PENNSYLVANIA VEGETABLE MARKETING AND RESEARCH PROGRAM
PENNSYLVANIA VEGETABLE GROWERS ASSOCIATION

Final Report for Research Grant "Factors Associated with Mold in Snap Beans" Funded in 2009

Report date: December 07, 2009

Research and Report done by:
Denis A. Shah, Ph.D.
Independent contractor – research & data analytics

Background

Several factors are listed as contributory to the overall risk for mold development in snap bean. How these factors interact, or how important they may be relative to each other, is still largely unknown. Given the uncertainty in predicting when and how bad mold may be in fields, it makes sense to adopt a strategy of minimizing the risk of catastrophic loss. Hence, most snap bean fields are sprayed at least once and sometimes twice with protectant fungicides. Growers do question whether the sprays were necessary in all cases (i.e. were the conditions even right for mold to begin with), or whether the spray timing was optimal.

This project used statistical learning methods (also called data mining) to search for predictors associated with the risk of white mold. The data set was collected by the author from 2006 to 2008, and consisted of observations made in processing snap bean fields across western NY and into north-central PA. In all, 1,546 observations were made in 419 fields. White mold was observed in 23% of the fields, and was the predominant mold, so the analysis focused on predictors of this disease.

Presentation of Results

The percent of plants with white mold was estimated in each field from a sample of at least 50 plants. **Figure 1** shows that in most fields (77%), no white mold was observed. Where white mold was seen, the levels were mainly less than 20%. As most fields are sprayed with fungicides at least once, the following assumption was made: protectant fungicides are not 100% effective, so if conditions were right for white mold, the disease would be present, though the level could be affected by the sprays. This project focused on identifying predictors from the observed data set (**Table 1**) that could be used to classify fields into eihter one of two groups: white mold present or white mold absent. The data were restricted to observations made within a few days of scheduled harvest, and so the results represent factors associated with white mold (or which contribute to disease occurrence) within a few days of harvest.

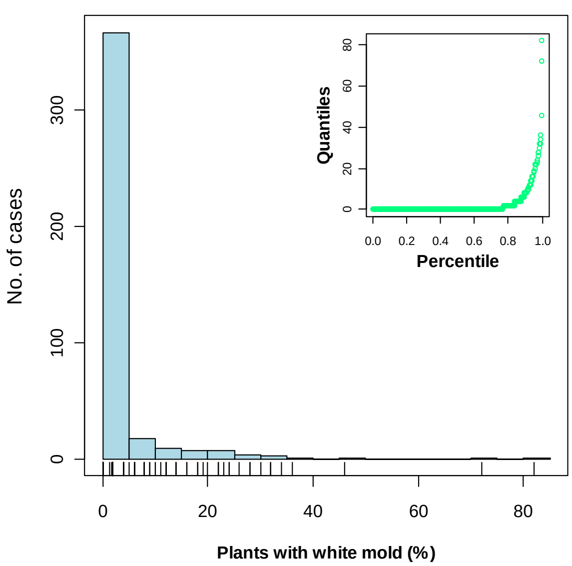
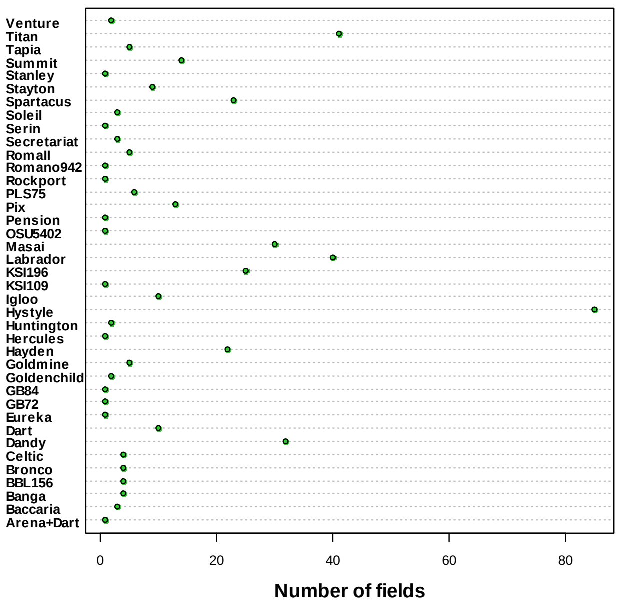


Figure 1. Histogram of plants with white mold (%)

Table 1. Some variables used to model white mold occurrence

Variable	Description	Mean and range, or categories
canclos	Open row space (cm)	23.51, 0-60.96
dap	Days after planting	58.58, 50-77
fmonth	Month in which bloom occurred	June, July, August, September
hybrid	Snap bean hybrid	39 different hybrids (see **Figure 2**)
pdateJ	Julian planting date	176.3, 134-215 [a]
rowor	Row orientation	North-South or East-West
smois	Soil moisture	0-25, 25-50, 50-75, 75-100%
vg	Hybrid group	Flat, large, whole, sieve-4

[a] Julian date 134 ≈ 14 May; Julian date 215 ≈ 03 August.



An underlying motivation in data collection was to measure variables as simply as possible, without having to rely on any advanced instrumentation. So, for example, soil moisture was assessed by using a trowel to dig into the top3 inches of the soil, and then using a soil moisture by feel method. Open row space was defined as the distance between the row canopies (tip to tip), and was measured with a yardstick.

A class of models called boosted regression trees (BRT) was used to find relationships in the data. BRT models try to 'learn' how the predictors can be used to classify fields with or without white mold. This is similar to how businesses such as Amazon.com find out what a customer is most interested in based on their previous purchases.

Figure 2. Hybrids used in commercial snap bean fields.

BRTs are a different way of looking at data compared with the methods you may have heard of before, such as regression and ANOVA. Regression and ANOVA assume a given model structure, and then try to fit the data to that model. A BRT on the other hand does not assume that the data adhere to a given model, but instead uses algorithms to learn the relationship between a reponse (here, white mold presence or absence) and predictors (**Table 1**).

The data set of 419 fields was split 70%-30% into a model training set (293 fields) and a separate testing set (126 fields) to be used for model evaluation. Over 100 different BRT models were explored and narrowed to four candidates. The four models developed from the training data set were quite similar in terms of the predictors identified and their relative importance (**Figure 3**). Of all the potential predictors of white mold, based on conditions within a few days of harvest, the two most important were planting date and canopy closure.

According to the data, it appears that in the years 2006-2008 fields planted during the first two weeks of June were at a higher risk of white mold compared with fields planted earlier or later than those weeks. Also fields planted after July 22 appeared to be at an increased risk of white mold (**Figure 4**). This is difficult to explain, but for the June plantings may be connected to the time required for sclerotia to mature and produce apothecia (Apothecia are the structures from which the spores are produced.The spores then infect bean blossoms, causing the white mold disease). For the late-season plantings (after July 22), it is possible that the cooler and wetter conditions of the late summer – early fall may promote white mold.

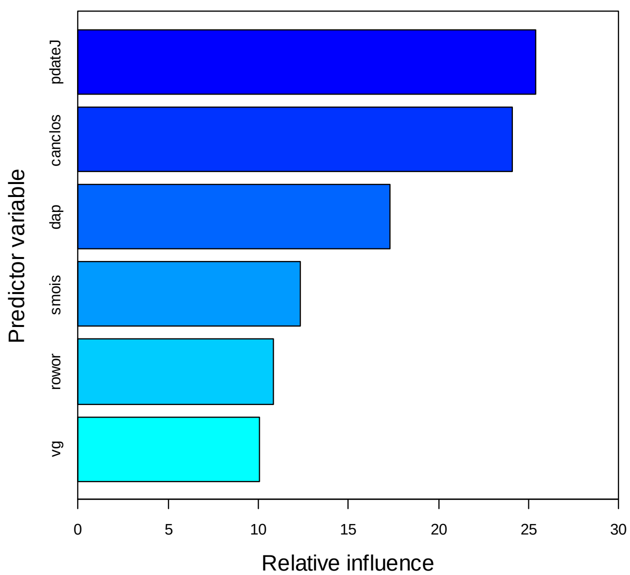


Figure 3. Relative contributions of predictor variables in boosted regression trees for white mold

The second most influential predictor was the degree of canopy closure. The more closed a canopy, the less open row space there is. Closed canopies are known to promote humidity and moisture, conditions generally favorable for mold infection.

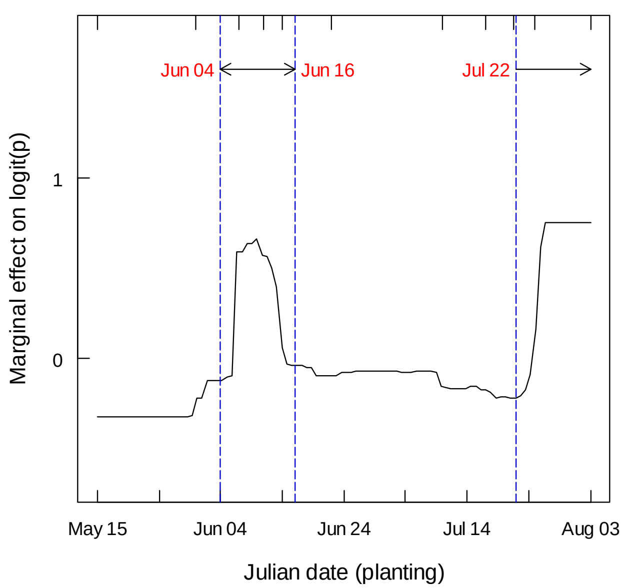


Figure 4. Modelled effect of planting date on the risk of white mold in snap bean fields.

The interesting finding was that the risk of white mold was substantially decreased when open row space was greater than about 34 cm (**Figure 5**). This may represent the point at which open canopies become largely unfavorable for spore production, dispersal and infection.

Other variables were associated with white mold to lesser extents. Higher soil moisture levels, as one may expect, were associated with elevated risk of white mold, but it appeared that the risk was about the same for any amount of soil moisture above 25%. Soil moisture itself was highly variable during the 60-day development of a given field, depending on rainfall, temperature, and the type of soil. However, I found that soil moisture tended to be higher in fields with closed canopies. This also illustrates the greater influence of canopy closure in white mold, as found by the BRTs.

Another factor of importance was row orientation. This is not a factor that one can control in general, as the row orientation largely depends on the field dimensions, which makes it more efficient to plant the rows in line with the longer dimension. Nevertheless, fields with rows oriented north-south were at higher risk for white mold than those fields with east-west rows.

The sieve-4 hybrids were at higher risk for white mold. This again may be linked to canopy, as sieve-4 hybrids tended to have more closed canopies, as shown in **Figure 6**.

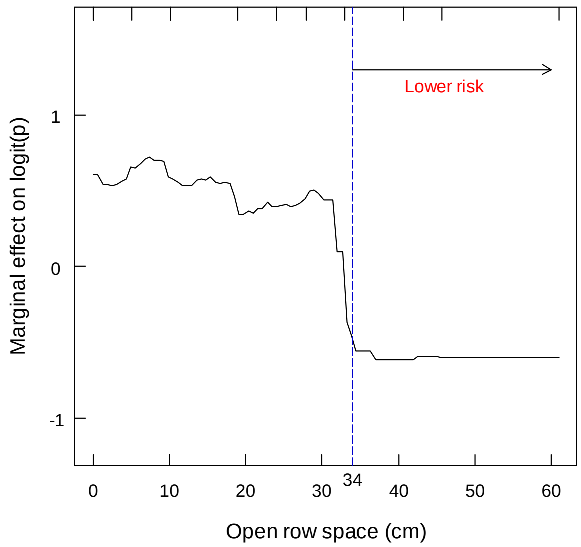


Figure 5. Modelled effect of open row space on the risk of white mold in snap bean fields.

Model evaluation. The BRT models were created with 70% of the data. These models were then tested for their ability to predict in the remaining 30% of the data whether a field would be positive for white mold. Table 2 shows the test data error matrix.

Table 2. Test data error matrix

	Predicted = 0	Predicted = 1
True status =0	94	3
True status = 1	22	7

There were 126 fields in the test data set. Of these, 29 had white mold. The BRT model correctly predicted most of the fields without white mold (**Table 2**). The overall misclassification error rate is 19.8%. However, 22 of the 29 fields with white mold were incorrectly classified.

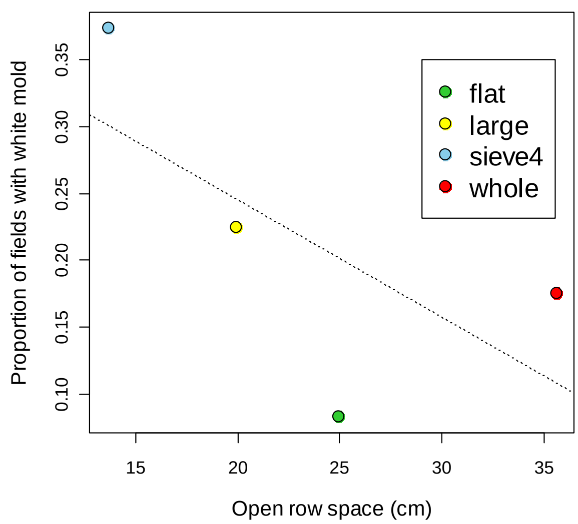


Figure 6. Presence of white mold as a function of open row space and hybrid type.

Summary

Boosted regression trees were used to 'learn' the occurrence of white mold in snap bean fields from a set of potential predictors. The two most influential factors were planting date and open row space. Other factors such as soil moisture and row orientation further impacted white mold risk, and sieve-4 type beans appeared to be at an elevated risk to white mold, probably because they also tended to have more closed canopies. In a test data set, the models could correctly classify 80.2% of fields, but less than 25% of fields with white mold were correctly classified. Further models will attempt to reduce the misclassification rate.