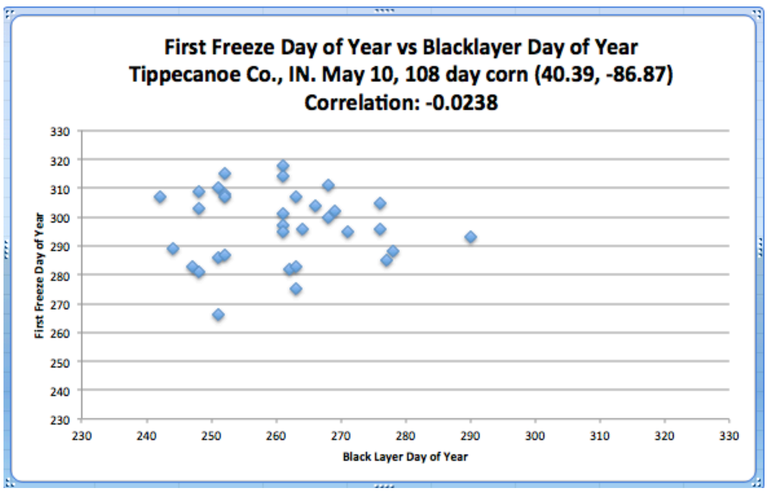
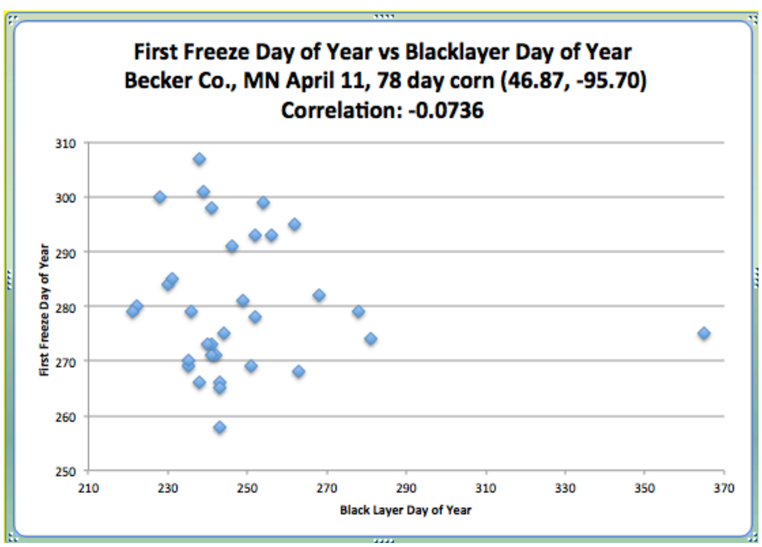
Probability of Freeze before Black Layer
(by Larry Biehl, 4/30/2015)

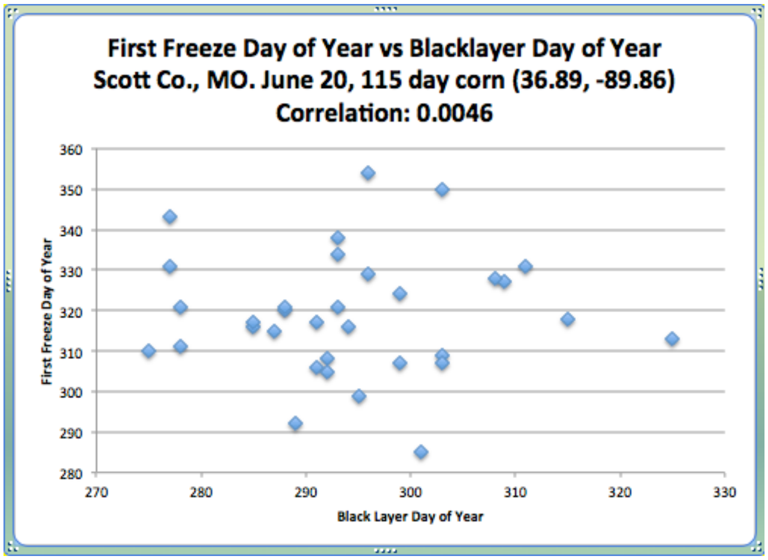
Based on a meeting with Purdue Statisticians on April 1, they recommended that one could compute this probability as the following:

P (First freezing day < Black Layer) = sum of [P (First freezing day = t) * P (Black Layer > t)] for t from first historical day of year for first freeze through the last historical day of year for first freeze. AS LONG AS there is confidence that the first freezing day and Black Layer are independent.

An analysis of the relationship of first freezing day and Black Layer day was conducted. The following figures illustrate the relationship for three scattered locations in the U2U area:







The above three figures illustrate that first freeze day and day to reach Black Layer appear to be independent.

Upon study of the information by the Statisticians, they indicated that the formula given above will be okay to us to use.

The following procedure will be an empirical approach to determine the probability of freeze before Black Layer. The Statisticians also outlined a non-parametric density estimation approach.

Programing Procedure

If the earliest observed historical first freeze day-of-year for the years since 1981 is after the latest observed day-of-year to reach Black Layer, the probability of freeze before black layer is 0%.

Else if the latest observed historical first freeze day-of-year for the years since 1981 is before the earliest observed day-of-year to reach Black Layer, the probability of freeze before black layer is 100%.

Else the probability of freeze before black layer is computed as the sum of the product of P(first freeze doy = t) times the P(Black Layer >=t) where:

P(first freeze day = t) = #{first freeze day = t} / n where n = number of first freeze observations.

P(Black Layer >=t) = #{Black Layer doy>=t} / m where m = number of black layer observations.

t will go from the earliest observed first freeze DOY to the earliest of [latest observed first freeze DOY and latest observed Black Layer DOY]. (If the first freeze DOY is after the latest observed Black Layer DOY, the probability for P(Black Layer >=t) is 0.)

The following is the document that was received from Ching-Wei Cheng (student in Purdue's Statisticians Consulting Office and reviewed by monitoring Professor) on April 29, 2015 concerning the probability of freeze before black layer.

No pattern is shown in the scatterplots, which suggests the first freezing day and Black Layer are independent. So the probability of the first freezing day before the Black Layer can be decomposed by

$$P(\text{First freezing day} < \text{Black Layer}) = \sum_t P(\text{First freezing day} = t) P(\text{Black Layer} > t).$$

Since the two events are independent, their probability distributions can be estimated separately. Suppose $x_1, \ldots, x_n$ are observed first freezing days and $y_1, \ldots, y_m$ are observed Black Layer days from the history (their sample sizes can be different).

Empirical probability: Estimate the two probability distributions by

$$\hat{P}(\text{First freezing day} = t) = \frac{\#\{\text{First freezing day} = t\}}{n}$$

and

$$\hat{P}(\text{Black Layer} > t) = \frac{\#\{\text{Black Layer} > t\}}{m}.$$

And then the above probability decomposition can be used by replacing the true probabilities P by the estimated ones $\hat{P}$.

Non-parametric density estimation: (I remember you mentioned non-parametric density estimation in the initial meeting and so I got another suggestion here.) Suppose $f(t)$ and $g(t)$ are the density estimations by some non-parametric approach (e.g., kernel estimation, smoothing spline, local regression, etc.) of the first freezing day (from $x_1, \ldots, x_n$) and the Black Layer (from $y_1, \ldots, y_m$), respectively. Then, since $f(t)$ and $g(t)$ are continuous, you should replace $\hat{P}(\text{First freezing day} = t)$ by $f(t)$ and $\hat{P}(\text{Black Layer} > t) = \int_t^\infty g(s)\,\mathrm{d}s$ and the above probability decomposition should be modified as

$$\begin{aligned}\hat{P}(\text{First freezing day} = t) &= \int_{\mathcal{T}} \hat{P}(\text{First freezing day} = t)\hat{P}(\text{Black Layer} > t)\,\mathrm{d}t \\ &= \int_{\mathcal{T}} \int_t^\infty f(t) g(s)\,\mathrm{d}s\,\mathrm{d}t,\end{aligned}$$

where $\mathcal{T} \in \mathbb{R}$ is the set of possible range of t.