REGULAR ARTICLE

Combining partially ranked data in plant breeding and biology: II. Analysis with Rasch model

Ivan Simko[1]*, John M. Linacre[2]

[1] USDA-ARS, Crop Improvement and Protection Research Unit, 1636 East Alisal Street, Salinas, CA 93905, USA.
[2] Winsteps.com, 14525 SW Millikan Way # 27615, Beaverton, OR 97005, USA.
* Corresponding author: Ivan Simko, E-mail: Ivan.Simko@ars.usda.gov

ABSTRACT

Many years of breeding experiments, germplasm screening, and molecular biologic experimentation have generated volumes of sequence, genotype, and phenotype information that have been stored in public data repositories. These resources afford genetic and genomic researchers the opportunity to handle and analyze raw data from multiple laboratories and study groups whose research interests revolve around a common or closely related trait. However, although such data sets are widely available for secondary analysis, their heterogeneous nature often precludes their direct combination and joint exploration. Integration of phenotype information across multiple studies and databases is challenging due to variations in the measurement instruments, endpoint classifications, and biological material employed by each investigator. In the present work, we demonstrate how Rasch measurement model can surmount these problems. The model allows incorporating data sets with partially overlapping variables, large numbers of missing data points and dissimilar ratings of phenotypic endpoints. The model also enables quantifying the extent of heterogeneity between data sets. Biologists can use the model in a data-mining process to obtain combined ratings from various databases and other sources. Subsequently, these ratings can be used for selecting desirable material or (in combination with genotypic information) for mapping genes involved in the particular trait. The model is not limited to genetics and breeding and can be applied in many other areas of biology and agriculture.

Key Words: *Aggregated ranking; Bradley-Terry model; combining data; rank-order.*

INTRODUCTION

In genetics, biology, or breeding, it is often useful to combine observations from numerous experiments into a single dataset. For example, in association mapping studies, phenotypic data used for the analysis are often collected from experiments conducted over the course of several years or are obtained from germplasm databases (e.g. Simko et al., 2004a; Simko et al., 2004b; Simko and Hu, 2008).

However, combining data from different years, locations, laboratories, and databases is challenging, because not all of the independent variables (e.g. plant accessions) will be common across all experiments. Moreover, laboratories often use their own ratings scales that cannot be combined by standard statistical approaches. This situation creates the need to develop methodologies that would allow combining datasets with only a partial overlap and dissimilar rating scales

To combine data from dissimilar rating scales into a single aggregated rating, the absolute values from each test might be replaced with relative rankings. If there are two or more rankings of the same elements, then there may be enough information to construct interval measures of the distances between elements (Linacre, 2006). Simko and Pechenick (2010) compared several rank-aggregating methods for combining partially ranked data from plant breeding trials and concluded that methods based on the Bradley-Terry (Bradley and Terry, 1952) and Rasch (Rasch, 1993) models performed better than the other tested methods when factors such as fitness of aggregate rankings, time required for analyses, and the ability to analyze weak rankings were considered. In this paper we show how the Rasch model can be used to combine data from numerous sources into a single dataset while allowing for the quantification of heterogeneity among the combined data. The advantage of the approach is that the final rating is not affected by 1) the method or rating scale with which the original values were obtained, and 2) the number of trials in which each individual accession was evaluated. Since the Rasch analysis of rank-ordered data exhibits robustness against missing data, it does not require that every accession be evaluated in every trial. The final rating is based on a combination of an accession's performance in trials, and on the mean performance of other accessions included in the same trials.

In the next section we provide a brief description of the Rasch model for rank-ordered data. We also illustrate application of the method on sets of real data in four examples. These examples demonstrate how the model can be used for combining and analyzing 1) sparse data, 2) extreme scores, 3) unexpected observations, and 4) disparate tests.

MATERIALS AND METHODS

PARTIAL CREDIT MODEL

Consider a situation where only a subset of accessions is tested in several independent trials and their performance in each trial is rated on a different scale. Suppose that we wish to combine results from these trials into an overall linear rating that estimates performance of all accessions, even those that were never tested together. We can replace absolute ratings in each trial with a relative ranking of accessions. As long as there are some overlapping accessions across rankings observed in different trials, the combined final rating (and ranking) can be constructed with the Rasch model approach (Rasch, 1993). Assuming that there are up to as many score categories as accessions in each trial, and that each score category is occupied by a single accession (or multiple accessions, if tied rankings are present), then a measurement model for this conceptualization is the same as the partial credit model (Wright and Masters, 1982). This type of an unidimensional latent trait model for polytomously-scored responses has long been applied in education and psychology (Masters, 1982).

MODEL FOR RANK-ORDERED DATA

In the present work we use a modification of the partial credit model for rank-ordered data (Linacre, 1992). This new model includes comparison of multiple objects while allowing for the constrains imposed by a ranking:

$$\log (P_{nij} / P_{ni(j-1)}) = B_n - D_i - F_{ij} \tag{1}$$

where P_{nij} is the probability that accession n tested in trial i is observed in category j; $P_{ni(j-1)}$ is the probability that accession n tested in trial i is observed in category j-1; B_n is the performance measure of the accession n; D_i is the difficulty measure of trial i (the mean performance of the accessions included in trial i); and F_{ij} is the Rasch-Andrich threshold between categories j and j-1, or in other words the calibration measure of category j relative to category j-1 in trial i.

This overall rating places each accession at its performance measure on a latent variable, which is marked out in logits. The standard error, S.E.$_n$, of the measure, B_n, is:

$$S.E._n = 1 / \sqrt{\sum_{n \forall i} \left(\sum_{j=1}^{m_i} \left(j^2 P_{nij} \right) - \left(\sum_{j=1}^{m_i} j P_{nij} \right)^2 \right)} \tag{2}$$

where the accessions in trial i are ranked sequentially in categories from 1 to m_i and ties are allowed. Similarly calculated is the standard error of trial i, S.E.$_i$.

If two accessions were never directly compared to each other, they can be compared through their performance against the other accessions. Because of the overlap of accessions across trials, it is possible to create a single scale of accession performance from these rankings. The wider the reported the measure range, the more unexpected statistically are the irregular results, and the more expected are the predicted results (Linacre, 2006). Under these conditions, the highest possible precision for any measure is obtained when the data fits the Rasch model.

TEST OF HETEROGENEITY

Fitting data to the model allows estimating of an accession performance from trials containing different number of different accessions. The Rasch model approach also enables the fit statistics to be calculated for the evaluation of the consistency of performance of each accession as reflected in its ranking in trials. Fit statistics can further report the degree to which rank ordering in each trial is consistent with the estimated measures based on the overall rankings. An identified deviant ranking might be due to an accession × trial interaction, or potential experimental error. However, it is necessary to point out that the Rasch model depends on a certain level of stochastic disagreement in order to construct linear measures from ordinal observations. Perfectly ordered observations would lack information that would allow for the construction of distances between accessions (Linacre, 1989).

TEST ON REAL DATA

For demonstration purposes, we implemented the rank-order approach on real data that were obtained from analyses of processing quality, plant morphology, and response to pathogen infection. The four examples cover aspects frequently observed in biological datasets.

Sparse data: The first example deals with combining very sparse data. To illustrate the application of the Rasch model, the approach is applied to chip quality data collected from the online databases of 10 laboratories. In this example only 10% of the data-points (from the complete matrix) is available for the analysis (Supplementary table S1: http://agrobiol.sggw.waw.pl/~cbcs/articles/5_1_8/Supplementary_Table_S1.xls).

Extreme scores: In the second example we use the ear emergence date (Supplementary table S2: http://agrobiol.sggw.waw.pl/~cbcs/articles/5_1_8/Supplementary_Table_S2.xls) measured on 19 accessions of winter wheat (Giles, 1990; Piepho, 2003). This example

demonstrates how data-sets with extreme scores can be combined into a single rating. The example also allows for comparison of the results obtained from non-parametric approach based on the Rasch analysis and parametric approach based on the least square estimate.

Unexpected observations: Besides combining data, the Rasch analysis allows for the detection of unexpected observations. To demonstrate this feature of the Rasch model approach, in the third example we analyze resistance of potato accessions to late blight (Supplementary table S3: http://agrobiol.sggw.waw.pl/~cbcs/articles/5_1_8/Supplementary_Table_S3.xls). Resistance data were obtained from 29 potato accessions tested in seven locations (Haynes et al., 2002).

Disparate tests: The fourth example illustrates how the Rasch analysis can identify consistency of rank-order from fit statistics. The analysis is applied on tuber blight data obtained from testing 137 potato clones with five isolates of the pathogen (Supplementary table S4: http://agrobiol.sggw.waw.pl/~cbcs/articles/5_1_8/Supplementary_Table_S4.xls).

STATISTICAL ANALYSIS

All analyses were carried out using the Winsteps 3.65.0 computer program. For better illustration, the final measures of each example were linearly adjusted to the same scale ranging from 0 to 100, where 0 corresponds to the extreme low score and 100 corresponds to the extreme high score. The extreme score is reached when an accession is consistently rated as either the best (lowest ranked, highest measure = 100) or worst (highest ranked = 1, lowest measure = 0) in every trial where it is tested.

RESULTS

In the following part we describe details of analyses and results obtained with the Rasch model approach for rank-ordered data that was applied to combine heterogeneous datasets. The actual datasets and the final combined values can be seen at the supplementary tables S1 to S4.

EXAMPLE 1. – SPARSE DATA

Potato chip quality is an economically important trait that is regularly evaluated in breeding programs worldwide. The best chipping cultivars are those that produce light-colored, uniform chips. There are several different rating scales that are used to grade the quality of chips. For example, some laboratories arbitrarily use five-, nine-, and 10-category rating scales where lower numbers suggest chips of better quality. Other laboratories measure chip color on the Agtron scale (0 to 100 range) in which higher readings indicate a lighter, more desirable chip color, while still others express chipping quality as the percent of tested chips that passed a certain internal standard.

To illustrate the application of the Rasch model, the approach was applied to chip quality data collected from the online databases of 10 laboratories (only named cultivars were considered in the present analysis). In total, 63 cultivars were evaluated across 157 trials; however, those data were highly heterogeneous (Supplementary table S1). While only one cultivar was evaluated in 154 trials, seven cultivars were evaluated in a single trial only. The number of evaluated cultivars per trial ranged from three to 16. In total, only 994 (10%) data-points were available from the matrix that would have contained 9,891 data-points (63 × 157) if all the cultivars had been evaluated in all the trials (Table 1).

The application of the Rasch model allowed us to combine all 63 accessions into a single rating scale (Figure 1a). The model explained 74.0% of the variation in cultivar ranking. The individual accession effect explained 53.9% of the variation in the data while the effect of trials explained 20.1% of the variation in the observed rankings (Table 1). Cultivar Snowden, known for producing chips of high quality, was used as standard in 136 (87%) trials and had the third highest rating score (61.5 ± 0.6). The only two cultivars that were rated above Snowden (Beacon Chipper with 66.6 ± 7.3 and Navan with 61.9 ± 7.5) were tested in two trials and one trial only; thus they have high standard errors. Overall, the worst rated

cultivars were Molli (30.5 ± 4.9), Stampede Russet (30.2 ± 5.6), Fabula (19.3 ± 9.9), and Milva (8.4 ± 10.1). These four cultivars were always rated the worst — or tied for the worst — in the respective chipping trials.

Table 1. Description of the four examples that were used in the demonstration of combining heterogeneous datasets with the rank-order approach.

Example	Number of accessions	Number of trials	Data-points in complete matrix	Available data-points	Trials per accession			Accessions per trial			Explained variation (%)			Unex-plained variation (%)
					Min.	Mean	Max.	Min.	Mean	Max.	Total	Acce-ssions	Trials	
No. 1	63	157	9,891	994	1	15.8	154	3	6.3	16	74.0	53.9	20.1	26.0
No. 2	19	12	228	73	3	3.8	5	4	6.1	9	91.0	72.8	18.2	9.0
No. 3	29	7	203	196	6	6.8	7	22	28.0	29	84.8	83.9	1.0	15.2
No. 4	137	5	685	543	1	4.0	5	87	108.6	121	65.7	16.6	49.2	34.3

EXAMPLE 2. – EXTREME SCORES

Another example illustrates how Rasch analysis deals with extreme scores. An accession reaches the extreme score if it achieves either the lowest or the highest ranking across all trials in which it is evaluated. Since the final measure that corresponds to an extreme score is not estimable by direct application of a Rasch model, extreme scores are typically adjusted to (slightly, but reasonably) non-extreme scores during the estimation process. Wright (1998) describes various approaches of data adjustment. These methods usually involve subtracting a fractional score point value from the maximum (or adding it to the minimum) score. Common extreme score corrections are in the range from 0.167 to 0.5. Winsteps corrects extreme scores by 0.3 score points by default. Another approach involves data augmentation with non-extreme responses that have minimal impact on the measurement system. Once the set of measures is estimated, the augmented data can be dropped, and standard error and fit statistics are computed from the observed data.

Ear emergence date were observed on 19 accessions of winter wheat in 12 independent trials (Giles, 1990; Piepho, 2003) (Supplementary table S2). The number of accessions per trial ranged from four to nine, and each accession was tested in three to five different trials. In total, 73 data points (32%) were available from the matrix that would contain 228 data points (19 × 12) if all accessions were evaluated in all trials (Table 1). Two accessions achieved the minimum possible score. One of those accessions was tested in three trials (Ac-277), while the other one was tested in five trials (Ac-2098). Because of the occurrence of extreme scores, augmentation was performed prior to data analysis (Wright, 1998).

The final rating for 19 accessions ranged from 3.7 ± 8.4 (Ac-277) to 88.0 ± 4.8 (Ac-988) (Figure 1b). The model explained 91.0% of the total variation, of which 72.8% was explained by the effect of accessions and 18.2% by the effect of trials (Table 1). Since ear emergence was measured in the same way in all trials, it was possible to calculate the composite rating also with a parametric approach. The rating based on a least square estimate (Piepho, 2003) significantly correlated (r = 0.94, P < 0.0001) with the rating based on the Rasch model. The most important difference between the two methods was in rating of the two accessions with extreme score. When the ranking approach was applied, the lowest rated accession was Ac-277, preceded by Ac-2098. The final order of the two accessions was switched when analysis was performed with the parametric approach. When Ac-2098 was excluded from the data analysis, the correlation between the two methods slightly increased to r = 0.97.

EXAMPLE 3. – UNEXPECTED OBSERVATIONS

In this example, we applied the Rasch model to the data for potato resistance to late blight obtained from trials performed across the USA (Haynes et al. 2002). Potato late blight, caused by *Phytophthora infestans*, is one of the most devastating diseases of potato worldwide. Twenty-nine potato accessions, consisting of 12 breeding lines, 10 cultivars, and seven late blight differentials, were tested for resistance in seven locations (Florida, Maine, Michigan, Minnesota, New York, North Dakota, Pennsylvania). Plots in Florida and Maine depended on natural infection while plots in the other locations were inoculated with *P. infestans*. Reaction to infection was assessed as mean AUDPC (Area Under the Disease Progress Curve) for each accession at each location. A near-complete (97%) data matrix was available for analysis, because all accessions were tested in all locations except for seven differentials that were not tested in Florida (Supplementary table S3, and Table 1). Though the AUDPC score was used to assess the resistance in each location, the non-parametric analysis of ranked data was recommended for data analysis due to a high heterogeneity of resistance scores (Haynes et al., 2002).

The Rasch model explained 84.8% of the variation in resistance ranking. Almost all of the variation (83.8%) was explained by the effect of individual accessions, and only 1.0% was explained by individual trials (Table 1). Once the data were fit to the Rasch model, Winsteps software was used to calculate the expected rank for each accession at each location. The observed and expected rankings were subsequently used to identify the most unexpected ranks. The unexpectedness of a rank was calculated as the difference between the observed and expected rank, divided by the square root of the variance. Our analysis indicates that the most unexpected rankings were rank 7 for line B1004-8 in Maine (expected 2.15), rank 24 for differential R_{multi} in Maine (expected 27.65), and rank 12 for line B0811-13 in Florida (expected 4.78). These results are not surprising since only these two locations relied on natural infection with the pathogen. Moreover, testing carried out in Maine showed limited late blight infection due to lack of inoculation, a lack of favorable environment conditions for *P. infestans*, or both (Haynes et al., 2002). Therefore, the unexpected results might either indicate a possible genotype × environment interaction or errors. More tests would be needed in these two locations to determine if the rankings are consistent. The final resistance rating for 29 accessions ranged from 22.4 ± 2.8 (B1004-8 – most susceptible) to 72.8 ± 3.1 (R_{multi} – most resistant) (Figure 1c).

EXAMPLE 4. – DISPARATE TESTS

Data for this example come from observations of late blight resistance on potato tubers (unpublished results from the W. E. Fry's laboratory at Cornell University). Potato tubers from 137 clones were inoculated in separate tests with up to five isolates of late blight and percent of tubers developing symptoms were recorded. Sixty-six clones were inoculated with all five isolates, 26 clones were inoculated with four isolates, 27 clones were inoculated with three isolates, 10 clones were inoculated with two isolates, and 8 clones were inoculated with only a single isolate (Supplementary table S4). The data distribution in all trials was skewed toward resistance. These data also contain many tied rankings, especially among the most resistant clones. In total, 543 data-points (79%) were available from the matrix that would contain 685 data-points (137 × 5) if all clones had been tested with all isolates (Table 1).

The final rating for 137 clones ranged from 20.9 ± 36.0 for the clone number CLN-061 (most susceptible) to 81.3 ± 10.7 for the clone number CLN-083 (most resistant) (Figure 1d). An unusually large standard error was estimated for CLN-061, because this clone was tested only once and happened to be the most susceptible material in the trial. The Rasch model explained 65.7% of the variation in ranking of the clones. However, only 16.6% of the variance was explained by the effect of individual clones, while 49.2% were explained by the effect of isolates (Table 1). The principal components/contrast analysis of model standardized residuals shows large contrast loading (> 0.99) for isolate #4 (Figure 2),

indicating that clones react differently to this isolate than to the other four isolates. Testing of the isolates on R-gene differentials revealed that four isolates (#1, #2, #3, and #5) are members of the US-8 clonal lineage, while the isolate #4 belongs to a Mexican lineage with a different set of avirulence genes (Rauscher et al., 2006).
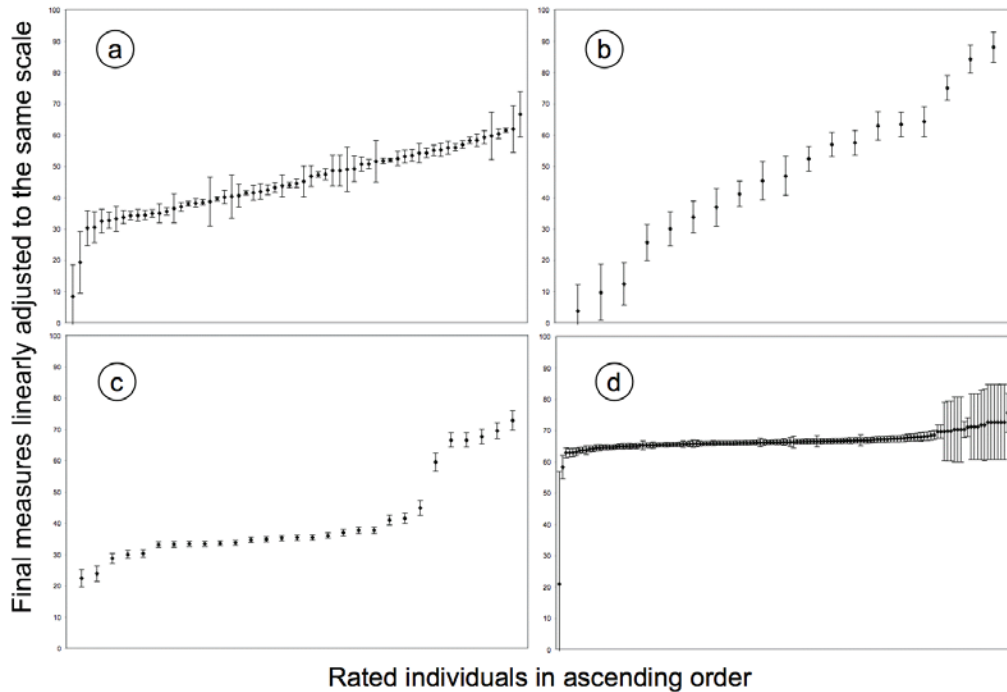

Figure 1. Final measure and corresponding standard error for individuals from the examples 1 (a), 2 (b), 3 (c), and 4 (d). The values were obtained with the rank-order approach analysis that is based on the Rasch model. The original rating values and the final measures for all individuals are available in the Supplementary tables S1, S2, S3, and S4. Black diamond indicates a final measure in ascending order and vertical lines show standard error of the measure.
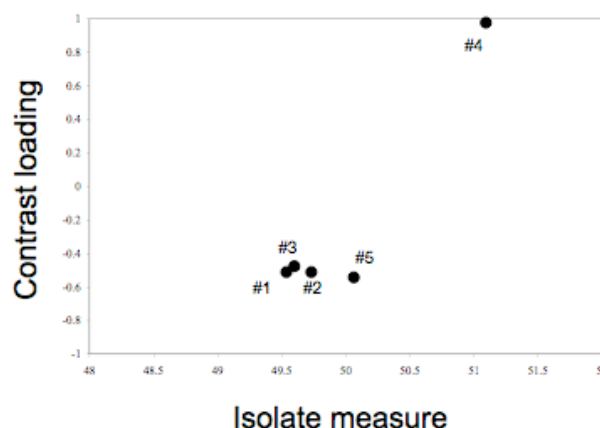

Figure 2. Plot of standardized residual contrasts for the five *Phytophthora infestans* isolates tested on 137 potato clones (example 4). The principal components/contrast analysis shows the largest contrast loading for the isolate #4, indicating that tested clones react differently to this isolate than to the other four.

In a view of the different makeup of avirulence genes in the isolate #4, resistance data from this isolate should not be combined into an overall rating with data obtained from the other four isolates (we combined them for illustration purposes only). This example shows how the Rasch model can detect data originating from different sub-groups and separate those sub-groups. When isolate #4 was excluded from the analysis, the Rasch model explained less (49.7%) of the observed variance. However, while variation explained by isolates decreased 5-fold to 9.9%, variation explained by individual clones increased 2.4-fold to 39.8% (data not shown).

DISCUSSION

The Rasch analysis focuses on constructing linear measures from ordinal data (rank order, rating scales). Successful application of the model requires that the ordering in the data correspond to qualitative advances along the intended latent variable. This requirement appears to be successfully approximated in biological datasets. For empirical demonstration, the Rasch model was applied to combine data from several studies as shown in the four examples.

The rank-order method worked well for integrating data across studies even when 90% of data were missing from the complete matrix (example 1). The ability to combine datasets with only a partial overlap is an advantage of the Rasch model approach. Due to the robustness of the model, combining sparse data is usually not a problem, and even datasets with 99% of missing data (relative to the complete matrix) can be successfully merged into a single dataset (Linacre, 2010). Also, mixing different types of ordered data (with or without ties) is usually straightforward and trouble-free (example 4).

The biggest challenge in combining biological datasets with the Rasch model appears to be the presence of extreme scores (Simko and Pechenick, 2010). However, augmentation of data with non-extreme responses that have minimal impact on the measurement system (Wright, 1998) worked well, and the approach provided a reasonable estimate of the final ratings (examples 2 and 4). Nevertheless, ratings for extreme scores are less inferentially secure than ratings for non-extreme scores (Linacre, 2006). Besides allowing for the combination of data from dissimilar studies, the advantage of the Rasch model is the possibility of quantifying heterogeneity across the combined data. In example 3, statistical analysis based on the Rasch model revealed the most unexpected data that may indicate genotype × environment interaction (or potential experimental error), while in example 4 the isolate with a different set of avirulence genes was detected correctly.

Previously, other methods were suggested for combining sparse and unbalanced data. Piepho (2003) compared two parametric methods for combining heterogeneous data such as evaluations performed on germplasm collections. The method of Giles (1990) removed bias by fitting a regression line to the bivariate scatter plot by principal component analysis. Though this method allowed calculating an adjusted mean, the resulting means were not invariant due to the arbitrary order of alignments. To avoid this type of ambiguity, Piepho (2003) advocates the use of an alternative method based on a least square estimate approach. When the Rasch model approach was applied to the same data (example 2), the final ratings significantly correlated (r = 0.94, P < 0.0001) with the estimates obtained by the least square approach (Piepho, 2003). The high correlation indicates that the non-parametric rank-order based method yields a rating very similar to that from the parametric test based on absolute values. However, neither the principal component analysis nor the least square approach can be used to combine data when the rating scale across trials is not identical. To deal with different rating scales, Hartung and Piepho (2005) proposed a threshold model. This model may be used to derive mean scores on any one of the rating scales when data from a long-term series of evaluations are available.

Simko and Pechenick (2010) compared performances of several rank-aggregating methods for combining data from plant breeding trials. Tests carried out on simulated and real data indicated that the best performing methods were based on Rasch and Bradley-Terry models, which are related to Thurstone's model (Thurstone, 1927). In the Bradley-Terry model (Bradley and Terry, 1952), the rankings from each trial are decomposed into (n × (n – 1) / 2) paired comparisons. For example, the rank-ordering of 10 accessions would yield 45 paired comparisons (10 × 9 / 2). The results obtained by the paired comparison test are virtually identical to those from the rank-order test (Bramley, 2005; Linacre, 2006). The only difference in the final rating is that the paired comparison approach appears to create a longer scale, especially at the extremes of the scaling range (Linacre, 2006).

Since the rank-order approach uses only information about the order of the data but discards the actual values, this non-parametric procedure can never be as powerful (able to detect existing differences) as its parametric counterparts when parametric tests can be used (Whitley and Ball, 2002). However, the advantage of non-parametric measures is that they are distribution-free and, therefore, are less sensitive to errors of measurements than the parametric statistics. Our work demonstrates that the Rasch model-based rank-order approach is suitable for combining phenotypic data from diverse datasets and dissimilar rating scales. We propose using this method in a data mining process to produce an aggregate rating from available databases and other sources that hold phenotypic information on biological material. A relative rating of phenotypes can be subsequently used together with molecular marker information in appropriate gene-mapping analyses (Simko, 2004; Simko et al., 2009; Simko and Pechenick, 2010) or for selecting desirable material.

Although the focus in our paper has been on use of the Rasch model for breeding and genetics, it should find application in many other areas of biology, including agriculture. Examples would include evaluation of herbicides, fertilizers, and other agricultural practices for efficacy.

REFERENCES

Bradley, R.A., Terry, M.E. (1952). Rank analysis of incomplete block designs: The method of paired comparisons. *Biometrika* 39, 324–345.

Bramley, T. (2005). A rank-ordering method for equating tests by expert judgment. *Journal of Applied Measurement* 6, 202–223.

Giles, R. (1990). Utilization of unreplicated observations of agronomic characters in a wheat germplasm collection. In: Srivastava, J.P., Damania, A.B. (Eds.) *Wheat Genetic Resources: Meeting Diverse Needs*, John Wiley & Sons, Chichester, UK. 113–130.

Hartung, K., Piepho, H.P. (2005). A threshold model for multiyear genebank data based on different rating scales. *Crop Science* 45, 1045–1051.

Haynes, K.G., Christ, B.J., Weingartner, D.P., Douches, D.S., Thill, C.A., Secor, G., Fry, W.E., Lambert, D.H. (2002). Foliar resistance to late blight in potato clones evaluated in national trials in 1997. *American Journal of Potato Research* 79, 451–457.

Linacre, J.M. (1989). Rank ordering and Rasch measurement. *Rasch Measurement Transactions* 2, 41–42.

Linacre, J.M. (1992). Objective measurement of rank-ordered objects. In: Wilson, M. (Ed.) *Objective Measurement: Theory into Practice*. Ablex, Norwood, NJ. 195–209.

Linacre, J.M. (2006). Rasch analysis of rank-ordered data. *Journal of Applied Measurement* 7, 129–139.

Linacre, J.M. (2010). Predicting responses from Rasch measures. *Journal of Applied Measurement* 11, 1–10.

Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika* 47, 149–174.

Piepho, H.P. (2003). Model-based mean adjustment in quantitative germplasm evaluation data. *Genetic Resources and Crop Evolution* 50, 281–290.

Rasch, G. (1993). *Probabilistic models for some intelligence and attainment tests*. MESA Press, Chicago, IL.

Rauscher, G.M., Smart, C.D., Simko, I., Bonierbale, M., Mayton, H., Greenland, A., Fry, W.E. (2006). Characterization and mapping of R_{Pi-ber}, a novel potato late blight resistance gene from *Solanum berthaultii*. *Theoretical and Applied Genetics* 112, 674–687.

Simko, I. (2004). One potato, two potato: Haplotype association mapping in autotetraploids. *Trends in Plant Science* 9, 441–448.

Simko, I., Costanzo, S., Haynes, K.G., Christ, B.J., Jones, R.W. (2004a). Linkage disequilibrium mapping of a *Verticillium dahliae* resistance quantitative trait locus in tetraploid potato (*Solanum tuberosum*) through a candidate gene approach. *Theoretical and Applied Genetics* 108, 217–224.

Simko, I., Haynes, K.G., Ewing, E.E., Costanzo, S., Christ, B.J., Jones, R.W. (2004b). Mapping genes for resistance to *Verticillium albo-atrum* in tetraploid and diploid potato populations using haplotype association tests and genetic linkage analysis. *Molecular Genetics and Genomics* 271, 522–531.

Simko, I., Hu, J. (2008). Population structure in cultivated lettuce and its impact on association mapping. *Journal of the American Society for Horticultural Science* 133, 61–68.

Simko, I., Pechenick, D.A. (2010). Combining partially ranked data in plant breeding and biology: I. Rank aggregating methods. *Communications in Biometry and Crop Science* 5, 41-55.

Simko, I., Pechenick, D.A., McHale, L.K., Truco, M.J., Ochoa, O.E., Michelmore, R.W., Scheffler, B.E. (2009). Association mapping and marker-assisted selection of the lettuce dieback resistance gene *Tvr1*. *BMC Plant Biology* 9, 135.

Thurstone, L.L. (1927). A law of comparative judgment. *Psychological Review* 34, 273–286.

Whitley, E., Ball, J. (2002). Statistics review 6: Nonparametric methods. *Critical Care* 6, 509–513.

Wright, B.D. (1998). Estimating measures for extreme scores. *Rasch Measurement Transactions* 12, 632–633.

Wright, B.D., Masters, G.N. (1982). *Rating scale analysis*. MESA Press, Chicago, IL.