JOURNAL OF CHEMOMETRICS
J. Chemometrics 2000; **14**: 379–380

Software Review

N-WAY TOOLBOX FOR MATLAB by Claus A. Andersson, and Rasmus Bro, Royal Veterinary and Agricultural University, Department of Dairy and Food Science–Food Technology, Rolighedsvej 30, DK-1958 Frederiksberg C, Denmark. Free MATLAB® toolbox available on the net at http://www.models.kvl.dk/source/nwaytoolbox/.

The *N*-way toolbox for MATLAB is a set of MATLAB routines designed to perform multiway data analysis. This collection of freely available programs can be considered pioneering work in the area and efficiently covers the lack of a specific software compilation for the application of multiway analysis methods to chemical data sets. Although these mathematical tools have been fairly well explored and applied by psychometricians, this has not been the case in chemistry, where the performance and adaptation of each model and method to the variety of chemical examples are still controversial. The authors should then be congratulated for offering a useful tool related to an evolving area, in which the fundamentals are not as well established as in other chemometric fields.

The exhaustive variety of models and algorithms covered by the *N*-way toolbox includes parallel factor analysis (PARAFAC) [1,2], *N*-mode partial least squares regression (*N*-PLS) [3], generalized rank annihilation method (GRAM) [4], direct trilinear decomposition (DTD) [5] and the family of TUCKER models [6]. These methods allow the chemist to carry out exploratory analysis, calibration and resolution of multiway arrays. The eventual transformation of the raw data sets has been foreseen and several options of multiway data preprocessing, such as centering and scaling, are incorporated. All the algorithms are also prepared to handle data sets with missing values.

The *N*-way toolbox can be downloaded from the net at http://www.models.kvl.dk/source/nwaytoolbox/. The routines in the toolbox are compatible with both MATLAB v. 5.x and MATLAB v. 4.2.c. The necessary information about installation and how to get started with the toolbox is found in a 'readme' file. A list of the tasks performed by each source file (*m*-file) is obtained by typing the 'help contents' command in the MATLAB environment. Every *m*-file has also a 'help' message about its own input and output. This information is enough for experienced MATLAB users familiarized with three-way analysis methods. Those who do not belong to this group should not worry, because two interactive courses are also freely available at http://www.models.kvl.dk/courses/. This didactic material describes in a clear and summarized way most of the methods used and provides simulated and real data sets representative of different chemical problems to play with. In using these data sets as proposed, these courses also become a good 'hands-on' tutorial for the *N*-way toolbox. Given the often large size of multidimensional arrays, the toolbox improves its performance when used in computers with moderate to high numerical processing capability. Despite the PC at hand, users should be aware of the intrinsic slowness associated with the application of the iterative methods in the toolbox.

An attractive feature in the implementation and use of the multiway methods in the toolbox is the large choice of options to run the different programs when the characteristics of the related method allow for this freedom, as is the case in the PARAFAC and TUCKER models. This flexibility fits the diversity of multiway chemical arrays and reflects the desire of the authors to offer software that can adapt as much as possible to different situations and user criteria. The parameters that may be chosen start with the initialization of the methods, which can use SVD-based profiles, DTD results, random orthogonalized profiles, the best of several models obtained using few iterations, or any other initial estimate input by the user. A key point in the use of these models for resolution purposes is the optional selection of constraints. Both PARAFAC and TUCKER routines can constrain the profiles related to each mode in the data array in different ways. Both the PARAFAC and the TUCKER routines offer non-negativity, unimodality and orthogonality as possible constraints. Related to the structure of the TUCKER model, the associated routine allows the constraint of the core matrix. To this end, an external core matrix can be input and fixed during the iterative process, or an array sized as the core can be input to define the relationship among the profiles in the different modes, i.e. which elements in the core should be equal to zero and which should not. When this information is not available, the TUCKER model provided is rotated to maximize the simplicity of the core and hence the interpretation of the final results. Known modes can be fixed during optimization in both TUCKER and PARAFAC programs. Typical parameters related to iterative optimization methods, such as the convergence criterion or the number of iterations allowed, can also be selected. In any case, users who are more for the 'black box' philosophy, or simply those who do not have a clear criterion to choose from among the parameters mentioned above, can run the algorithms with the default options set by the authors.

Despite the wide applicability and range of options of this toolbox, this software should not be considered a finished product. Indeed, the authors continuously offer new and updated versions. Far from being a negative feature, the dynamic nature of this toolbox responds to the continuous evolution of multiway chemometrics and ensures that the

product will not be out of date in a couple of years. The evolution of this toolbox is mainly due to the experience of the authors and to their interest in receiving feedback from users, who can send doubts and suggestions through the net. In this sense the 'on-line' service to answer queries about both the programs and the multiway methods can be qualified as fast and efficient.

As an additional user, I list a few personal suggestions and wishes that the authors may perhaps consider in the future. These comments are mostly focused on the inclusion of additional constraints for resolution purposes, the incorporation of additional algorithms in the toolbox, and the input/output format.

Despite the evident concern to include constraints in the iterative algorithms, there is no way to input explicitly the presence of selectivity in the data set, which is essential for resolution purposes. In the current version it is possible to keep fixed whole mode-related matrices. Freedom to fix the profiles in the modes completely or partially could help in this point. Selectivity could then be implemented by fixing some elements in the profiles, e.g. in selective regions the elements in profiles of absent species could be set to be lower than a very small threshold value close to zero. This would also allow the inclusion of some invariant profiles in a mode, e.g. spectra of identified compounds.

The extensive work of the authors in multiway analysis is not completely presented in the toolbox. The integration of recently implemented algorithms in the toolbox, such as the PARAFAC2 model [7,8], which can be downloaded from the net separately, would be welcome.

In general, the default options set by the authors in the algorithms are clearly presented and users can easily decide if they are suitable for their own data set. This point could perhaps be improved in the selection of a convergence criterion for iterative algorithms. In the current version there is an absolute threshold value (10^{-6}) for the difference in fit between consecutive iterations. Since the scales in all data sets are different, it is hard to find a default value, and the inclusion of this option in relative terms may be more advisable. Besides, average users are more likely to know whether they allow a 0·1% fit difference between consecutive iterations rather than knowing which absolute number corresponds to this percentage. The same could be applied to the fit parameter; though the square sum of the residuals is an informative parameter and should be present, so is the percentage explained variance associated with this numerical value. This relative parameter, which appears sometimes on the screen, could be defined as an additional storable output argument.

Still on the input/output format, the authors have opted for vectorized arrays in some parameters, which are by nature data matrices. This holds for input parameters, e.g. matrices of external estimates, and for the scores and loadings given by the PARAFAC, TUCKER and *N*-PLS related programs. Replacing a data matrix by a long array of appended profiles does not seem the most intuitive way to handle input parameters or to interpret an output. Even though a routine called 'fac2let' is provided to reshape output arrays, the average user would appreciate finding a more direct correspondence between the mathematical entities present in the theoretical models and the information needed or provided by the related programs. The task of unfolding the input matrices and folding the output vectors could be incorporated in the programs that may need it.

Although this review should be limited to the *N*-way toolbox, it is worth mentioning again the quality of the interactive courses related to it. With them, users may learn about the theory and practice of multiway analysis and assess their degree of understanding with the many test proposals in this material. A last wish would be to find a closing chapter with guidelines on which method is the most suitable for each purpose. Though this is still a matter of open debate, many members of the multiway community would appreciate the opinion of these experienced authors on this point.

In summary, the *N*-way toolbox and related material should be considered a very versatile and user-friendly tool for multiway analysis. This software compilation will surely encourage many non-'*multiwayers*' to get started fearlessly in this area and to find many applications similar to their everyday chemical problems. For those who are already in the field, the *N*-way toolbox is an extensive and practical compilation of algorithms that will certainly complement their habitual software.

REFERENCES

1. Harshman RA. *UCLA Working Papers Phonet*. 1970; **16**: 1–84.
2. Carroll JD, Chang J. *Psychometrika* 1970; **35**: 283–319.
3. Bro R. *J. Chemometrics* 1996; **10**: 47–61.
4. Sánchez E, Kowalski BR. *Anal. Chem.* 1986; **58**: 496–499.
5. Sánchez E, Kowalski BR. *J. Chemometrics* 1990; **4**: 29–45.
6. Tucker LR. The extension of factor analysis to three-dimensional matrices. In *Contributions to Mathematical Psychology*, Frederiksen N, Gulliksen H (eds). Holt, Rinehart Winston: New York, 1964; 110–182.
7. Kiers HAL, ten Berge J, Bro R. *J. Chemometrics* 1999; **13**: 275–294.
8. Bro R, Andersson CA, Kiers HAL. *J. Chemometrics* 1999; **13**: 295–309.

ANNA DE JUAN
Chemometrics Group
Dept. de Química Analítica
Universitat de Barcelona
E-08028 Barcelona, Spain

