Week 3.

Stat 222, Spatial Statistics. Lecture MWF 9am, Math-Sci 5203.
Professor: Rick Paik Schoenberg, frederic@ucla.edu, www.stat.ucla.edu/$\sim$frederic

DAY SIX. Monday, 4/16/01.

1) Misc.

a) SEL/EMS library.
b) Read up to page 75 in Cressie.
c) Remember, no class Friday 4/20/01.

2) Nuggets and Sills

Recall that the "nugget effect" $= c_0 = \lim_{h\to 0} \gamma(h)$, where $\gamma(h)$ is the semivariogram. A related quantity is the "sill" $= C(0)$, where $C(h)$ is the covariogram.

[**Question**: Suppose a process $Z(s)$ is 2nd order stationary. Does the sill equal the nuggest effect? If not, think of an example of a process where $c_0 \neq C(0)$.]

Answer: $2\gamma(h) = V[Z(s+h) - Z(s)] = 2C(0) - 2C(h)$ [see top of p5 of week 1 notes], so $2c_0 = 2C(0) - 2\lim_{h\to 0} C(h)$, so $c_0 = C(0)$ iff. $\lim_{h\to 0} C(h) = 0$. Thus any function $C(h)$ without this property will do, e.g. $C(h) = exp(-|h|)$. Note that $C(h)$ need not be continuous at $h = 0$ and indeed is very often not continuous at zero in practice. Also note that if Z is white noise, then $C(h) = 0$ for all $h \neq 0$, so in this case the nugget effect and sill are equivalent.

3) Separable covariograms.

Cressie (p68) says a process Z in $\mathbf{R}^d$ has a separable covariogram if, for any vector h, $C(h) = C_1(h_1) \times C_2(h_2) \times \ldots \times C_d(h_d) = \prod_{i=1}^{d} C_i(h_i)$.

What does this mean? I'd prefer to express this as: $C(h) = \prod_{i=1}^{d} C_i(P_i h)$, where $P_i h$ means the projection of h onto axis i.

For example, suppose Z is a process in the plane, $\mathbf{R}^2$. By P_1 we mean projection onto the x-axis, and by P_2 we mean projection onto the y-axis. Suppose $h = (x, y)$. If Z is separable, then we can write $C(h) = f(x) \times g(y)$. Then instead of estimating the 2-d function C, we can estimate the 1-d functions f and g, separately, which is often much easier.

DAY SEVEN. Wednesday, 4/18/01.

1) Misc.

a) No class Friday 4/20. On Monday, bring book or reader to class, and have read (and understand, or be ready to ask questions about) Cressie's coal-ash data analysis example, especially the figures on pages 32, 34, 37, 39, 41.

b) Question about rescaling by x^λ.

On the bottom of p65 of Cressie, he notes that if the region D can be divided into pieces D_i so that Z is intrinsically stationary on each D_i, and further if on region i, Z has variogram $2\gamma_Z^{(i)}(h)$ such that $2\gamma_Z^{(i)}(h) \times [\mu_i^{\lambda-1}]^2$ is independent of i, then one can rescale the process Z by letting $Y(s) = g[Z(s)]$, where instead of $g(x) = log(x)$, now let $g(x) = x^\lambda$. Then Y is "approximately" intrinsically stationary.

What does Cressie mean by saying "$2\gamma_Z^{(i)}(h) \times [\mu_i^{\lambda-1}]^2$ is independent of i"? This means that $2\gamma_Z^{(i)}(h) = \phi(h)/[\mu_i^{\lambda-1}]^2$, where $\phi(h)$ is some function of h, which is the same no matter what i is, i.e. no matter what portion of the space D we're in.

From our previous formula (see Rescaled Processes, page 2 of week 2 notes), the variogram of the rescaled process Y on the subspace D_i is given by

$2\gamma_Y^{(i)}(h) = [g'(\mu_i)]^2 2\gamma_z^{(i)}(h)$,

and now $g'(x) = \lambda x^{\lambda-1}$, so $[g'(\mu_i)]^2 = \lambda^2[\mu_i^{\lambda-1}]^2$.

Therefore $2\gamma_Y^{(i)}(h) = \lambda^2[\mu_i^{\lambda-1}]^2\phi(h)/[\mu_i^{\lambda-1}]^2 = \lambda^2\phi(h)$, which does not depend on i.

Since the variogram of Y does not depend on i, it would seem that Y is intrinsically stationary. Howevere, to be intrinsically stationary, Y must also have constant (and finite) mean. Although $E\{Z(s)\} = \mu_i < \infty$, where s is in D_i, $E\{Y(s)\} = E\{[Z(s)]^\lambda\}$, which we cannot guarantee is finite and which will likely depend on i. That is why we say Y is "approximately" intrinsically stationary.

2) More about separability.

In the plane, separable covariogram means $C(h) = C_1(h_1) \times C_2(h_2)$. That is, $C(3,5) = C_1(3) \times C_2(5)$. Why is this useful? If the process is isotropic, then $C(h) = C(|h|)$, which is nice because that means one can get a handle on the covariogram simply by looking at the plot of $C(h)$ versus $|h|$. *Even if the process is NOT isotropic*, one can examine the covariogram by looking at simple plots like this, provided the process has a separable covariogram: one simply looks at plots of $C_1(h_1)$ versus h_1 and $C_2(h_2)$ versus h_2. $C_1(h)$ indicates how values are correlated as you move h units in the horizontal direction, and $C_2(h)$ indicates how values are correlated as you move h units in the vertical direction.

3) Non-parametric estimation of the variogram.

a) Method of moments (classical).

Use $2\hat{\gamma}(h) =$ mean squared difference between $Z(s_i)$ and $Z(s_j)$, over all pairs of locations (s_i, s_j) such that $|s_i - s_j| = h$.

$\hat{\gamma}$ has the nice feature that $\hat{\gamma}(-h) = \hat{\gamma}(h)$.

Two problems with $\hat{\gamma}$:

(*) What if there are no (or few) pairs of observations exactly h apart?

(**) Not robust to outliers. If one value happens to be very large and another very small, the difference between them will get squared and have an inordinate influence on the estimated variance.

b) Smoothed estimate [to solve problem (*)]. Two types.

(i) Take the mean squared difference not just over pairs of locations that are exactly h apart, but over all pairs of locations that are approximately h apart. That is, let $2\gamma^+(h) =$ mean of $[Z(s_i) - Z(s_j)]^2$ over a "tolerance region" $T(h) = \{(s_i, s_j) : h - \epsilon < |s_i - s_j| < h + \epsilon\}$. May want to take a weighted mean, weighting pairs where $|s_i - s_j|$ is very close to h more than those where the difference is further from h. The value of ϵ must be selected to determine the size of the tolerance region $T(h)$. Rule of thumb suggested in Cressie is to ensure that $T(h)$ contains at least 30 pairs of locations.

(ii) Kernel smoothing.

Let $2\hat{\gamma}^K(h) = \sum_{i,j}[Z(s_i) - Z(s_j)]^2 K(|s_i - s_j| - h)$, where $K(x)$ is some kernel function, typically a function that is symmetric around $x = 0$, e.g. the normal density function.

c) Robust estimate [to solve problem (**)]. Again, two types.

(i) Instead of taking the **mean** of $[Z(s_i) - Z(s_j)]^2$, as in a), take the **median**. This estimate is denoted $2\tilde{\gamma}(h)$.

Equivalently, could write $2\tilde{\gamma}(h)$ as $[median\sqrt{|Z(s_i) - Z(s_j)|}]^4/B(h)$, where $B(h)$ is simply some correction so that the estimate is unbiased. This suggests the following alternative estimate:

(ii) $2\bar{\gamma}(h) = [mean\sqrt{|Z(s_i) - Z(s_j)|}]^4/B(h)$, where now $B(h)$ is simply $.457 + .494/N(h)$, where $N(h)$ is the number of pairs you're taking the mean over.

Somehow taking square roots of absolute values of differences results in a more stable estimator when compared to squared differences. We'll see this later in the context of variogram clouds and square root differences clouds.

4) Proof of (and intuition about) the formula for the variance of $Y(B_1) - Y(B_2)$. (Cressie p66, or bottom of p2 of week 2 notes.)

The formula was:
$V[Y(B_1) - Y(B_2)] =$

$$-\int_{B_1}\int_{B_1} \gamma(s-u)dsdu/|B_1|^2 - \int_{B_2}\int_{B_2} \gamma(s-u)dsdu/|B_2|^2 + \int_{B_1}\int_{B_2} \gamma(s-u)dsdu/[|B_1| \times |B_2|], \quad (1)$$

where $Y(B)$ is the integral of Z over the region B, and Z is any intrinsically stationary process.

The intuition: the last term on the right hand side of equation (1) is like the whole integrated variation across B_1 and B_2. This is the SUM of the variation within B_1 and within B_2 (the first two terms on the right hand side) plus the overall variation *between* B_1 and B_2, which is $V[Y(B_1) - Y(B_2)]$.

Proof.

Since the mean of Z is constant, $E[\int_{B_1} Z(s)ds/|B_1| - \int_{B_2} Z(s)ds/|B_2|] = 0$. Thus

$$\begin{aligned} V[Y(B_1) - Y(B_2)] &= V[\int_{B_1} Z(s)ds/|B_1| - \int_{B_2} Z(s)ds/|B_2|] \\ &= E[\int_{B_1} Z(s)ds/|B_1| - \int_{B_2} Z(s)ds/|B_2|]^2 \\ &= E[\int_{B_1} Z(s)ds]^2/|B_1|^2 + E[\int_{B_2} Z(s)ds]^2/|B_2|^2 \\ &- 2E[\int_{B_1}\int_{B_2} Z(s)Z(u)dsdu]/|B_1||B_2|. \quad (2) \end{aligned}$$

Similarly, writing out $\int_{B_1}\int_{B_2} \gamma(s-u)dsdu/[|B_1| \times |B_2|]$ we get that it is equal to
$E[\int_{B_1} Z(s)^2ds]/|B_1| + E[\int_{B_2} Z(u)^2ds]/|B_2| - 2E[\int_{B_1}\int_{B_2} Z(s)Z(u)dsdu]/|B_1||B_2|$.
And,

$$\int_{B_1}\int_{B_1} \gamma(s-u)dsdu/|B_1|^2 = E[\int_{B_1} Z(s)^2ds]/|B_1| - E[\int_{B_1} Z(s)ds]^2/|B_1|^2.$$

Similarly,

$$\int_{B_2}\int_{B_2} \gamma(s-u)dsdu/|B_2|^2 = E[\int_{B_2} Z(s)^2ds]/|B_2| - E[\int_{B_2} Z(s)ds]^2/|B_2|^2.$$

Combining these with (1) and (2), and cancelling a bit, gives the desired result.

(Note that some of the steps in this proof are condensed a bit; see or email me if you'd like further explanation.)