BUILDING IMPROVED MODELS OF SUGAR MAPLE MORTALITY

Charles H. Perry and Patrick L. Zimmerman[1]

Abstract.—The decline of sugar maple (*Acer saccharum* Marsh.) in the northern United States is causing concern, and several studies have identified soil properties that are linked to the observation of dead/dying trees. Unfortunately, the sample of trees supporting these studies is purposive in nature; soil properties are assessed only on those plots where dead trees are observed. In this study, we used the U.S. Forest Service's Forest Inventory and Analysis database (FIADB) to conduct an exploratory analysis of a broader population of sugar maple (live and dead) across a wide range of soil types. This population of plots has a highly skewed, zero-inflated distribution: the number of plots in the sample without dead trees is an order of magnitude greater than the number of plots with dead trees. One effective method of analysis is a hurdle—or conditional—model approach. In the first phase, the response variable is the presence or absence of dead sugar maple and the inferential space is the entire population of plots with sugar maple trees. The second phase uses the relative abundance of dead sugar maple as the response variable; in this case, inference is restricted to those plots where dead sugar maple trees are observed. In both sets of models, basal area and geology are significant predictors of dead sugar maple, but the most significant soil variables vary between these two inferential spaces. Our study highlights important analytical considerations when using FIADB for analysis of forest health conditions and presents simple methods to create a more comprehensive space for statistical inference.

INTRODUCTION

Several studies of sugar maple (*Acer saccharum* Marsh.) mortality exist (e.g., Horsley et al. 2000, Long et al. 2009), but most evaluations focus on an area of known decline from Pennsylvania to New Hampshire. Sampling of sugar maple decline in these and related studies tends to be purposive in nature and evaluates only those plots with dead sugar maple.

The U.S. Forest Service Forest Inventory and Analysis Program (FIA) collects field data to describe the status and trends of forests across the United States. It focuses on live trees and live-tree observations vastly outnumber those of dead trees in the FIA database (FIADB). To wit, the ratio of live-tree to dead-tree observations for the complete 2011 5-year inventory of the Great Lakes states of Minnesota, Wisconsin, and Michigan was 5.8:1.[2] However, the inventory is not biased systematically against dead trees. Dead trees are recognized as particularly important ecologically (Woodall et al. 2009), and standing dead trees are the subject of specific reporting since Field Guide 2.0 was published in 2004 (USDA Forest Service 2004).

Joint observations of live and dead trees contain important ecological information and increasing the size of the sample population also increases the resulting inferential space. However, a joint analysis of live and dead trees in FIADB yields a zero-inflated population, and statistical inference which does not

[1] Research Soil Scientist (CHP), U.S. Forest Service, Northern Research Station, 1992 Folwell Avenue, St. Paul, MN 55108; Ph.D. Candidate (PLZ), University of Minnesota, Minneapolis, MN. CHP is corresponding author: to contact, call 651-649-5191 or email at charleshperry@fs.fed.us.

[2] Calculations may be made using FIA's online tools available at http://fiatools.fs.fed.us.

account for zero-inflation is likely to be erroneous (Martin et al. 2005). Zero-inflation often can be accommodated by hurdle and mixture models when the additional zeros are "true zeros" (Martin et al. 2005). Hurdle models (also known as conditional models) treat the problem in two stages: first, the analyst determines the probability of a species or property being present or absent in a binary outcome; second, and conditional on its presence, the relative abundance of said species/property is found (Cameron and Trivedi 1998). Mixture models attempt to answer the same two questions in one model, but the resulting parameters are more challenging to interpret (Martin et al. 2005).

In this paper, we outline the application of a hurdle model approach to sugar maple mortality in the northern United States. Twenty states were included in the analysis: Connecticut, Delaware, Illinois, Indiana, Iowa, Maine, Maryland, Massachusetts, Michigan, Minnesota, Missouri, New Hampshire, New Jersey, New York, Ohio, Pennsylvania, Rhode Island, Vermont, West Virginia, and Wisconsin. Our study highlights important analytical considerations when using FIADB for analysis of forest health conditions and presents simple methods to create a more comprehensive space for statistical inference.

METHODS

Forest and soil inventory plots in the Northern United States were joined and extracted from the FIADB (Woudenberg et al. 2010). These data were collected between 2000 and 2006. Plots were included in the analysis if at least three sugar maple trees with d.b.h. greater than 5 inches were measured on the plot.

Plot information included state and county, and latitude, and longitude. Plot latitude and longitude were used to link plots to spatially explicit geologic databases describing the origin of surface material (Fullerton et al. 2003). Forest-level attributes included the basal area of live and dead sugar maple, ecological subsection, forest-type group, stand age, stand-size code (a classification of the predominant diameter class of live trees), slope, aspect, physiographic class (e.g., xeric, mesic, or hydric), and the presence/absence of disturbance on the plot. Soil plot information focused on the suite of soil chemistry variables extracted from mineral soil samples (O'Neill et al. 2005, Woodall et al. 2010) and their derivatives.

Statistical analyses were conducted in three stages: (0) ordinary linear regression on all plots to demonstrate the impact zero-inflation; 1) logistic regression on the presence and absence of dead sugar maple; and 2) ordinary linear regression of those plots with dead sugar maple. Given the exploratory nature of our investigation using the suite of variables available in FIADB, analyses were completed using stepwise techniques in R (R Development Core Team 2011). Appropriate variable transformations were suggested by Box-Cox analyses. Zeroes cannot be log transformed, so a very small number (0.001) was added to variables as required.

RESULTS

Stage 0

A total of 219 plots were selected that met the defined criteria of at least three sugar maple trees and the collection of soil chemistry data. A number of terms were available as predictors (Table 1). Our first effort was directed at modeling the fraction of dead sugar maple basal area as the response in a multiple regression model. If successful, this would be a simple and complete model of sugar maple mortality. This investigation collapsed because of the zero-inflated distribution; too many plots had zero dead sugar maple (Fig. 1).

Stage 1

Given our trouble with the zero-inflated fraction of dead sugar maple basal area in stage 0, we adopted hurdle modeling. Using the hurdle model, we modeled the data in two stages. In stage 1, we modeled the presence or absence of dead sugar maple using logistic

Table 1.—Variables available to predict sugar maple death across the northern United States

Site characteristics	Soil characteristics
Latitude, Longitude `{lat, lon}`	pH
Drought index `{di}`	sqrt(ECEC) `{secec}`
Ecoprovince `{eco}`	log(Ca:Al ratio) `{lca.al}`
Forest-type group `{forest}`	log(Mg:Al ratio) `{lmg.al}`
Basal area `{ba}`	log(Mg:Mn ratio) `{lmg.mn}`
Stand age `{age}`	log(Exchg. K percentage) `{lekp}`
Stand-size class `{size}`	log(Exchg. Na percentage) `{lesp}`
Site class `{site.class}`	log(Exchg. Ca percentage) `{lecp}`
Slope `{slope}`	log(Exchg. Mg percentage) `{lemp}`
Aspect `{aspect}`	log(Exchg. Al percentage) `{leap}`
Disturbance `{dist}`	
Geology `{geo}`	

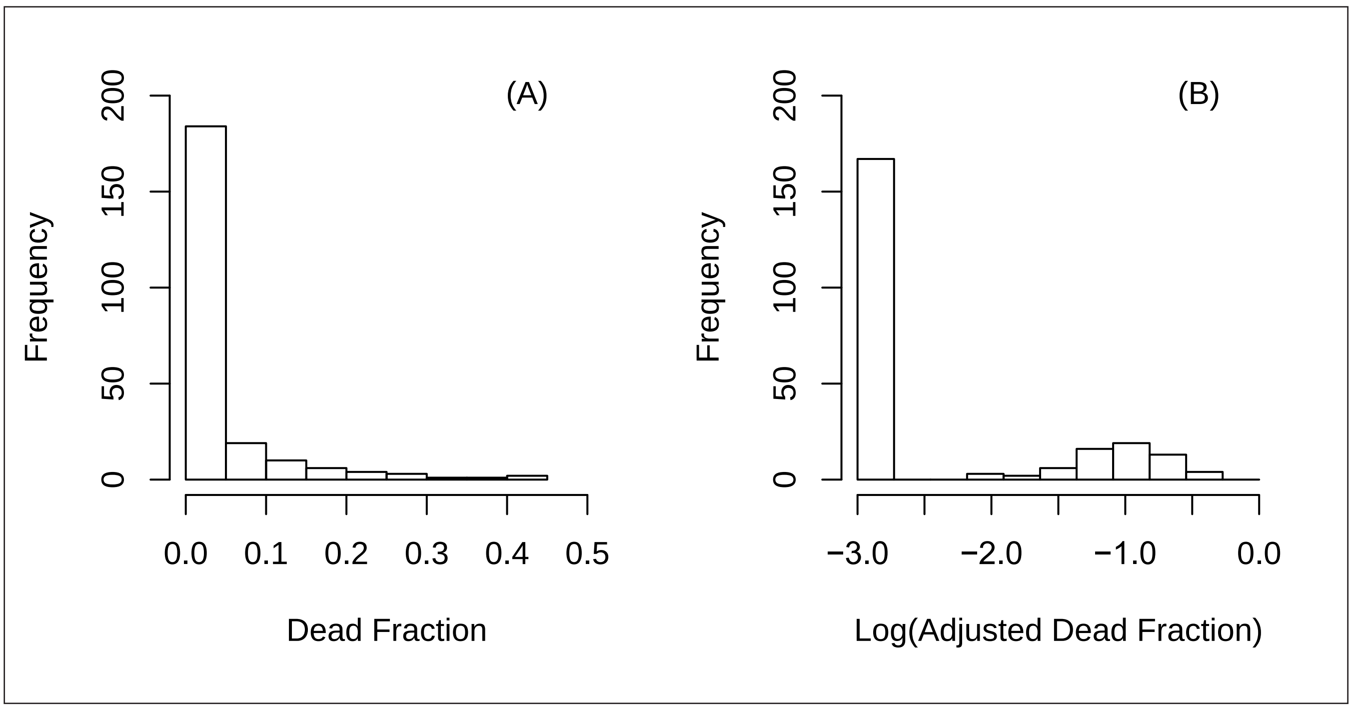

Figure 1.—Histograms of the observed dead fraction of sugar maple basal area on all plots (A) before and (B) after logarithmic transformations.

regression. In stage 2 (below), we modeled abundance given dead trees were present. Modeling presence or absence was accomplished using the binomial family of `glm()`. As before, 219 plots were available to parameterize the model, and the same terms were used as predictors (Table 1).

Exploratory binomial models were built using two starting points: 1) intercept only; and 2) a full model. Stepwise regression found the best model using AIC as the selection criteria. Similar models were selected from different starting points. The coefficients of the most plausible stage 1 model are included in Table 2.

Table 2.—Parameters for the most plausible model of sugar maple death using logistic regression with all plots

Variable Coefficient	Estimate	Std. Error	Z	Pr(>|z|)
ba	0.0197	0.0057	3.47	0.0005
lmg.mn	-0.5959	0.2198	-2.71	0.0067
geo:glacial	-0.3715	0.5001	-0.74	0.4576
geo:till	-1.0628	0.3698	-2.87	0.0041
geo:non-glacial	-2.5894	0.5593	-4.63	3.67e-06

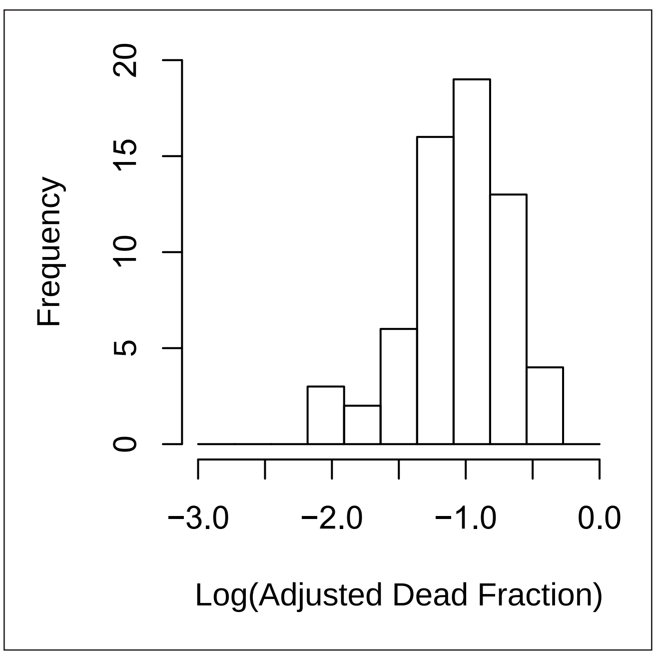

Figure 2.—Histogram of the observed dead fraction of sugar maple basal area for those plots with dead trees.

The interpretation of the intercepts in logistic regression is done using log-odds. Each unit increase in basal area increases the odds of dead basal area by a factor of 1.02. Each 10x increase in Mg:Mn reduces the odds of dead basal area to 55 percent of that for the original landscape. A till landscape reduces the odds of dead basal area to 50 percent of that for other glacial landscapes. A nonglacial landscape reduces the odds of dead basal area to 10 percent of that for a glacial (non-till) landscape.

Stage 2

In the second stage of the hurdle model, we modeled the amount of dead sugar maple basal area found on those plots that have dead sugar maple. We focused on the 58 points where dead sugar maple was observed (Fig. 2), representing 26 percent of the population of plots with sugar maple. The parameters from the most plausible model (below) are included in Table 3.

Table 3.—Parameters for the most plausible model of sugar maple death using linear regression with only those plots including dead trees

Variable Coefficient	Estimate	Std. Error	t value	Pr(>|t|)
Intercept[a]	-0.675	0.587	-1.150	0.257
lat	0.018	0.014	1.334	0.190
secec	0.077	0.031	2.473	0.018
lca.al	0.027	0.013	2.066	0.046
lmg.mn	-0.018	0.011	-1.636	0.110
lesp	0.080	0.022	3.674	0.001
lemp	-0.078	0.033	-2.362	0.023
forest (MBB)	-0.110	0.077	-1.428	0.162
forest (OH)	0.213	0.085	2.501	0.017
forest (Other)	-0.140	0.104	-1.340	0.188
age	0.002	0.001	1.660	0.105
size (Medium)	0.055	0.042	1.318	0.196
size (Small)	0.157	0.115	1.366	0.180
site.class (4)	0.025	0.103	0.244	0.808
site.class (5)	0.187	0.107	1.743	0.089
site.class (6)	0.111	0.111	1.004	0.322
dist	0.065	0.056	1.164	0.252
geo:nonglacial	0.045	0.098	0.455	0.652
geo:till	0.089	0.039	2.245	0.031
ba	-0.001	0.001	-2.543	0.016

[a]The model intercept includes forest (AB), size (Large), site (3), and geo (glacial, not till). Multiple R-squared: 0.6118, Adjusted R-squared: 0.4177, F-statistic: 3.152 on 19 and 38 DF, p-value: 0.001269.

The exploratory model developed in Stage 2 presents results affirming and challenging previous evaluations of sugar maple decline. Rising Mg levels (lmg.mn and lemp) are associated with declines in death, but contrary to expectations, increases in other forms of mineral soil nutrition available to trees (lecec, lca.al, and lesp) are associated with increasing death of sugar maple (Horsley et al. 2000, Long et al. 2009).

DISCUSSION AND CONCLUSIONS

Statistical inference in a hurdle model approach is complicated by the use of two stages of model building. In stage 1, we analyzed the full dataset using logistic regression, and log (odds) can be difficult to interpret. The model developed in the second stage is constructed more traditionally—by linear regression—so interpretation of the resulting coefficients is relatively straightforward. Additionally, while these two sets of models are similar, they are not identical.

Our emphasis here is to outline a process whereby more comprehensive datasets (namely those including both live and dead trees) can be used to evaluate the likelihood of sugar maple death across the species' range, so additional interpretations are being set aside for more thorough consideration in a subsequent manuscript. Given our use of AIC, multi-model inference will be a useful tool for assessing predictors within and potentially between the two stages (Burnham and Anderson 2002). Our key point is that hurdle models offer an opportunity to model comprehensive, zero-inflated datasets, like those collected by FIA, where the zero-inflation results from the presence of true zeros in the dataset (Martin et al. 2005).

LITERATURE CITED

Burnham, K.P.; Anderson, D.R. 2002. **Model selection and multimodel inference: a practical information-theoretic approach. 2nd ed.** New York, NY: Springer. 488 p.

Cameron, A.C.; Trivedi, P.K. 1998. **Regression analysis of count data.** Cambridge, UK: Cambridge University Press. 411 p.

Fullerton, D.S.; Bush, C.A.; Pennell, J.N. 2003. **Map of surficial deposits and materials in the eastern and central United States (east of 102 west longitude).** U.S. Geological Survey Geologic Investigations Series I-2789. Denver, CO: U.S. Geological Survey.

Horsley, S.B.; Long, R.P.; Bailey, S.W.; Hallett, R.A.; Hall, T.J. 2000. **Factors associated with the decline disease of sugar maple on the Allegheny Plateau.** Canadian Journal of Forest Research. 30: 1365-1378.

Long, R.P.; Horsley, S.B.; Hallett, R.A.; Bailey, S.W. 2009. **Sugar maple growth in relation to nutrition and stress in the northeastern United States.** Ecological Applications. 19(6): 1454-1466.

Martin, T.G.; Wintle, B.A.; Rhodes, J.R.; Kuhnert, P.M.; Field, S.A. [et al.]. 2005. **Zero tolerance ecology: improving ecological inference by modelling the source of zero observations.** Ecology Letters. 8(11): 1235-1246.

O'Neill, K.P.; Amacher, M.C.; Perry, C.H. 2005. **Soils as an indicator of forest health: a guide to the collection, analysis, and interpretation of soil indicator data in the Forest Inventory and Analysis program.** Gen. Tech. Rep. NC-258. St. Paul, MN: U.S. Department of Agriculture, Forest Service, North Central Research Station. 53 p.

R Development Core Team. 2011. **R: a language and environment for statistical computing.** Vienna, Austria: R Foundation for Statistical Computing.

USDA Forest Service. 2004. **Forest Inventory and Analysis national core field guide. Volume 1: Field data collection procedures for phase 2 plots, version 2.0 [online].** St. Paul, MN. Available at http://fia.fs.fed.us/library/database-documentation/. (Accessed August 15, 2012).

Woodall, C.W.; Conkling, B.L.; Amacher, M.C.; Coulston, J.W.; Jovan, S. [et al.]. 2010. **The Forest Inventory and Analysis database version 4.0: Database description and users manual for phase 3.** Newtown Square, PA, U.S. Department of Agriculture, Forest Service, Northern Research Station. 180 p.

Woodall, C.W.; Smith, J.E.; Miles, P.D. 2009. **Attributes of standing dead trees in forests of the United States.** In: McRoberts, R.E.; Reams, G.A.; Van Deusen, P.C.; McWilliams, W.H., eds. Proceedings of the eighth annual Forest Inventory and Analysis symposium; 2006 October 16-19; Monterey, CA. Gen. Tech. Rep. WO-79. Washington, DC: U.S. Department of Agriculture, Forest Service: 191-196.

Woudenberg, S.W.; Conkling, B.L.; O'Connell, B.M.; LaPoint, E.B.; Turner, J.A. [et al.]. 2010. **The Forest Inventory and Analysis database: database description and users manual for version 4.0 for phase 2.** Gen. Tech. Rep. RMRS-245. Fort Collins, CO: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 336 p.