A PRACTICAL RESEARCH DIGEST FOR TURF MANAGERS

TURFGRASS TRENDS

Section II • Volume 11, Issue 8 • **August 2002**

www.turfgrasstrends.com

SEED TESTING

Scientists Look For Ways To Improve Turf Evaluations

By J.S. Ebdon

Evaluating turfgrass quality is by no means an exact science. As it stands now, the NTEP results, which are based on Grand Mean scores averaged across three replications at different locations, are the most accurate way to predict how turfgrass cultivars will behave on your course.

As with all scientific endeavors, however, we in academia are always looking to refine our evaluation methods to produce even more accuracy than is currently available. That's why several different statistical models are currently being tested to see if we can improve on NTEP's accuracy even further in the next five years.

One such family of models that I, along with my colleague Hugh Gauch Jr. at Cornell University, consider to be more appropriate for NTEP data are Additive Main effect and Multiplicative Interaction (AMMI) models.

Why accuracy is important

Monthly performance data, which is collected by university and industry researchers and breeders, is summarized in NTEP Final Reports. These reports contain turf-quality data for each variety for trials repeated at many locations that are then evaluated over several years.

The data, derived from arithmetical means, are then used by turf-extension specialists in making recommendations on what turf varieties work in which parts of the country. The varieties that ranked near the top based on past performance using NTEP means, are then recommended for planting at specific locations.

It's assumed that using arithmetical means to summarize turfgrass performance provide the most accurate prediction of future performance, but statistical theory and recent research indicate otherwise. Arithmetical means are, in fact, less accurate than other options that will be at the disposal of turfgrass researchers as their methodologies are refined.

Accuracy is an integral part of agricultural science and research. Furthermore, accuracy is also important to the turfgrass manager because greater accuracy ensures more reliable prediction and better turfgrass recommendations.

Therefore, predictive accuracy is important to both the science and management of turf. Improved accuracy can be achieved using different strategies, though some are economically more sensible than others.

Improving accuracy

The major purpose of NTEP variety trials is to provide accurate estimates of turfgrass quality for each varieties growing in different environments. The growing environment is represented by the various NTEP locations, which usually range between 20 to 30 depending on the test.

Since turfgrasses are perennials, the evaluation period is conducted over several years at each location. NTEP selects locations of diverse soil textures, pHs, cultural practices and climatic conditions to test genotypes in specific environments. By evaluating trials over several years, it's possible to capture information caused by year-to-year variation in the growing season. Consequently, genotypes are evaluated over many location-by-year combinations to account for as much information as is economically feasible to ensure the most reliable recommendations.

> If NTEP wanted to double the accuracy of turf-quality estimates, co-operators would plant 12 replicates instead of the usual three. This strategy is neither economical for seed companies nor practical for investigators.

NTEP relies on raw means, observations averaged over replications, to estimate and to predict turfgrass performance for genotypes growing in different environments. Replication is used as the basis for quantifying the variety trials' experimental accuracy. Statistical theory tells us that the accuracy of a mean to predict issues such as turf quality increases with the number of replications and is related to the square root of the number of replications (Snedecor and Cochran, 1989).

NTEP uses three replications for each variety to estimate turfgrass quality. Therefore, if NTEP wanted to double the accuracy of turf-quality estimates, co-operators would then be encouraged to plant 12 replicates instead of the usual three. This strategy, however, of planting four times the number of replicates is neither economical for seed companies nor practical for investigators.

AMMI predictions are more accurate because they use more data. Secondly, AMMI means are more accurate than the raw means because the analysis is able to partition out noise in the data.

Turfgrass quality is a visual assessment (1 to 9 scale, with 9 representing the highest quality) used by NTEP evaluators to measure turfgrass performance. Such evaluations are subject to the personal biases and preferences of the evaluator, which introduces error into the measurements and reduces accuracy.

Additionally, error (noise) can be caused by natural variation in soil texture or may be caused by nonuniform fertilization or irrigation practices. Noise is especially problematic for large NTEP trials. For example, it's not unusual for NTEP to evaluate at least 100 genotypes that are planted in 100 or more environments, resulting in 10,000 treatment combinations and 30,000 observations.

In such large research trials, noise accumulates rapidly and the pattern quickly becomes buried. Without noise, all replicates would be identical, and turfgrass quality estimates for various treatment combinations would be exact.

Noise (error) causes serious discrepancies between turf-quality estimates and accurate prediction of future performance. Near optimal predictions can be achieved then by minimizing noise to its lowest possible level. This gain in accuracy by noise reduction is independent of the gains in accuracy that can be achieved by increasing replications as described earlier.

The ability to reduce noise statistically was not possible in agricultural research until recently. It is important to understand that these gains in accuracy come at little or no cost to NTEP compared to the alternative strategy of increasing replications.

Assessing predictive data accuracy

The NTEP model (i.e., cell means model) considers as relevant data the mean of three observations for predicting the future performance of genotypes growing in specific environments. NTEP variety trials are replicated both in time and space, and therefore genotypes may be growing in 100 or more environments. However, the cell-means model does not consider this data to be relevant in predicting the future performance of genotypes growing in environments. Because the future is uncertain, it would not be good advice to bet on future performance based on only three observations while ignoring all other relevant data.

TABLE 1

Model accuracy in predicting future turfgrass performance (turfgrass quality): Validation results for the 1990 NTEP Kentucky bluegrass and perennial ryegrass trials.

	KENTUCKY BLUEGRASS TRIAL		PERENNIAL RYEGRASS TRIAL	
Model†	Statistical efficiency‡	Free observations§	Statistical efficiency‡	Free observations§
AMMI	2.05	27,169	5.60	101,844
Cell means	1.00	0	1.00	0

†Most accurate AMMI model compared to the cell means model (means averaged over replicates). NTEP bases predictions of future turfgrass performance using the cell means model.

‡Statistical efficiency is the gain in accuracy (in equivalent replications) afforded by the most accurate AMMI model relative to the standard NTEP model (cell means).

§Free observations are the required number of additional observations needed using the standard NTEP model (cell means) to achieve the same accuracy provide by the most accurate AMMI model.

TABLE 2

Summary of turf quality (TQ) differences between cell means (means averaged over replicates) and AMMI estimates for the 1990 Kentucky bluegrass variety trial.

	MODEL	
Statistic	**Cell means (NTEP model)**	**Adjusted means (AMMI model)**
Environments (no.)	69	
Winners (no.)†	29	18
Avg.. TQ loss (-) or gain (+)	-0.4‡	+0.3§
Maximum TQ gain	–	+1.4¶
Same winners	19 of 69 environments (27.5%)	
Different winners	50 of 69 environments (72.5%)	

†Total number of winners across all environments based on individual winning genotypes (top ranked within each environment) identified by a corresponding model (AMMI vs. cell means).

‡Average loss in turfgrass quality by choosing NTEP winners (based on raw means) over AMMI winners.

§Average gain in turfgrass quality by choosing AMMI winners over NTEP winners (based on raw means).

¶Maximum gain in turfgrass quality by choosing AMMI winners over NTEP winners (based on raw means).

NTEP relies on the cell-means model that is simplistic, but is often complicated by noise. This method elevates some genotypes to a higher position and increases the number of winning genotypes unnecessarily, which complicates recommendations. There are, however, alternative statistical models that may be more effective in achieving NTEP goals by providing more reliable recommendations.

The AMMI models have advantages over ordinary cell-means models in that AMMI considers the entire data set to be relevant in predicting future performance by fitting a multivariate model to calculate its turfgrass quality estimate. Where the cell-means model uses a simplistic arithmetic mean that can be calculated with a hand calculator, the AMMI prediction of future performance requires millions of mathematical steps, which requires only a few seconds of microcomputer time. As a result, AMMI predictions are more accurate because they use more data.

Secondly, AMMI means are more accurate than the raw means because the analysis is able to partition out noise in the data, thus improving accuracy by noise reduction.

Because of this reduction, AMMI predictions are adjusted appropriately. The statistical principles underlying AMMI theory and gains in accuracy have been available as early as the 1960s. However, it was not until the introduction of high-speed microcomputers did AMMI theory become accessible for turfgrass research. AMMI gains in predictive accuracy have a strong and compelling theoretical basis, but this theory has also been validated by equaling convincing empirical evidence.

Predictive accuracy of AMMI vs. NTEP model

Raw data from the 1990 NTEP Kentucky Bluegrass test and the 1990 Perennial Ryegrass test were kindly provided by NTEP to test the predictive accuracy of AMMI vs. competing cell-means model used by NTEP.

In the Kentucky bluegrass data set, 25,875 observations were analyzed (125 genotypes in 69 environments over three replicates). A similarly large data set was analyzed using the perennial ryegrass trial and included 22,140 GER observations (123 genotypes in 60 environments over three replicates).

The simplest and most effective strategy for assessing accuracy of various models is by data splitting resulting in two parts: modeling data and validation data. The modeling data is used to construct the model (i.e., cell means vs. AMMI) and in turn to estimate turfgrass quality. The validation set is used to assess accuracy. The data splitting into two parts is conducted by randomization for each treatment by selecting two observations for modeling and one observation for validation.

Randomization ensures each observation is given equal chance for selection. The model that comes closest to the validation set is the model that is the most accurate in predicting future performance.

For each data set (bluegrass and ryegrass), this valida-

TABLE 3

Summary of turf-quality (TQ) differences between cell means (means averaged over replicates) and AMMI estimates for the 1990 perennial ryegrass variety trial.

	MODEL	
Statistic	**Cell means (NTEP model)**	**Adjusted means (AMMI model)**
Environments (no.)	60	
Winners (no.)†	34	18
Avg. TQ loss (-) or gain (+)	-0.4‡	+0.2§
Maximum TQ gain	–	+0.9¶
Same winners	6 of 60 environments (10%)	
Different winners	54 of 60 environments (90%)	

†Total number of winners across all environments based on individual winning genotypes (top ranked within each environment) identified by a corresponding model (AMMI vs. cell means).

‡Average loss in turfgrass quality by choosing NTEP winners (based on raw means) over AMMI winners.

§Average gain in turfgrass quality by choosing AMMI winners over NTEP winners (based on raw means).

¶Maximum gain in turfgrass quality by choosing AMMI winners over NTEP winners (based on raw means).

tion procedure was repeated with thousands of different randomizations, so more than 8 million separate validation steps were performed. The results were then averaged. The AMMI gain in accuracy was expressed relative to the predictive accuracy of the standard model used by NTEP.

Estimates of noise in the data showed that the bluegrass data set had approximately 33 percent noise while the ryegrass data set was 41 percent noise (Ebdon and Gauch, 2002). Further analysis also showed that AMMI was effective in removing most of this noise. Therefore, it provided more accurate data. To that end, Table 1 summarizes the gains in accuracy with AMMI relative to the competing cell-means model.

To achieve the same accuracy without AMMI, NTEP would have to double the number of replications from three to six for the bluegrass trial (AMMI statistical efficiency of 2, Table 1). Similarly for the ryegrass trial, the number of replicates would have to be increased by a factor of 5.6 (from three to 17 replications).

These gains in accuracy with AMMI come at no additional cost and are equivalent to more than 25,000 free observations for bluegrass and more than 100,000 free observations for ryegrass (Table 1). The cost over the entire evaluation test to achieve the same accuracy by increasing the number of replications is equivalent to $271,690 (for bluegrass) and $1,018,440 (for ryegrass), assuming $10/observation/year.

As mentioned earlier, increasing accuracy by increasing the number of replicates is not free. Therefore AMMI offers a potential cost-efficient alternative. Furthermore, the AMMI gain in accuracy for the perennial ryegrass trial more than doubles the accuracy possible using the standard NTEP practice.

To achieve maximum efficiency and accuracy, these studies suggest that priority then should be given to the more accurate AMMI prediction of turfgrass quality. Greater accuracy with AMMI implies better selection of superior genotypes and more reliable recommendations. Therefore, by giving priority to AMMI winners over raw data winners increases in turfgrass quality would be expected by planting AMMI winning genotypes.

Tables 2 and 3 summarize turf quality differences between the various models for the 1990 Kentucky Bluegrass and Perennial Ryegrass trials. These gains in turf quality by planting AMMI winners over raw data winners shown in Tables 2 and 3 can be as large as .9 to 1.4 on a rating scale of 1 to 9. While these differences focus on winning genotypes identified by competing models, differences can also be found throughout the entire roster. AMMI estimates often lead to different ranking of genotypes within each environment (Figure 1, on the last page).

For example, in the ryegrass trial (Table 3), the two models picked the same winners in only 10 percent of the 60 environments. Similar inconsistencies were also observed between models for the bluegrass trial (Table 2). Because noise increases the complexity of the data by increasing the number of winning genotypes, it's not surprising that the noise-rich cell-means model identified 1.5 to three times as many winners compared to the AMMI model (Tables 2 and 3).

So, which statistical model predictions are to be more trusted? The answer is AMMI, because:

The cost over the entire evaluation test period to achieve the same accuracy by increasing the number of replications is equivalent to $271,690 for bluegrass and $1,018,440 for ryegrass.

- AMMI predictions are based on sound statistical theory;
- the theory is validated by compelling experimental evidence; and
- the AMMI results reported here with turf variety trials are consistent with results observed from yield trials (Gauch, 1992).

The next stage in this research is to conduct comparative studies in the field in order to validate these findings. Based on field validation studies from yield trials (Annicchairico, 1992), research indicates that AMMI winners should provide superior turf quality in comparison to those winners suggested by NTEP means.

Acknowledgment

The author thanks NTEP for financial support and for providing the raw data that made this research possible. Special thanks to my colleague Hugh Gauch Jr. for his invaluable mentoring.

J. S. Ebdon is an assistant professor of turfgrass management at the University of Massachusetts in Amherst, Mass.

REFERENCES

Annicchiarico, P. 1992. "Cultivar adaptation and recommendation from alfalfa trials in northern Italy." *J. Genetics and Breeding*. 46:269-277.

Ebdon, J. S., and H. G. Gauch, Jr. 2002. "AMMI analysis of national turfgrass performance trials: II. Cultivar recommendations." *Crop Sci.*, 42:497-506).

Gauch, H. G., Jr. 1992. *Statistical analysis of regional yield trials: AMMI analysis of factorial designs.* Elsevier, Amsterdam, The Netherlands.

Snedecor, G. W., and W. G. Cochran. 1989. *Statistical methods*. Eighth edition. Iowa State University Press. Ames, Iowa.

Figure 1. Accuracy alters genotype ranking and turfgrass selection. NTEP predictions (means averaged over replicates) versus the more predictively accurate AMMI estimates for perennial ryegrass turfgrass quality (for Pullman, WA in 1994). Pullman, WA represents an average (typical) NTEP location-year combination. Weighted lines identify individual genotypes. Lynx perennial ryegrass wins based on NTEP predictions while AMMI ranks Prelude II first. For this location-year, genotype Danilo shows the largest loss in ranking (and turf quality) from 8th to 105th (out of 123 genotypes). Similar results were also observed for the Kentucky bluegrass variety trial.