Hotspot Detection for Chestnut Oak Regeneration

Songlin Fei
The Pennsylvania State University
7 Ferguson Bldg.
University Park, PA 16802
Email: feisl@psu.edu

ABSTRACT
Chestnut oak regeneration hotspots were investigated in 52 mature mixed-oak stands in the central Appalachians. Four methods: ranking, SaTScan, ClusterSeer, and classification tree were applied to detect chestnut oak regeneration hotspots. Ranking method provided a quick assessment of the regeneration hotspot; SaTScan and ClusterSeer compromised the spatial correlation among the plots; classification tree not only identified the regeneration hotspots but also presented the associated biotic and abiotic variables simultaneously. The results from these four different methods were similar to each other, and no method seems statistically and ecologically superior to the others. . Inclusive knowledge of the relationships between regeneration abundance and biotic and abiotic factors can help resource management agencies, such as Forest Service and Park Service to maintain a healthy, diverse, and compositionally stable forest ecosystem.

Categories and Subject Descriptors
J. Computer Applications; J. 3 Life and Medical Sciences

General Terms
Biology and Genetics

Keywords
Forest health, Ranking, SaTScan, Classification tree, ClusterSeer, Oak regeneration,

1. INTRODUCTION
Oaks (*Quercus* spp.) are the most prevalent forest types in the eastern deciduous forests. However, natural regeneration of oaks is often difficult to obtain. Spatial and temporal hotspot detection and surveillance have been applied to critical issues, such as homeland security, public health, disaster management, and ecosystem health. We stand to gain further knowledge if we can identify these hotspots and study their associated environmental conditions. In this study, four hotspot detection methods (ranking, SaTScan [3], classification tree, ClusterSeer [5]) were applied and compared to identify hotspots for chestnut oak (*Q. montana*) regeneration.

2. DATA SOURCE
Field measurements on 52 mixed-oak stands (1,433 plots) were performed in the central Appalachians. Chestnut oak regeneration was recorded by height class. Both biotic and abiotic variables were used to explain variation in oak regeneration. Cumulative height [2] was used to describe chestnut oak seedling population abundance.

3. HOTSPOT DETECTION

3.1 Ranking
The ranking method simply assigned ranks to all the surveyed plots based on the cumulative height of each plot. To make the result comparable to the other methods, first 190 plots were defined as the regeneration hotspots (approximately 15 percent of the total number of plots). Hotspots defined by ranking method were scattered on 28 stands, with an average cumulative height of 0.7 m/m^2. Within the 28 stands, percentage of plots classified as the chestnut oak regeneration hotspot ranged from three to 91 percent, and eight stands had hotspot percentage greater than 50 percent. Chestnut oak regeneration hotspots were concentrated in the lower-left part of the Ridge and Valley province, and only about two percent of the hotspots were located on the Allegheny Plateau province.

3.2 SaTScan
SaTScan can be used to detect areas of significantly high or low rates and to evaluate reported spatial or space-time clusters [4]. By assuming that chestnut oak regeneration in an area is Poisson distributed, a Poisson-based model was used and a maximum potential regeneration population was assumed with cumulative height of eight m/m^2. To make the analysis comparable to the above ranking method, maximum spatial cluster size was set as 15 percent of the potential population.

In total, 186 plots were detected as the regeneration hotspots for chestnut oak. These regeneration hotspots were distributed on 12 stands. Seven stands had all their plots identified as hotspots, and one stand only had 22 percent of its plots identified as hotspots. Within the 186 hotspots, six statistically significant clusters were recognized. Cluster 1 had the highest average cumulative height (1.9 m/m^2) with a maximum of over 6.4 m/m^2, while Cluster 3 had the largest number of plots (109). Chestnut oak regeneration hotspots were concentrated in the lower-left part of the Ridge and Valley, and no hotspot was detected on the Allegheny Plateau physiographic province.

3.3 ClusterSeer
ClusterSeer provides spatial, temporal and spatial-temporal clustering methods to evaluate disease clusters and non-disease

events. Since cumulative height represents group-level oak regeneration abundance, Besag and Newell's [1] method was used to detect global spatial clusters. Same as the SaTScan method, a maximum potential regeneration population was assumed with cumulative height of eight m/m^2. Cluster cutoff size was set at 30, which not only provided enough power to detect the clusters, but also made the total number of plots included in these clusters comparable to the former two methods.

Across the study area, 216 statistically significant clusters were detected by ClusterSeer. Since clusters defined by this software can share the same centroid (plots), a total number of 197 plots were identified as the centroids of these clusters. These centroids were considered as the regeneration hotspots. They were scattered on 14 stands, in which seven stands had all their plots identified as hotspots, and one stand only had 10 percent of its plots identified as hotspots. Similar to the result of SaTScan, chestnut oak regeneration hotspots were concentrated in the lower-left part of the Ridge and Valley province, and no hotspot was detected on the Allegheny Plateau physiographic province.

3.4 Classification Tree

SPSS 13.0 [4] was used to run the classification tree analysis. Since the response variable, cumulative height is an ordinal variable, chi-square for determining node splitting and category merging was calculated using the likelihood-ratio method (CHAID criterion).

The classification tree had two major branches split by the two physiographic provinces. The Allegheny Plateau branch consisted of three terminal nodes, and the Ridge and Valley branch of eight terminal nodes. Deer browsing, soil series, and overstory tree composition were the factors that were significantly associated with regeneration abundance in the Ridge and Valley province. To match with the former methods, plots in the three terminal nodes that had the highest average chestnut oak cumulative height were defined as the regeneration hotspots, which included a total number of 191 plots. These hotspots were distributed on 27 stands. Six stands had all their plots identified as hotspots, and six stands had less than 10 percent of their plots identified as hotspots. Chestnut oak regeneration hotspots were again concentrated in the Ridge and Valley, and no hotspot was detected on the Allegheny Plateau physiographic province.

4. COMPARISON

Oak regeneration hotspots defined by one method were compared with the other three methods and percentages of common hotspots were listed in Table I. Overall, hotspots defined by ranking method had the least frequency matched with the hotspots defined by other methods, and hotspots defined by ClusterSeer method had the highest frequency matched with hotspots defined by other methods.

The regional distribution of the hotspots defined by the four methods had some commonalities and dissimilarities. All the hotspots defined by the four different methods were located in the Ridge and Valley province, except the ranking method defined two percent hotspots on the Allegheny Plateau. Hotspots distribution based on naïve ranking was similar to the classification tree method, which was scattered throughout the Ridge and Valley province; while the distribution based on SaTScan was similar to the ClusterSeer method, which was concentrated at the lower-left corner in the Ridge and Valley province.

Table I. Percentage of Hotspots Identified by Each Method

	Ranking	SaTScan	ClusterSeer	Tree
Ranking	100	49	62	51
SaTScan	49	100	64	58
ClusterSeer	62	64	100	66
Tree	51	58	66	100

Ranking method provides a quick and simple assessment, but it lacks the ability to detect spatial association among plots. Classification tree defines the regeneration hotspots and presents the associated biotic and abiotic variables at the same time. However, it also ignores the spatial association among plots. SaTScan and ClusterSeer implement the consideration of spatial association among plots. However, both methods use circles as the scanning window, which unavoidably includes plots with low regeneration as hotspots.

5. CONCLUSION

Comparisons among the four hotspot detection methods indicated that these methods provided relatively similar results. Each method has its advantages and limitations, and no methods can claim to be statistically and ecologically superior to the others. Therefore, it is beneficial to apply as many methods as possible in order to capture the true hotspots.

6. ACKNOWLEDGMENTS

I would like to thank Professor Kim Steiner of the Penn State School of Forest Resources, my Ph.D. in Ecology advisor, and the Pennsylvania Bureau of Forestry for support. I would also like to thank Professor G.P. Patil of the Penn State Department of Statistics, my M.A. in Statistics advisor. The NSF Digital Government Program Grant to the Center for Statistical Ecology and Environmental Statistics with Professor Patil as the PI provided substantive opportunity and timely support to carrying out the cross disciplinary research investigation that has led to this paper of potential interest to digital governance.

7. REFERENCES

[1] Besag, J. and Newell, J. The detection of clusters in rare diseases. *J. Royal Stat. Soc. Series A*, 154 (1991), 143-155.

[2] Fei, S. *Early Development of Mixed-oak Forest Stands.* Ph.D. thesis. The Pennsylvania State University. 2004.

[3] Kulldorff M. and Information Management Services, Inc. SaTScan 5.1 http://www.satscan.org/. 2004.

[4] SPSS Inc. SPSS 13.0. http://www.spss.com/. 2004.

[5] TERRASEER Inc. ClusterSeer2.0. http://www.terraseer.com/. 2004.