autoregressive (CAR) Model

- This model uses the concept of Markov property
- We consider \( f(Z(s_i)|Z(s)_{-i}) \) where \( Z(s)_{-i} \) denotes the vector of all the data except \( Z(s_i) \)
- Specifically we assume each of the conditional distributions is Gaussian and we let

\[
E(Z(s_i)|Z(s)_{-i}) = m(s_i)\beta + \sum_{j=1}^{n} c_{ij}(Z(s_j) - m(s_i)\beta)
\]

\[
\text{Var}(Z(s_i)|Z(s)_{-i}) = \sigma_i^2, \quad i = 1, \ldots, n
\]

- Here \( c_{ij} \) is nonzero only if \( s_i \in N_i \) and \( c_{ii} = 0 \)
Conditional Autoregressive (CAR) Model

- Given conditional distributions, it is not easy to construct joint distributions to do estimation and inference.
- SAR model is only defined for multivariate Gaussian distribution while CAR model may not.