3. Spatial Stationarity

Observe that all regressions done up to this point, including the regression model in expression (2.2.1) above, are based on the classical "ordinary least squares" assumption of *independently* and *identically distributed* residuals $(\varepsilon_i : i = 1,..,n)$. We shall consider this classical model in more detail in Section ?? below. But for the present it important to note that when the underlying index set $\{1,..,n\}$ has *structure* (such as proximity relations among points in time or space) then this structure itself can induce *statistical dependencies*. In particular, unobserved influences on the dependent variable of interest [such as $\ln PCB_i$ in (2.2.1)] often vary smoothly in space (and/or time), so that values ε_i and ε_j with i "close" to j tend to be more similar than would be expected under statistical independence (as in the example of Section 3.1 below).

Moreover, such statistical dependencies often have little substantive relation to the main phenomena of interest. In terms of our basic modeling framework, $Y(s) = \mu(s) + \varepsilon(s)$, in (1.2.1) above, we are usually more interested in the global structure of the spatial process, as represented by $\mu(s)$, than in the specific relations among unobserved residuals $\{\varepsilon(s_i) : i = 1,..,n\}$ at any given set of sample locations $\{s_i : i = 1,..,n\}$. These relations are sometimes referred to as "second-order" effects in contrast to the "first-order" effects represented by $\mu(s)$. Hence it is often desirable to model such second-order effects in a manner that will allow the analysis to focus on the first-order effects, while at the same time taking these unobserved dependencies into account. This general strategy can be illustrated by the following example.

3.1. Example: Measuring Ocean Depths

Suppose that one is interested in mapping the *depth* of the sea floor over a given region. Typically this is done by taking echo soundings (sonar measurements) at regular intervals from a vessel traversing a system of paths over the ocean surface. This will yields a set of *depth readings*, $\{D_i = D(s_i) : i = 1,..,n\}$, such as the set of measurements is shown in Figure 3.1 below:

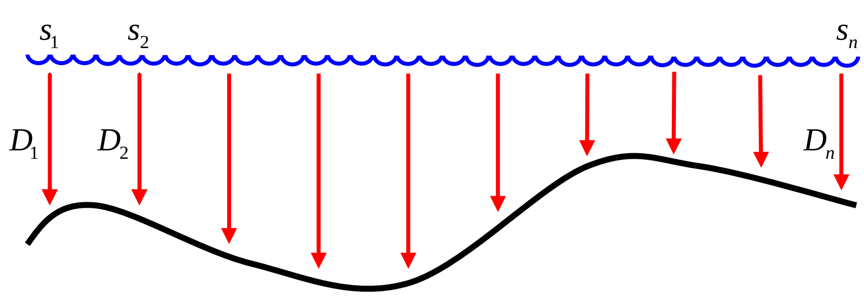

Figure 3.1. Pattern of Depth Measurements

However, the ocean is not a homogeneous medium. In particular, it is well known that such echo soundings can be influenced by the local concentration of zooplankton in the region of each sounding. These clouds of zooplankton (illustrated in Figure 3.2 below) create interference called "ocean volume reverberation".

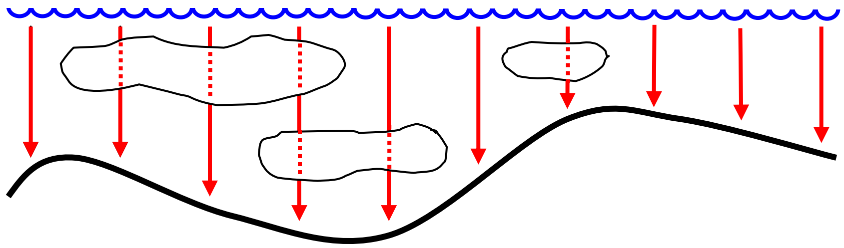

Figure 3.1. Zooplankton Interference

This tends to vary from location to location, and even from day to day (much like the way in which sunlight is affected by cloud patterns).[1] So actual readings are random variables of the form,

(3.1.1) $$D(s_i) = d(s_i) + \varepsilon(s_i) \ , \ i = 1,..,n$$

where in this case the actual depth at location s_i is represented by $d(s_i) = E[D(s_i)]$, and $\varepsilon(s_i)$ represents measurement error due to interference.[2] Moreover these errors are *statistically dependent*, since plankton concentrations at nearby locations will tend to be more similar than at locations widely separated in space. Hence to obtain confidence bounds on the true depth at location s_i, it is necessary to postulate a statistical model of these joint interference levels, $[\varepsilon(s_i) : i = 1,..,n]$. Now one could in principle develop a detailed model of zooplankton behavior, including their patterns of individual movement and clustering behavior. However, such models are not only highly complex in nature, they are very far removed from the present target of interest, which is to obtain accurate depth measurements.[3]

[1] Actual variations in the distribution of zooplankton are more diffuse than the "clouds" depicted in Figure 3.1. Veritical movement of zooplankton in the water column is governed mainly by changes in sunlight, and horizontal movement by ocean currents.

[2] In actuality, such measurement errors include many different sources, such as the reflective properties of the sea floor. Moreover, depth measurements are actually made indirectly in terms of the *transmission loss*, $L_i = L(s_i)$, between the signal sent and the echo received. The corresponding depth, D_i, is obtained from L_i by a functional relation, $D_i = \phi(L_i, \theta)$, where θ is a vector of parameters that have been calibrated under "idealized" conditions. For further details, see Urick, R.J. (1983) *Principles of Underwater Sound*, 3rd ed., McGraw-Hill: New York, and in particular the discussion around p.413.

[3] Here it important to note that such detailed models can be of great interest in other contexts. For example, acoustic signals are also used to estimate the volume of zooplankton available as a food source for sea creatures higher in the food chain. To do so, it is essential to relate acoustic signals to the detailed behavior

So what is needed here is a statistical model of spatial residuals that allows for *local spatial dependencies*, but is simple enough to be estimated explicitly. To do so, we will adopt the following basic assumptions of *spatial stationarity*:

(3.1.2) [*Homogeneity*] Residuals, $\varepsilon(s_i)$, are *identically distributed* at all locations s_i.

(3.1.3) [*Isotropy*] The joint distribution of distinct residuals, $\varepsilon(s_i)$ and $\varepsilon(s_j)$ depends only on the *distance* between locations s_i and s_j.

These assumptions are loosely related to the notion of "isotropic stationarity" for point processes discussed in Section 2.5 of Part I. But here we focus on the joint distribution of random variables at selected locations in space rather than point counts in selected regions of space. To motivate the present assumptions in the context of our example, observe first that while zooplankton concentrations at any point of time may differ between locations, it can be expected that the range of possible concentration levels over time will be quite similar at each location. More generally, the *Homogeneity* assumption asserts that the *marginal distributions* of these concentration levels are the same at each location. To appreciate the need for such an assumption, observe first that while it is in principle possible to take many depth measurements at each location and employ these samples to estimate location-specific distributions of each random variable, this is generally very costly (or not even feasible). Moreover, the same is true of most spatial data sets, such as the set of total rainfall levels or peak daily temperatures reported by regional weather stations on a given day. So in terms of the present example, one typically has a *single* set of depth measurements $[D(s_i):i=1,..,n]$, and hence only a *single joint realization* of the set of unobserved residuals $[\varepsilon(s_i):i=1,..,n]$. Thus, without further distributional assumptions, it is impossible to say *anything* statistically about these residuals. From this viewpoint, the fundamental role of the Homogeneity assumption is to allow each joint realization, $[\varepsilon(s_i):i=1,..,n]$, to be treated as multiple samples from a *common population* that can be used to estimate parameters of this population.

The *Isotropy* assumption is very similar in spirit. But here the focus is on statistical dependencies between distinct random variables, $\varepsilon(s_i)$ and $\varepsilon(s_j)$. For even if their marginal distributions are known, one cannot hope to say anything further about their *joint* distribution on the basis of a single sample. But in the present example it is reasonable to assume that if a given cloud of zooplankton (in Figure 3.1) covers location, s_i, then it is very likely to cover locations s_j which are sufficiently close to s_j. Similarly

of such microscopic creatures. See for example, Stanton T.K. and D. Chu (2000) "Review and recommendations for the modelling of acoustic scattering by fluid-like elongated zooplankton: euphausiids and copepods", *ICES Journal of Marine Science*, 57: 793–807.

for locations that are very far apart, it is reasonable to suppose that clouds covering s_i have little to do with those covering s_j. Hence the Isotropy assumption asserts more generally that similarities between concentration levels at different locations depend only on the distance between them. The practical implication of this assumption is that all pairs of residuals, $\varepsilon(s_i)$ and $\varepsilon(s_j)$, separated by the same distance, $h = \|s_i - s_j\|$, must exhibit the *same* degree of dependency. Thus a collection of such pairs can in principle provide multiple samples to estimate the degree of statistical dependency at any given distance, h. A second advantage of this Isotropy assumption is that it allows simple models of "local spatial dependency" to be formulated directly in terms of this single distance parameter. So it should be clear that these two assumptions of spatial stationarity do indeed provide a natural starting point for the desired statistical model of residuals.

But before proceeding, it should also be emphasized that while these assumptions are conceptually appealing and analytically useful – they may of course be *wrong*. For example, it can be argued in the present illustration that locations in shallow depths (Figure 3.1) will tend to experience lower concentration levels than locations in deeper waters. If so, then the Homogeneity assumption will fail to hold. Hence more complex models involved "nonhomogeneous" residuals may be required in some cases.[4] As a second example, suppose that the spatial movement of zooplankton is known to be largely governed by prevailing ocean currents, so that clouds of zooplankton tend to be more elongated in the direction of the current. If so, then spatial dependencies will depend on direction as well as distance, and the Isotropy assumption will fail to hold. Such cases may require more complex "anistropic" models of spatial dependencies.[5]

3.2. Covariance Stationarity

In many cases the assumptions above are stronger than necessary. In particular, most of our subsequent analyses will assume that residuals are *multinormally* distributed (as discussed further in Section ?? below). Since these joint distributions are determined entirely by their means and covariance structures, it suffices to assume stationarity of first and second moments. More formally, a spatial stochastic process, $\{Y(s) : s \in R\}$, is said to be *covariance stationary* if and only if the following two conditions hold for all $s_1, s_2, v_1, v_2 \in R$:

(3.2.1) $$E[Y(s_1)] = E[Y(s_2)]$$

(3.2.2) $$\|s_1 - s_2\| = \|v_1 - v_2\| \;\Rightarrow\; \mathrm{cov}[Y(s_1), Y(s_2)] = \mathrm{cov}[Y(v_1), Y(v_2)]$$

[4] For example, it might be postulated that the variance of $\varepsilon(s)$ depends on the unknown true depth, $d(s)$, at each location, s. Such nonstationary formulations are complex, and beyond the scope of these notes.

[5] This possibility will be touched on in the discussion of "anisotropic variograms" in Section ?? below.

These conditions can be stated more compactly by observing that (3.2.1) implies the existence of a *common mean value*, μ, for all random variables. Moreover, (3.2.2) implies that covariance depends only on distance, so that for each distance, h, and pair of locations $s,v \in R$ with $\|s-v\| = h$ there exists a *common covariance value*, $C(h)$, such that $\text{cov}[Y(s),Y(v)] = C(h)$. Hence, process $\{Y(s): s \in R\}$ is *covariance stationary* if and only if (iff) the following two conditions hold for all $s,v \in R$,

$$E[Y(s)] = \mu \tag{3.2.3}$$

$$\|s-v\| = h \;\Rightarrow\; \text{cov}[Y(s),Y(v)] = C(h) \tag{3.2.4}$$

Note in particular from (3.2.4) that since $\text{var}[Y(s)] = \text{cov}[Y(s),Y(s)]$ by definition, and since $\|s-s\| = 0$, it follows that these random variables must also have a *common variance*, σ^2 given by

$$\text{var}[Y(s)] = C(0) = \sigma^2 \quad , \quad s \in R \tag{3.2.5}$$

At this point it should be noted that while the most important application of covariance stationarity for our purposes will be to model residual distributions, the above definition is given in terms of an arbitrary spatial stochastic process, $\{Y(s): s \in R\}$. But (3.2.3) together with (1.2.1) imply that every covariance stationary process can be written as

$$Y(s) = \mu + \varepsilon(s) \tag{3.2.6}$$

so that each such process is associated with a unique *residual process*, $\{\varepsilon(s): s \in R\}$. Moreover, since $\text{cov}[Y(s),Y(v)] = \text{cov}[\varepsilon(s),\varepsilon(v)] = E[\varepsilon(s)\cdot\varepsilon(v)] - E(s)\cdot E(v)$, we see that $\{\varepsilon(s): s \in R\}$ must satisfy the following more specialized set of conditions for all $s,v \in R$:

$$E[\varepsilon(s)] = 0 \tag{3.2.7}$$

$$\|s-v\| = h \;\Rightarrow\; E[\varepsilon(s),\varepsilon(v)] = C(h) \tag{3.2.8}$$

These are the appropriate covariance stationarity conditions for *residuals* that correspond to the stronger Homogeneity (3.1.2) and Isotropy (3.1.3) conditions in Section 3.1 above.[6]

[6] At this point it should be noted that many standard references focus on a weaker form of stationarity in which the Isotropy assumption is relaxed by requiring that covariances dependent only on the *difference* between locations, i.e., that for all $h = (h_1, h_2)'$, $s - v = h \;\Rightarrow\; \text{cov}[Y(s),Y(v)] = C(h)$. So the covariogram

Note finally that even these assumptions are too strong in many contexts. For example (as mentioned above) it is often convenient to relax the isotropy condition implicit in (3.2.4) and (3.2.8). [See for example ([BG], p.162) and Cressie (Sections 2.2.1 and 2.3).] However, we shall treat *only* the isotropic case [(3.2.3),(3.2.4)], and shall use these assumptions throughout.

3.3 Covariograms and Correlograms

Note that since the above covariance values, $C(h)$, are unique for each distance value, h, in region R, they define a function, C, of these distances which is designated as the *covariogram* for the given covariance stationary process.[7] But as with all random variables, the values of this covariogram are only meaningful with respect to the particular units in which the variables are measured. Moreover, unlike mean values, the values of the covariogram are actually in squared units, which are difficult to interpret in any case. Hence it is often more convenient to analyze dependencies between random variables in terms of (dimensionless) correlation coefficients. For any stationary process, $\{Y(s): s \in R\}$, the (product moment) *correlation* between any $Y(s)$ and $Y(v)$ with $\|s-v\| = h$ is given by the ratio:

$$\rho[Y(s),Y(v)] = \frac{\mathrm{cov}[Y(s),Y(v)]}{\sqrt{\mathrm{var}[Y(s)}\sqrt{\mathrm{var}[Y(v)}} = \frac{C(h)}{\sqrt{C(0)}\sqrt{C(0)}} = \frac{C(h)}{C(0)} \tag{3.3.1}$$

which is simply a normalized version of the covariogram. Hence the correlations at every distance, h, for a covariance stationary process are summarized by a function called the *correlogram* for the process:

$$\boxed{\rho(h) = \frac{C(h)}{C(0)} \quad , \quad s \in R} \tag{3.3.2}$$

in this more general setting is two-dimensional, and allows for directional differences in covariances. See for example Cressie (1993, p.40) and Waller and Gotway (2004, p.273).

[7] To be more precise, if the set of all distances associated with pairs of locations in region R is denoted by $h(R) = \{h : \|s-v\| = h \text{ for some } s,v \in R\}$, then the covariogram, C, is a numerical function on $h(R)$.