Date: **11/8/10**

To: **Michael Levins and Scott Holan**

From: Noel Cressie

Re: Grain-germ (growth-attrition) model (2 pages) - A **FURTHER** UPDATE

In what follows, I shall put my new material in red.

My 8/13/10 memo has been rewritten and extended to be consistent with notation in Cressie (1993), SSD, pp. 753-802. Warning: The notation here is different from the 8/13/10 memo.

Consider $\mathbf{s} \in A$, a spatial domain of interest (e.g., North America) and $t = t_0, t_0 + 1, \ldots,$ where t_0 is the start time of the process (e.g., $t_0 = 8402$, the first of a sequence of 3-hourly intervals in 1970). Let $X_t \subset A$ denote the union of all pixels above the thresholds, $\{k(\mathbf{s}) \colon \mathbf{s} \in A\}$. That is, if $Y_t(\mathbf{s})$ is the rainfall in pixel $\mathbf{s}$ at time t, then

$$X_t \equiv \{\mathbf{s} \colon \mathbf{s} \in A \text{ and } I(Y_t(\mathbf{s}) > k(\mathbf{s})) = 1\},$$

where $k(\cdot)$ is given and $I(\cdot)$ is the indicator function. We would initially like to model

$$X_{t+1} \text{ given } X_t,$$

but later we could also model rainfall *amounts* in these exceedance regions.

Let $\{\mathbf{x}_{ti} \colon i = 1, 2, \ldots, n_t\}$ denote the individual pixels in X_t. That is, $\mathbf{x}_{ti} \in X_t$ iff

$$I(Y_t(\mathbf{x}_{ti}) > k(\mathbf{x}_{ti})) = 1; \quad i = 1, \ldots, n_t.$$

Let $\{\mathbf{v}_t(\mathbf{s}) \colon \mathbf{s} \in A\}$ denote the velocity vectors (in km/hr) given by NARCCAP at time t. Then

$$\mathbf{u}_t(\mathbf{s}) \equiv 3\mathbf{v}_t(\mathbf{s})$$

is a displacement vector. **The multiplier "3" is chosen because we are considering three-hourly time steps.** Define the sets,

$$L_t(\mathbf{s}) \equiv \{c\mathbf{u}_t(\mathbf{s}) \oplus \mathbf{s} \colon -1 \le c \le 1\}; \quad \mathbf{s} \in A,$$

and

$$M_t \equiv \cup\{L_t(\mathbf{x}_{ti}) \colon i = 1, \ldots, n_t\}.$$

This choice of M_t "fattens" the set X_t both backward and forward along the wind vector at each element of X_t. However, we now consider a physically more defensible way to "fatten" X_t.

Instead, we define M_t by using some fraction, f, of the threshold:

$$M_t \equiv \{\boldsymbol{\omega} \colon \boldsymbol{\omega} \in A \text{ and } I(Y_t(\omega) > f \cdot k(\boldsymbol{\omega}))\}; \quad 0 \le f < 1.$$

Now use "Bernoulli thinning" on M_t. After thinning, denote the resulting set as D_t; that is, each pixel is or is not in D_t according to iid Bernoulli random variables with probability p_t; $t = t_0, t_0 + 1, \ldots$. Denote

$$D_t \equiv \{\mathbf{s}_{t,j} \colon j = 1, \ldots, |D_t|\} ,$$

where $\{\mathbf{s}_{t,j}\}$ are the "germs" (foci of growth). The "grains," $\{Z_{t,j}\}$, are iid random sets (here, they are random **lines** centered at $\mathbf{0}$ with random length and random orientation). That is,

$$\{Z_{t,j} \colon j = 1, \ldots, |D_t|\} \text{ are iid as } Z_t .$$

Finally,

$$X_{t+1} \equiv \cup\{Z_{t,j} \oplus \mathbf{s}_{t,j} \colon j = 1, \ldots, |D_t|\} ,$$

which we can write as $X_{t+1} \equiv \{\mathbf{x}_{t+1,i} \colon i = 1, \ldots, n_{t+1}\}$.

Hence, attrition can occur if p_t is small and **the length, $|Z_t|_1$, of Z_t** is small. Growth can occur if **the length, $|Z_t|_1$, of Z_t** is large. Notice that M_t *expands* the set of possible pixels, and then the thinning field is used to *randomly delete* some of these pixels. This describes a type of "birth and death" process.

Notes

1. **Define $A \oplus B \equiv \{\mathbf{a} + \mathbf{b} \colon \mathbf{a} \in A,\ \mathbf{b} \in B\}$. Then if A is a set centered at $\mathbf{0}$ and B is a singleton $\mathbf{b}$, the set $A \oplus B$ is simply a translation of A so that it is centered at $\mathbf{b}$. There is a very useful equivalent way to write set addition:**

$$A \oplus B = \cup_{\mathbf{b} \in B}(A \oplus \mathbf{b}) .$$

2. Data are the sequence, $X_{t_0}, X_{t_0+1}, \ldots$.

3. Given the dynamical nature of the model, we should be able to estimate p_t, $E(|Z_t|_1)$, and $\mathrm{var}(|Z_t|_1)$ from the pair (X_t, X_{t+1}); $t = t_0, t_0 + 1, \ldots$.

4. A reference for this type of model in a cancer-cell-growth context is:
Cressie, N. and Hulting, F.L. (1992). A spatial statistical analysis of tumor growth. *Journal of the American Statistical Association*, **87**, 272-283.
More details can be found in Cressie, N. (1993). *Statistics for Spatial Data*, rev. edn., Wiley, NY, pp. 753-802.

5. The hitting function is given by (9.7.4) on p. 777 of Cressie (1993). **That formula is:**

$$T_{X_{t+1}}(K) = 1 - \exp\{-\lambda E(|\check{Z}_t \oplus K) \cap X_t|\} ,$$

where $\check{Z} \equiv \{-\mathbf{z} \colon \mathbf{z} \in Z\}$ is the reflection of Z about the origin. Choice of K in **this** formula is determined by whether a closed-form calculation is possible. **The**

presence of "$\cap X_t$" is potentially problematic, so I would try to finesse it as follows:

$$
\begin{aligned}
|(\check{Z}_t \oplus K) \cap X_t| &= |(\bigcup_{\mathbf{z}\in Z_t} K \oplus \{-\mathbf{z}\}) \cap X_t| \\
&= |\bigcup_{\mathbf{z}\in Z_t} [(K \oplus \{-\mathbf{z}\}) \cap X_t]| \\
&= |\bigcup_{\mathbf{z}\in Z_t} [K \cap (X_t \oplus \mathbf{z})]| \\
&= |K \cap [\bigcup_{\mathbf{z}\in Z_t} (X_t \oplus \mathbf{z})]| \\
&= |K \cap (X_t \oplus Z_t)|
\end{aligned}
$$

Recall that Z_t is a random line, so $(X_t \oplus Z_t)$ might be easier to work with than $(\check{Z}_t \oplus K)$. I suggest you try to use this formula to calculate the hitting function.