A PRACTICAL RESEARCH DIGEST FOR TURF MANAGERS

TURFGRASS TRENDS

Volume 10, Issue 9 • September 2001

SEED SELECTION

How Predictable is NTEP Data for Your Particular Site?

By Doug Brede, Ph.D.

Last fall I worked with a client in Edmonton, Alberta, Canada, who was planting a golf course to Kentucky bluegrass. I asked if he had consulted data from the National Turfgrass Evaluation Program (NTEP) to aid in his decision, and he said he had. The client explained that he had chosen a handful of varieties from the top of the Grand Mean column and wanted to construct a blend.

Sound familiar? This same scenario plays out in locations across the continent all the time. Contractors, landscape architects, and turf managers consult the NTEP listings as a routine part of their planting plans. But the question remains: Is this the best way to pick varieties for your site?

In this article I'm going to examine some of the relationships buried inside the NTEP data. Most people who use NTEP data look at just the single column of Grand Mean averages for recommendations. But is this the right thing to do? Or are there idiosyncrasies hidden within the statistics that may paint a misleading picture? I will show you what some of these rating values really mean by examining underlying interrelations among the variables.

First, I'm going to explain some of the more confusing concepts within NTEP, such as the differences and similarities between such things as density and texture. (Does anyone really know the difference between those two?) By doing so, I'll provide insights into the thought-processes of the raters and the meaning of their results.

Next, I'll show you why you may be making a giant mistake by following the Grand Mean Quality results for your variety recommendation needs – as my Edmonton client later discovered.

Hidden interrelationships in NTEP data

Whenever I tell one of my non-turf colleagues about the NTEP trials – our "yardstick" of turf breeding – the question invariably comes up: What kind of meters do you use to take the readings? Most scientists are accustomed to carrying gadgets and gizmos with them to measure things. My non-turf colleagues are always surprised to learn that there are no such gadgets with turf. Every measurement in the NTEP trial is based on eyeball estimates.

To those of you familiar with the process, this comes as no surprise. But it may surprise you to learn that some of these visual estimates are strongly interrelated. Many are highly correlated: Factor A influences the rater's judgment on Factor B.

To explore these interrelationships, I downloaded tables from the 2000 results of the 1995 Kentucky bluegrass trial from NTEP's web site (www.ntep.org). I used a software package called Statistica to analyze the data. However you can do many of the same manipulations with Microsoft Excel on your desktop.

AN ADVANSTAR ★ PUBLICATION

IN THIS ISSUE

TURFGRASS TRENDS

Executive Editor
Sue Gibson
440/891-2729; 440/891-2675 (fax)
sgibson@advanstar.com

Managing Editor
Curt Harler
440/238-4556; 440/238-4116
curt@curtharler.com

On Line Editor
Lynne Brakeman

Senior Science Editor
Dr. Karl Danneberger

Group Editor
Vern Henry

Production Manager
Rene' Fall
218/723-9352; 218/723-9223 (fax)
rfall@advanstar.com

Senior Graphic designer
Jeff Landis
440/891-2702; 440/891-2675 (fax)
jlandis@advanstar.com

Circulation Manager
Cheryl Beeman
218/723-9271; 218/723-9433 (fax)
cbeeman@advanstar.com

Group Publisher
John D. Payne
440/891-2786; 440/891-2675 (fax)
jpayne@advanstar.com

Corporate & Editorial Office
7500 Old Oak Blvd.
Cleveland, OH 44130-3369

New Subscriptions
888/527-7008

Abstracts: 800/466-8443
Reprint: 440/891-2744
Permission: 440/891-2742
Single copy or back issues:
Subscription/Customer Service
888/527-7008; (fax) 218/723-9437
Web site address:
www.landscapemanagement.net

ADVANSTAR
COMMUNICATIONS

Chairman and Chief Executive Officer
Robert L. Krakoff

Vice Chairman & CTO
James M. Alic

President & COO
Joe Loggia

VP-Finance, CFO & Secretary
David W. Montgomery

Executive Vice Presidents
Alexander S. DeBarr
Jean-Marc Levy
Daniel M. Phillips

Vice President & General Counsel
Eric I. Lisman

Treasurer and Controller
Adele D. Hartwick

Are raters color-blind?

One of the classic relationships in NTEP is between color and quality. Raters I've spoken with take pride in the fact that they don't let the color of a grass taint their judgment when rating turfgrass quality. Most assert that a dense, pest free, light green grass would be rated just as highly as a good dark one. Or are they swayed? When I plotted the genetic color versus the Grand Quality Mean, a strong relationship appeared (Fig. 2a).

The graphs in Figure 2 display data points of all 103 varieties in the 1995-2000 trial. I labeled a handful of landmark varieties to establish mileposts in the sea of dots: KenBlue (a common-type variety), Classic and Baron (two older, intermediate types), Eclipse and Glade (top varieties from the 1980's), Limousine (a high density variety), and Award and Midnight (varieties presently at the top of the quality charts).

The computer did not draw a straight line for the relationship of color versus quality but one with a bowed center (Fig. 2a). Nonetheless, darker color does appear to influence higher quality scores. KenBlue, one of the lightest colored varieties in the trial, also had the lowest turf quality. Award and Midnight both had high quality and dark color.

Certainly there are other explanations for this connection of color and quality. An argument can be made that dark color has a physiological benefit to the plant. A darker plant, it's been shown, contains more chlorophyll – the energy compound in plants. With more energy, darker green varieties are able to grow faster, produce a denser turf, and regrow foliage lost to mowing, disease, and wear. Therefore two associations are at play: A preference by the raters for darker color, and a physiological advantage to the plant from more chlorophyll.

Are raters dense?

Turf density and texture are among the most misunderstood ratings in NTEP. In theory, density reflects the number of plants per square inch. But no one actually gets down on their hands and knees to count. We stand and judge. And by doing so, we get confounding results. For example, how do you tell if a plot has more plants per square inch, or whether it has more leaves per plant? You can't.

Leaf texture is an evaluation of the width of individual blades. In leaf texture ratings, finer-bladed varieties are scored higher. Again, no one gets out a ruler to measure leaf width (which would be the logical but time-consuming way to approach the problem). Instead, we stand and judge.

One misleading assumption in leaf texture ratings is that finer texture is more desirable. After all, why would finer texture ratings have a higher number if it didn't mean narrower is better? Personally, I prefer a variety with a leaf texture rating of about a "7" (on a 1 to 9 ratings scale). I think it is entirely possible for a variety to be too finely bladed, possibly sacrificing toughness, wear tolerance, or mixability with others. Other evaluators may feel differently.

In Figure 2b you can observe the tight clustering of leaf texture with turf density about the slope line. As the example in Figure 1 shows, a tighter adherence to the slope line indicates a stronger relationship and better predictability. The main difference between texture and density is in point spread: Leaf texture has a 3-point ratings spread (from 5 to 8) from best to worst, while density has only a 1-point spread. Could it be that the raters are more comfortable with the concept of texture than density? It's hard to say for sure, but that's a possibility.

Turf density has two other interesting associations, those being with ground coverage and disease. The skin-tight clustering of the points about the slope line (Fig. 2d) indicates that density and ground coverage are virtually synonymous. Over the years I've questioned whether the "ground coverage" rating was even necessary. These results suggest that either the raters can't distinguish between the two, or that density so affects ground coverage as to make it superfluous.

I must admit, at first sight, the relationship between turf density and disease resistance (Fig. 2c) caught me by surprise. Classical plant pathology says that stands with higher plant densities tend to get more disease. That's because typically, denser stands have smaller, frailer plants, easily prone to fungal attack and spread. While there is a fairly good association between density and leafspot resistance (as evidenced by the clustering), the surprise was that slope of the line was positive, not negative. If a denser stand was truly more disease prone, the line would slope downward not up.

ASSIGNING A NUMBER TO PREDICTABILITY

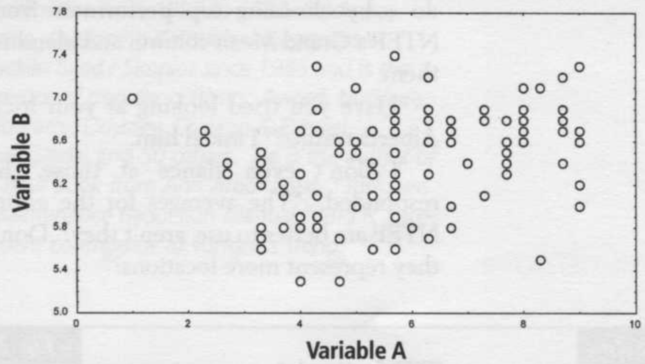
The science of statistics is all about assigning numbers to things that happen in nature. Like anything else, predictability can be quantified and assigned a meaningful value. Take, for example, my golf scores. Based on my past research, my success rate for driving straight down the fairway tends to increase with the hole number. I nearly always flub the first tee shot. But the longer I play, the better I get and the greater my success of hitting a straight drive.

This relationship can be graphed, showing a fairly straight line between my tee-off success rate and the hole number. Of course, not every data point falls on a straight line. Towards the end of the match, my drives again tend to stray, as fatigue and Miller Genuine Draft takes effect.

Back in high school, I remember my math teacher demonstrating a way to take a straightedge, estimate the best fit through a clump of data points, and draw a pencil line through the middle. The straight line represents the relationship between the hole number and the score. Statisticians have an even classier way of doing this, called the Least-Squares Method. Using a computer program, the computer digests the data points and constructs the straight-line relationship using a mathematical method of Best Fit. You find Least-Squares programs running behind the scenes in popular software programs such as Microsoft Excel's graphing routine.

Software can even estimate the Goodness of Fit or predictability of that line. If my 18 golf data points all fell exactly on a straight line, the percent fit would be 100% (see graph below). Of course, 100% predictability rarely happens in nature. More often, you have no relationship, or something in between. Where no relationship exists, the data points form a "glob" on the X-Y scatterplot. Better fits are illustrated by a tighter clustering to the slope line. The degree of fit can be expressed as a number value, or r2 value, expressing the percent predictability.

Scatterplot example of "No Relationship"



Scatterplot example of an "Excellent Relationship"

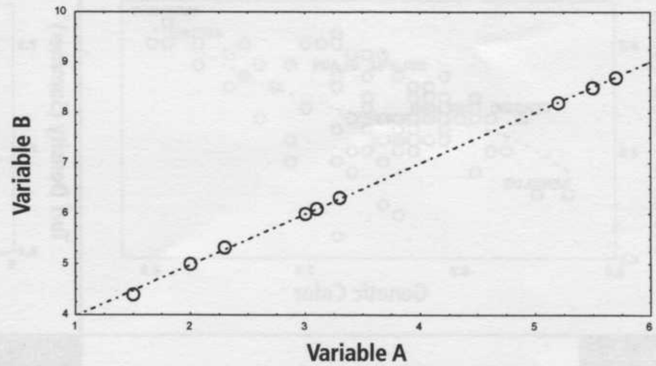


Figure 1. Scatterplots are handy ways for examining relationships between two sets of numbers. Let's say you survey 100 people, measuring their height and their cholesterol level. By plotting each point, with their height on one axis and their cholesterol reading on the other, you might get a graph like the one below (top chart). This scatterplot illustrates a classic "no relationship" response. In other words, taller people do not have a tendency toward higher cholesterol. Now let's say you surveyed the same people, recording the height at the top of their head versus the height of their shoulders. These two variables are obviously interrelated and illustrate a nice, straight-line relationship, as shown in the lower graph.

Here's what I believe is occurring: Disease-resistant cultivars are simply able to produce more shoots than susceptible ones. Varieties like Award and Midnight, which are nearly immune to leafspot, are not encumbered by the thinning of disease attack. These varieties help illustrate the real reason behind the positive relationship of density and disease resistance.

How useful is "grand quality mean?"

My Alberta client put together a bluegrass blend for his golf course the way most of us do – by choosing top performers from NTEP's Grand Mean column and blending them.

"Have you tried looking at your local Alberta results?" I asked him.

"I don't even glance at those," he responded. "The averages for the entire NTEP are better to use, aren't they? Don't they represent more locations?"

After that conversation, I did some thinking about whether his approach was right or wrong. Unable to reach a conclusion, I decided to let statistics help me find an answer. Using the same data mentioned above, I compared the individual State/Province averages of the 103 bluegrass varieties versus the Grand Quality Mean. I used Statistica to calculate the "predictability" of each State/Province versus the Grand Mean. A predictability value of 100% would indicate that the particular State's mean was exactly shadowing the Grand Mean and the user could consult either result with equal certainty.

Among the 26 sites in the trial, the NJ1 site (New Brunswick, NJ) gave the best correlation with the Grand Mean, with 68% predictability. Turf managers in New Jersey can probably utilize either their State results or the Grand Mean with fairly equal implications. (It makes you wonder why we don't just have Drs. Bill Meyer and Reed

Fig. 2a

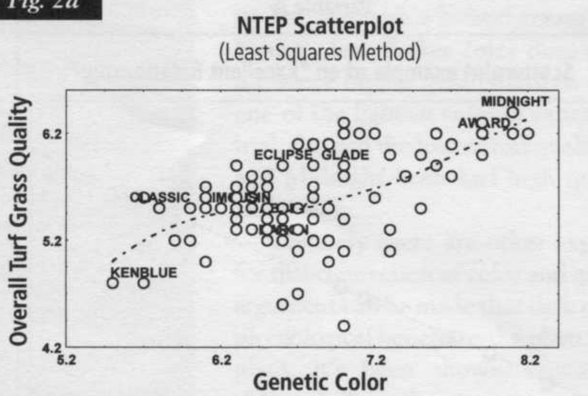


Fig. 2b

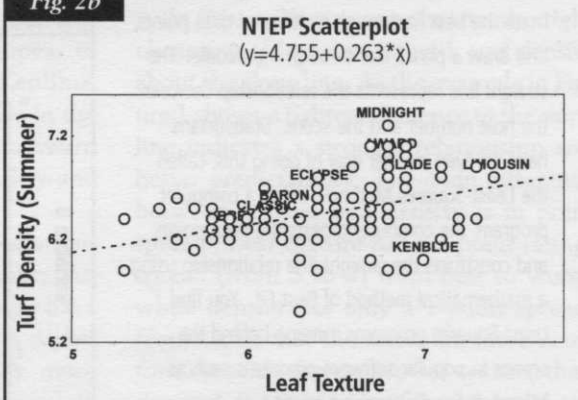


Fig. 2c

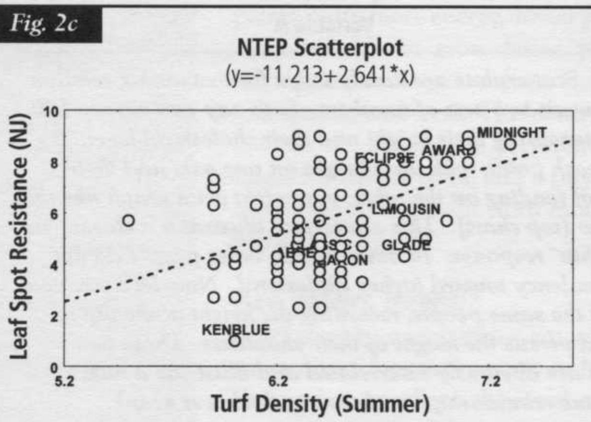
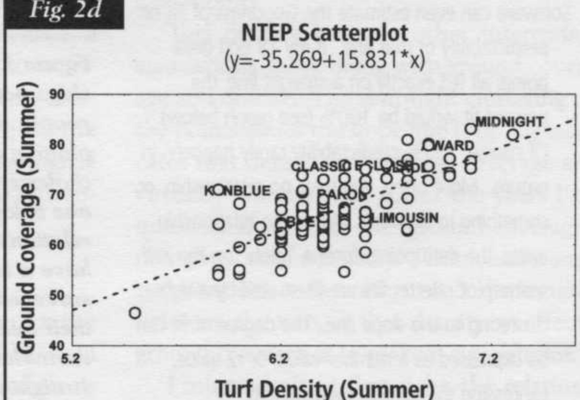


Fig. 2d



Funk at Rutgers do the whole NTEP there themselves!) Curiously, a second site in New Jersey (Adelphia) gave just 34% predictability. Minnesota also had a strong positive correlation. Years ago when I ran a similar analysis on the 1985-1990 trial, there was a negative correlation between the Minnesota site and the Grand Mean. In other words, varieties that did well in the Minnesota trial, tended to do poorly nationally. Strange but true.

But getting back to my Edmonton client, I found there was absolutely zero predictability between his local Alberta site and the NTEP Grand Mean. Yet, this fellow was taking the Grand Mean as gospel, downplaying the need to even glance at his local site results.

Examples like that force me to conclude that the Grand Mean may be more of an albatross than a benefit – especially when it misleads people more than it helps. Clearly half of NTEP's sites predict one-third or less of the variability in the Grand Mean. My Edmonton colleague would have been far better served to consult his local site data and not even glance at the national results.

Does that mean that certain State data are wrong or even bad? Not at all. It means the results are State specific. Data from New England and some Midwestern states correlated closely with the Grand Mean, showing high levels of association. Canada, the Mid-Atlantic region, Iowa, and the West correlated poorly with National averages. Turf managers in those areas should preferentially take the State readings over the Grand Mean.

NTEP is presently grappling over dispensing with the Grand Mean column and emphasizing individual State/Province results. My advice to you: If your state has a predictability of less than 50% (Fig. 3), I'd stick with your State results and forget about the Grand Quality Mean.

Doug Brede has had a long association with NTEP, dating back to 1979 when he attended a planning meeting at Rutgers University to establish the initial protocols for NTEP. Even before that time, he was an evaluator for Penn State University's plots of Project NE-57, which was the precursor of the modern NTEP trial. Brede was an evaluator and host site for NTEP trials from 1980 (NTEP's inception) until 1994, when trials at private companies were discontinued. He served on NTEP's Policy Committee from 1997 to 1999. Brede has been developing Kentucky bluegrasses at Jacklin Seed / Simplot since 1986 and is the creator of popular cultivars, Award, NuGlade, Liberator, Odyssey, Chicago II, Everest, EverGlade, and 50 others. He is the author of a new book from Ann Arbor Press, "Turfgrass Maintenance Reduction Manual," and is a frequent contributor to Turfgrass Trends