Spatial Statistics (Spring 2013) – Peter Craigmile
Variograms, intrinsic processes, and commonly used parametric models for geostatistical processes

• The variogram

• Intrinsic stationarity

• Writing the variogram in terms of covariances

• Writing the covariance in terms of the variogram?

• Properties of variograms

• Visualizing and describing variograms and covariance

• Parametric models for spatial variogram/covariances
  – 1. The exponential covariance/variogram
  – 2. The Gaussian covariance/variogram
  – 3. The power exponential covariance/variogram
  – 4. The spherical exponential covariance/variogram
  – 5. The wave exponential covariance/variogram

• An inappropriate model?

• Discussion
The variogram

- Another commonly-used measure of spatial dependence is the variogram, which measures the variance of the difference of a geostatistical process \( \{Z(s) : s \in D\} \) at two spatial locations \( s \) and \( t \):

\[
2\gamma_Z(s, t) = \text{var}(Z(s) - Z(t)).
\]

- When \( E[Z(s) - Z(t)] = 0 \), we have that

\[
2\gamma_Z(s, t) = E([Z(s) - Z(t)]^2).
\]

- Terminology: \( 2\gamma_Z(s, t) \) is called the variogram; \( \gamma_Z(s, t) \) is called the semi-variogram.

- Interpretation:
  - When there is little variability in the difference, \( Z(s) - Z(t) \), then \( Z(s) \) and \( Z(t) \) are more similar (more dependent);
  - When there is greater variability in \( Z(s) - Z(t) \), then \( Z(s) \) and \( Z(t) \) are less similar (less dependent).
Intrinsic stationarity

- A geostatistical process \( \{Z(s) : s \in D\} \) is intrinsic (stationary) when \( 2\gamma_Z(s + h, s) = \text{var}(Z(s + h) - Z(s)) \) only depends on the displacement \( h \) for all \( s \).

- When the process is intrinsic stationary we can denote the variogram by \( \gamma_Z(h) \).

- As with stationary processes we can have intrinsic stationary processes that are isotropic. Such processes are called homogeneous, and we can denote the variogram by \( \gamma_Z(||h||) \) for some norm \( ||\cdot|| \).

- Weakly stationary implies intrinsic stationary (but not vice versa).
Writing the variogram in terms of covariances

- Using rules for covariances we can relate the variogram to the covariance function:

\[ 2\gamma_Z(s, t) = \text{var}(Z(s) - Z(t)) \]

\[ = \text{cov}(Z(s) - Z(t), Z(s) - Z(t)) \]

\[ = C_Z(s, s) + C_Z(t, t) - 2C_Z(s, t). \]

- When the process is weakly stationary we can simplify to

\[ \gamma_Z(h) = C_Z(0) - C_Z(h). \]

- Thus, given the covariance function (when it exists), we can calculate the variogram.
Writing the covariance in terms of the variogram?

• This is possible when a stationary geostatistical process \( \{ Z(s) : s \in D \} \) is \textbf{ergodic}: the mean of a finite area of a geostatistical process tends to the mean of the entire area as the finite area gets larger and larger.

• In that case

\[
C_Z(h) \to 0,
\]

as ||\( h \)|| \to \infty.

• Then

\[
\lim_{||h|| \to \infty} \gamma_Z(h) = \lim_{||h|| \to \infty} [C_Z(0) - C_Z(h)] \\
= C_Z(0) - 0 = C_Z(0).
\]

• With this result we get that

\[
C_Z(h) = \lim_{||u|| \to \infty} \gamma_Z(u) - \gamma_Z(h).
\]
Properties of variograms

- Properties of the variogram for an intrinsic stationary process \(\{Z(s) : s \in D\}\):

1. \(\gamma_Z(0) = 0\).

2. \(\gamma_Z(h) \geq 0\) for all \(h\).

3. \(\gamma_Z(-h) = \gamma_Z(h)\) for each \(h\) (\(\gamma_Z(\cdot)\) is an even function).

4. \(\gamma_Z(h)\) is conditional negative definite:

   - For any \(n \geq 1\) consider any spatial locations \(s_1, \ldots, s_n\) and constants \(a_1, \ldots, a_n\) such that \(\sum_{j=1}^{n} a_j = 0\).

   Then

   \[
   \sum_{j=1}^{n} \sum_{k=1}^{n} a_j \gamma_Z(s_j - s_k) a_k \leq 0.
   \]
The sill is the limiting value of the variogram at $t \to \infty$.

The range is the distance at which the variogram reaches the sill (could be infinity).

The nugget is the limiting value of the variogram as $t \to 0$, from the right.

The partial sill $= \text{sill} - \text{nugget}$.
Visualizing and describing covariances

- The **sill** is the covariance at zero distance.

- The **range** is the distance at which the covariance reaches zero (could be infinity).

- The **partial sill** is the limit of the covariance as $t \to 0$, from the right.

- The **nugget** = sill − partial sill.
There are many parametric models for the variogram and covariance function that are used for geostatistical modelling.

Suppose that the geostatistical process \( \{Z(s) : s \in D\} \) is stationary and isotropic.

In these models:

- We rewrite the distance \( ||h|| \) as just \( t \).
- The parameter \( \tau^2 > 0 \) is the **nugget**.
- The parameter \( \sigma^2 > 0 \) is the **partial sill**.
- The parameter \( \phi > 0 \) is a **range parameter** (not the range, but measures how quickly the covariance decays to zero). Fixing the distances:
  
  * With a smaller value of \( \phi \) the covariance function decays to zero quicker.
  * With a larger value of \( \phi \) the covariance function decays to zero slower.

In the upcoming plots we set \( \tau^2 = 1/2 \) and \( \sigma^2 = 1 \). In all examples except the wave \( \phi = 1 \); for the wave \( \phi = 1/4 \).
1. The exponential covariance/variogram

- Certainly the most commonly used parametric model.

- The **exponential covariance** function is
  \[
  C_Z(t) = \begin{cases} 
  \sigma^2 \exp(-t/\phi), & t > 0; \\
  \tau^2 + \sigma^2, & t = 0,
  \end{cases}
  \]
  and the **exponential variogram** is
  \[
  \gamma_Z(t) = \begin{cases} 
  \tau^2 + \sigma^2 (1 - \exp(-t/\phi)), & t > 0; \\
  0, & t = 0.
  \end{cases}
  \]
2. The Gaussian covariance/variogram

- The **Gaussian covariance** function is

\[ C_Z(t) = \begin{cases} 
\sigma^2 \exp(-t/\phi^2), & t > 0; \\
\tau^2 + \sigma^2, & t = 0,
\end{cases} \]

and the associated variogram is

\[ \gamma_Z(t) = \begin{cases} 
\tau^2 + \sigma^2(1 - \exp(-t/\phi^2)), & t > 0; \\
0, & t = 0.
\end{cases} \]
3. The power exponential covariance/variogram

- This model is valid for $0 < r \leq 2$.

- The **power exponential covariance function** is

$$C_Z(t) = \begin{cases} 
\sigma^2 \exp(-|t/\phi|^r), & t > 0; \\
\tau^2 + \sigma^2, & t = 0,
\end{cases}$$

and the associated variogram is

$$\gamma_Z(t) = \begin{cases} 
\tau^2 + \sigma^2(1 - \exp(-|t/\phi|^r)), & t > 0; \\
0, & t = 0.
\end{cases}$$

- With $r = 1$ we get the exponential, and with $r = 2$ we get the Gaussian.
4. The spherical exponential covariance/variogram

- The **spherical covariance function** is valid in $\mathbb{R}^p$ for $p = 1, 2, 3$. Not valid for $p \geq 4$.

  $$C_Z(t) = \begin{cases} 0, & t > \phi, \\ \sigma^2 \left[1 - \frac{3}{2}(t/\phi) + \frac{1}{2}(t/\phi)^3\right], & 0 < t \leq \phi; \\ \tau^2 + \sigma^2, & \text{otherwise}, \end{cases}$$

  with

  $$\gamma_Z(t) = \begin{cases} \tau^2 + \sigma^2, & t > \phi, \\ \tau^2 + \sigma^2 \left[\frac{3}{2}(t/\phi) - \frac{1}{2}(t/\phi)^3\right], & 0 < t \leq \phi; \\ 0, & \text{otherwise}. \end{cases}$$

- The **range** in this model is finite and equal to $\phi$. 

5. The wave exponential covariance/variogram

- This is the only example of a process so far where the covariance can go up and down as a function of the distance $t$. We have

$$C_Z(t) = \begin{cases} \sigma^2 \left[ \frac{\sin(t/\phi)}{t/\phi} \right], & t > 0; \\ \tau^2 + \sigma^2, & t = 0, \end{cases}$$

and

$$\gamma_Z(t) = \begin{cases} \tau^2 + \sigma^2 \left[ 1 - \frac{\sin(t/\phi)}{t/\phi} \right], & t > 0; \\ 0, & t = 0. \end{cases}$$
An inappropriate model?

6. **Linear**:

\[
\gamma_Z(t) = \begin{cases} 
\tau^2 + \sigma^2 t, & t > 0; \\
0, & t = 0.
\end{cases}
\]

Since no covariance model exists, this process is intrinsic stationary, but not weakly stationary.
Discussion

• These notes introduce most of the popular covariance models used for modelling geostatistical data.

• One we have missed out at this point is the Matèrn covariance function. This covariance function includes some of the processes included above as special examples, and has a special role in prediction – we discuss this later in the notes.

• Now we consider exploratory data analysis (EDA) for geostatistical datasets.