My Multiway Analysis: From Jan de Leeuw to TWPack and Back

Pieter M. Kroonenberg
Leiden University & The Three-Mode Company

Abstract

This paper is a mixture of my personal experiences of Jan de Leeuw as a supervisor of my master's and Ph.D. theses, as well as a sketch of how three-way analysis, the subject Jan chose for me, developed over time. The emphasis is on where it is and was applied, and to what extent it stole the hearts of applied researchers in different disciplines. Furthermore, the paper contains some musings about how we should go about promoting the use of the techniques, especially in the social and behavioural sciences. Finally, an overview is provided of available software and attention is paid to how (three-way) software may be designed to encourage its use by the scientific community, as it befits a paper in the *Journal of Statistical Software*.

Keywords: Jan de Leeuw, three-way analysis, three-mode component analysis, applications, software, Fortran, Delphi.

1. The Lion and I

One of the recurrent plots in science fiction is the "What if ..." question. What would have happened to the universe if there had no been no Starship Enterprise? This is the type of speculation that hits me when I try to contemplate what would have happened to me if there had been no Jan de Leeuw, or John Lion if you prefer. (Hear how to properly pronounce his last name on his website at De Leeuw 2016c.) He certainly kick-started my scientific career, and rather than working on three-way methods I might have ended up writing a thesis on "Exact tests in contingency tables", a noble subject but not as fascinating as my current work. Three-way analysis has brought me into contact with many researchers and many disciplines, such as psychology, education, nursing, medicinal pharmacy, cancer research, and plant breeding concerned with crops such as maize, soya, wheat, adzuki beans, and peanuts, and, of course, the environmental state of affairs on the Great Barrier Reef.

Figure 1: Committee for the Ph.D. defense of Pieter Kroonenberg.

Jan was and is a charismatic figure who in ancient times would have been the stuff of legends. A probably apocryphal story goes that one evening after a prolific drinking session he was handed the 460-page book by Warren Torgerson on *Theory and Methods of Scaling* (Torgerson 1958) with the message that this would probably be something worthwhile for him to read. And sure enough, the next morning after apparently having finished the book (when exactly is still a mystery) he summarised the author's achievements and indicated places where interesting extensions of the procedures described in the book could be added.

Jan is a somewhat unconventional person who does not like formalities much ("undisciplined and wild" as he says in his own CV De Leeuw 2016b). As a full professor in Leiden he seemed to have appeared once at a Ph.D. defense with a red sweater under his gown, while a dark suit with a white shirt and black tie are compulsory; see Figure 1 for an appropriately dressed Jan de Leeuw during my Ph.D. ceremony.

Jan's influence on the world of psychometrics and beyond has been profound, as can be deduced from the papers, published and unpublished, which he wrote on all aspects of what was then already called *Data Theory*. The contributions in this volume and his bibliography (De Leeuw 2016a) are ample proof of this. He joined his thesis supervisor John P. van de Geer at the Leiden Department of Data Theory. This department became a shining research light in the Faculty of Social and Behavioural Sciences at Leiden University. Jan left Leiden in 1987 for the University of California, Los Angeles, but continued to supervise Dutch Ph.D. students. By 1990 there were fourteen of them, of whom eight went on to become full professors themselves.

As a student of applied mathematics I wrote my master's thesis under his direction. His direction consisted of giving me a few sheets of orange millimeter graph paper containing a series of formulas outlining the TUCKALS2 algorithm (downloadable as De Leeuw 1975). He added that I should read Tucker's 1966 paper, program the algorithm given on his sheets, apply it to an example, and this would lead to a Master's degree. I came to him because as his student assistant I had been programming a precursor of the present Homals algorithm; a program dumped as soon as it was ready, because Jan had since then devised a better algorithm. In addition, one of the professors at the Department of Mathematics had indicated that I could not seriously expect him to pull a subject out of the bag for every student who had the audacious idea of *doing* mathematics rather than *studying* it. I set to work on Jan's algorithm and 3000 IBM punch cards later the computer program was ready. I wrote the thesis on an electric typewriter with four "golf balls" that you could swap to get different fonts. Mistakes were covered up with correction tape; see also Jan's similar experiences on the CV page of his website (De Leeuw 2016b).

The research bulletin based on my master's thesis (Kroonenberg and De Leeuw 1977) was transformed into my first paper (Kroonenberg and De Leeuw 1978), which however was in Dutch. However, a couple of years later Jan wanted to know when the 'real' article for an English-language journal was forthcoming. He had even heard that there was some competition. And had he mentioned that I should really expand the three-way model from my master's thesis with a second, more well-known variant (now known as the Tucker3 model)? 'Otherwise it will not do'. So I did some more programming, this time on a real Teletype typewriter linked to an IBM mainframe, so there was no longer a need for punched cards and trips up and down to the Computer Centre. In 1979 I wrote the article that went with it, typed it up in IBM's **DisplayWrite3** on a word processor the size of a bookcase, showed it to Jan as the obvious co-author, photocopied it six times, and sent it by post to *Psychometrika*.

A few months later I received to my surprise a letter from the journal (no email yet at the time) containing two articles about multivariate analysis of three-way data that had been submitted to the journal. The editor wanted to know what I thought about those articles in relation to my own. I went to see Jan, somewhat shell-shocked: "What am I supposed to do with this?"; "Well, answer the questions of course," he said, "and tell them what you think". This is what they call "reviewing an article", but it was the first I'd heard of it. No matter, I rolled up my sleeves and got the job done.

A little after that I received a message to say our article had been accepted, conditional on certain additions and revisions. I set to work once more, but took my time. Until one day, now 35 years ago, an actual telegram arrived from the journal: "Where is your final version? We want to include your article in our March issue together with one of the others that you reviewed". I quickly finished my paper and sent it off. I am proud to say that now, early in 2016, this article (Kroonenberg and De Leeuw 1980) is number four on Jan's list of most cited papers.

Jan continued to play an important part in my academic life. Leiden University came to insist that all its staff should have a Ph.D. degree. Because I intended to continue with three-mode analysis as a research project, the obvious thing to do was to go back to Jan and ask him to be my supervisor, even though he was not even four years my senior. Jan, laid-back as always, approved, and let me draw up my own plan except that he urged me to include a chapter on a subject that I was not familiar with. I did not really want to include this chapter, but out came the graph paper again and Jan outlined the mathematical and algorithmic content

for the chapter. Apparently Jan had sufficient faith in me, so that we managed go through the 400-page thesis in only a few sessions. The thesis itself was published by the now-defunct DSWO Press (Kroonenberg 1983).

My next academic encounter with Jan was later in 1983 when he received a request to contribute a paper to the first multi-mode data analysis book (Law, Snyder Jr., Hattie, and McDonald 1984). He put this request on my desk and said that this was typically something for me to do (Kroonenberg 1984). In 2005 he also asked me to contribute to the *Encyclopedia of Statistics in Behavioral Science* (Kroonenberg 2005).

2. The multiway world

Jan set me on the path of my life-long fascination with three-way, and later multiway data analysis. In this section I will present a brief introduction into three-way and multiway data analysis and provide a condensed overview what happened in that field over the years[1].

2.1. A brief introduction to multiway component analysis

Even though multiway is more than three-way, in this paper I will consider the terms three-way and multiway component analysis as largely synonymous, because the fundamental step is actually from two-way to three-way analysis. The generalisation from three-way to multiway does not introduce any new mathematical or statistical intricacies, but primarily increases the notational and interpretational complexity. Thus, in this section I will concentrate on: "What is three-way component analysis?"

Before going into this, a word about the terms *way* and *mode*, both employed in this context. The first term, *way*, indicates the number of dimensions of a data box: one-way refers to a vector, two-way to a matrix, and three-way to an array: data that fit in a box. The second term, *mode*, refers to the content of a way. Within the social and behavioural sciences a standard two-way data matrix has rows (subjects) and columns (variables), and many three-way data boxes have rows (subjects), columns (variables), and tubes (conditions or time points). Hence, such data are three-way three-mode data. However, a set of correlation matrices is three-way but only two-mode, as the rows and columns both consist of the same entities, mostly variables. Thus, the word 'way' is more general than the word 'mode' and in what follows I will mainly use that.

Two-way principal component analysis is based on the singular value decomposition: a subject space spanned by the subject components, and a variable space spanned by the variable components plus the singular values (the square roots of the eigenvalues). Each subject component is uniquely linked to a variable component, so that they can occupy the same space and the size of the singular values indicates the strengths of the links between the two types of components.

Generalising this idea to the analysis of three-way data, we have to add a condition space spanned by the condition components. Again we can define links, but now for three sets of components; such links are referred to as *three-way singular values* or as elements of the core

---

[1]This and some of the next sections are partly based on material from a paper published in Blasius and Greenacre's book *Visualization and Verbalization of Data* (Kroonenberg 2014a) and from my farewell lecture (Kroonenberg 2014b).
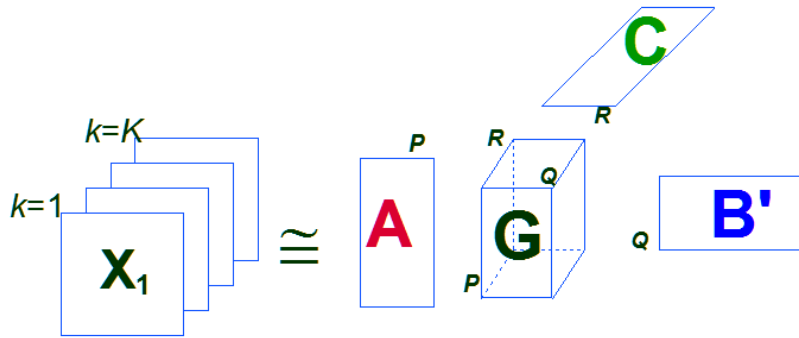
Figure 2: The Tucker3 model for the three-way data array $\underline{\mathbf{X}}$. $\mathbf{A}$ = Subject space with P components; $\mathbf{B}$ = Variable space with Q components; $\mathbf{C}$ = Condition space with R components; $\underline{\mathbf{G}}$ = Core array with the $P \times Q \times R$ links between the components from the three ways.

array. The three-way component model was introduced by Tucker (1963; 1966; 1972), who also coined the word *core* to indicate that the links in the core array play a fundamental role in modelling the relationships between the components of the three ways (see Figure 2).

The other major three-way model is the parallel factors (PARAFAC) model (Harshman 1970; Harshman and Lundy 1984b,a, 1994). This model is also referred to as the CP model, an abbreviation of the CANDECOMP-PARAFAC model recognising Carroll and Chang's 1970 work on what they called the canonical decomposition (CANDECOMP) model. Later, another meaning was given to the same CP abbreviation, namely canonical polyadic decomposition. The latter designation goes back to Hitchcock (Hitchcock 1927a,b) and was resurrected by researchers such as De Lathauwer, Comon, and Lek-Heng. In Figure 2 $P = Q = R$ are equal for the PARAFAC model, so that its core array is a cube which in addition is superdiagonal with only P elements.

The major differences between two-way and three-way models centre around the properties of uniqueness and complete decomposability. In the two-way component model the components can be transformed without losing fit of the model to the data. This transformational property is used extensively to enhance and facilitate interpretation. At the same time, the components are not unique, which hampers the theoretical identification of the solution.

As we have seen in three-way component analysis, there are several component models which can claim to be the generalisation of the two-way component model and the singular value decomposition. A very limited number of them have the uniqueness property, i.e., the solution is mathematically identified and cannot be transformed without losing fit of the model to the data. The downside of this is that such models, primarily the PARAFAC model and its descendants (Harshman and Lundy 1984b), cannot fit every single three-way data set and a complete decomposition of the data into its components is not always possible. In contrast, the models proposed by Tucker, such as the Tucker3 model, are not unique but can fully decompose any three-way data set. It is especially the uniqueness property of the PARAFAC models which has been of great value in analytical chemistry, signal processing, and other exact sciences which have physical three-way models describing the patterns underlying their three-way data.
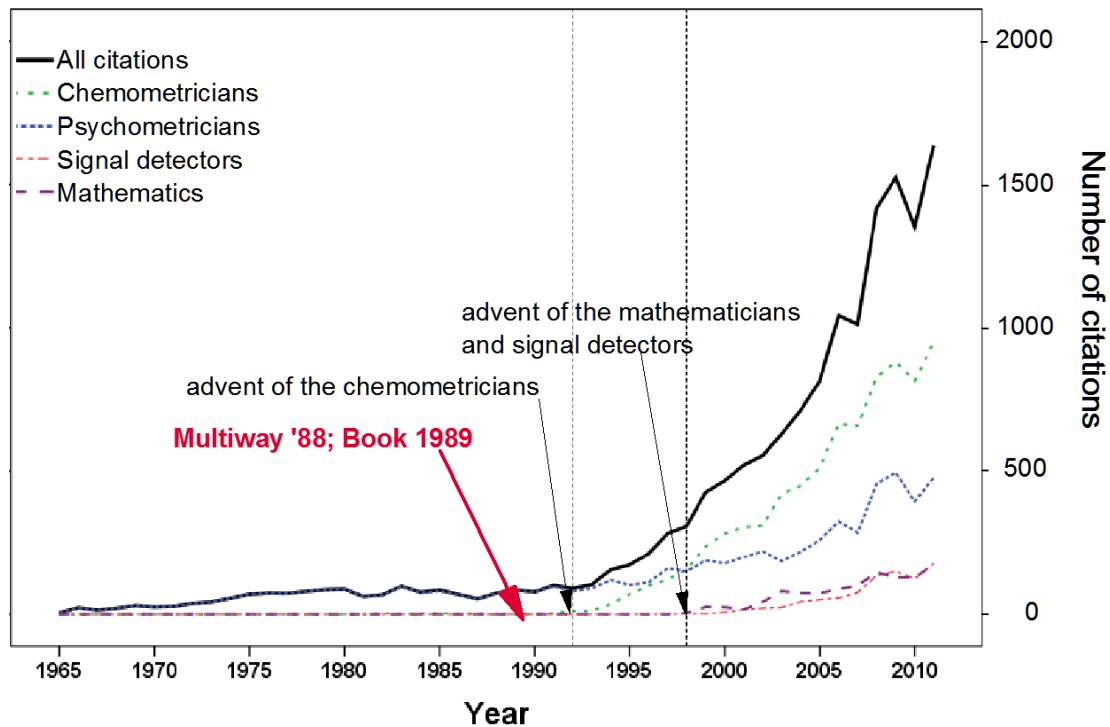
Figure 3: Citations to three-way and multiway papers by researchers from different disciplines. "Multiway 88; Book 1989" refers to the Multiway conference in Rome 1988; Coppi and Bolasco (1989) contains the proceedings.

2.2. The rise of multiway analysis

To get an idea of what went on in the multiway world we can take a look at what happened with the proposed fundamental models over the years; in particular, what happened since Tucker's introduction of three-mode data analysis in psychology (Tucker 1966)[2] and Harshman's ground-breaking work on the PARAFAC (or CP) model (Harshman 1970).

The curves in Figure 3 represent the number of citations to multiway articles written by authors from the fields of psychology, chemistry, signal detection, and mathematics, respectively. It is clear that there was a strong acceleration from 1992 onwards, and that for the time being there seems no end in sight. We can see that particularly in the field of chemistry the unbridled enthusiasm has continued unabated, but that researchers from other disciplines have become involved as well.

By the way, it is interesting to note that theoretical mathematicians such as Hitchcock came up with three-way models as early as 1927, but that for a long time mathematicians seemed to have forgotten about them. It was not until around 1998 that multiway models attracted their attention once more. It is noteworthy that psychometricians also started to publish in chemometric journals (e.g., Kiers and Harshman 1997; Kiers, Ten Berge, and Bro 1999; Harshman 2001; Kroonenberg and Van Ginkel 2012). This is primarily due to an ever-

---

[2]From Neil Dorans's interview with Tucker (Dorans 2004) I learned that Tucker's second initial R was not an initial at all. His parents gave him the single letter as a second name so that there should be no period following the R. Little did they know how important this letter was going to be in statistical programming.
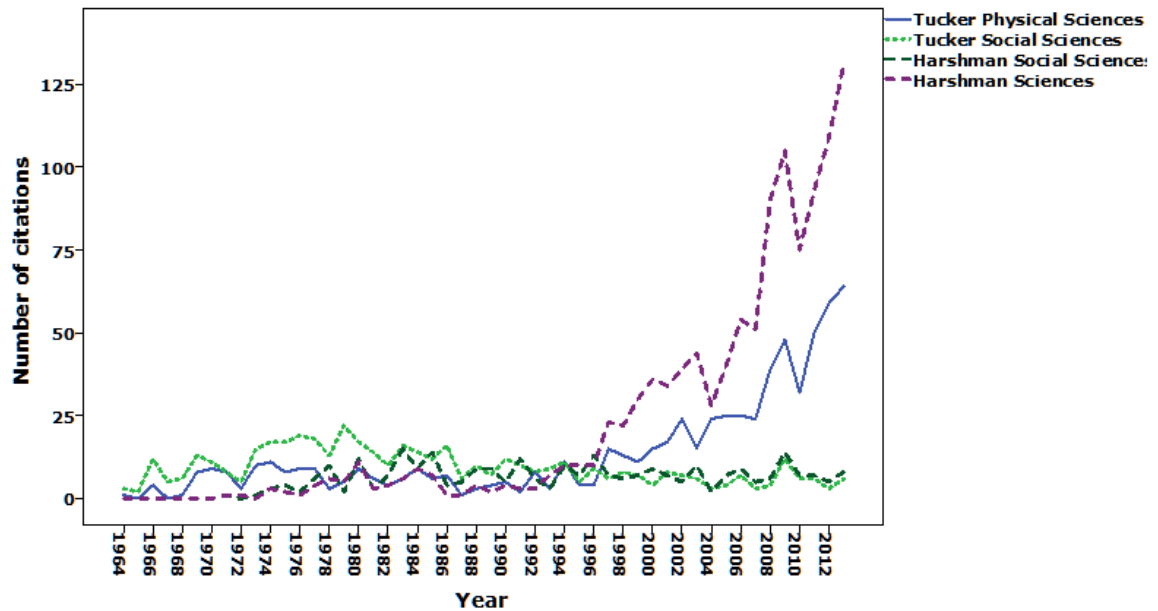
Figure 4: Citations to the work by L. R Tucker and R. A. Harshman in the social and behavioural, and in the physical sciences

increasing collaboration between psychometricians and chemometricians as a result of their triannual TRICAP (three-way methods in chemometrics and psychometrics) conferences; see, e.g., the early conference reports: Geladi (1994), Bro and Kiers (1997), and Van den Berg and Ceulemans (2003).

3. Multiway in the social and behavioural sciences

From Figure 3 one might get the impression that the increase in citations to psychometric three-way articles reflects an upsurge in their use in the social and behavioural sciences. Unfortunately, nothing is further from the truth. Figure 4 shows a very different picture. It is based on references to the work of the founding fathers Tucker and Harshman, split according to whether the citing author works in the natural sciences or in the social and behavioural sciences. We see that the increase in citations to the work of the two psychometricians is exclusively due to researchers from the natural sciences, rather than to researchers from their own discipline.

Why is this? It could be that describing patterns in data does not particularly appeal to social and behavioural scientists, but that they prefer testing new and well-founded theories about human behaviour. Such testing is done through inferential rather than descriptive statistics. It could also be that many social scientists focus on finding the causes of phenomena and predicting future behaviour. This is something three-way methods were not originally designed for, but the first steps in this direction have been made. Another factor that could play a part is a justifiable desire to keep analyses as simple as possible in order to prevent conflicts with journal editors and their reviewers. After all, the latter are not necessarily well-versed in advanced statistical methods. A further important aspect could be that multi-

variate descriptive techniques, and three-way techniques in particular, are mainly intended to model individual differences rather than test group characteristics such as averages, standard deviations, correlations, etc.

Moreover, social and behavioural scientists are missing something that in natural sciences such as chemistry is the motor for much three-way work: substantive theories that take the form of the statistical models used in three-way analysis. If such models are available, three-way techniques directly provide the estimates for the parameters of the formal physical models, and there is no need to wonder afterwards how parameters should be interpreted – which happens not infrequently in the social and behavioural sciences. This difference in modelling is discussed in, for instance, Hand's presidential address to the Royal Statistical Society in 2009 under the heading "iconic models" versus "empirical models". The physical sciences typically deal with the former, while the social and behavioural sciences have to make do with the latter. However, it should be understood that empirical modelling occurs in some way or another in nearly every kind of science, physical science included, as Hand convincingly demonstrates.

4. A cornucopia of applications

In this section I provide a brief and non-exhaustive overview of the breadth of topics that have been researched with multiway methods from domains other than the social and behavioural sciences (which are treated in a separate section). Lieven De Lathauwer mentioned, for example, that tensors, the mathematical elements that form the basis for many multi-way models, are the golden tools for "signal processing, array processing, data mining, machine learning, system modelling, scientific computing, statistics, wireless communication, audio and image processing, biomedical applications, bio-informatics and so forth" (De Lathauwer 2014), not even mentioning fields such as the social and behavioural sciences, environmental studies, and agriculture. Kroonenberg's 2008 book contains examples from several of these last ones.

4.1. Chemistry

Besides physical models, chemistry also employs 'linked' measuring techniques that work together to produce three-way data such as gas chromatography coupled with mass spectrometry. As a result, the chemometricians need techniques to analyse such data, and three-way and multiway methods have come to the rescue. This also resulted in three-way models for online process control in industry. A view of the breadth of applications in chemistry can be found in the review papers by Bro (2006), Açar and Yener (2009), Kolda and Bader (2009), the edited book by Muñoz de la Peña, Goicoechea, Escandar, and Olivieri (2015), and the books by Smilde, Bro, and Geladi (2004) and Olivieri and Escandar (2014).

4.2. Agriculture and food industry

In the field of agriculture, plant breeders want specific varieties of crops, say peanuts with specific attributes to suit specific purposes. They also want to know which crops can best be grown in which locations. Analysis of individual variation is of prime importance here: the industry wants peanuts that can easily be turned into peanut butter, while consumers want

king-size peanuts that feel right in the mouth. Bro's 1998 thesis provides a background to both chemical and food industry applications (see also Smilde *et al.* 2004), while Basford and Kroonenberg discuss applications in agriculture, especially plant breeding, in various papers (see, e.g., Basford, Kroonenberg, and Cooper 1996).

4.3. Signal detection

Three-way analysis is of prime importance for signal detection, particularly for keeping apart signals from different devices - blind source separation. Are there nowadays people who do not have a device on them that constantly detects signals, whether it is a mobile phone, tablet or sat-nav? It is in this field that there is a great upsurge of applications. Much about these developments can be found in the Handbook of Blind Source Separation (Comon and Jutten 2010, Chapter 13), the review paper by Cichocki, Mandic, De Lathauwer, Zhou, Zhao, Caiafa, and Phan (2015), and the book by Cichocki, Zdunek, Phan, and Amari (2009).

4.4. Systems biology

A very special area of application is systems biology - a discipline that aims to study organisms as a whole, with specialisms such as metabolomics (the systematic study of the unique chemical characteristics of specific cell processes) and proteomics (the large-scale study of proteins, particularly their structure and function). Examples of the use of multiway methods in these areas can be found in, for instance, Yener, Açar, Aguis, Bennett, Vandenberg, and Plopper (2008); Rubingh, Bijlsma, Jellema, Overkamp, Van der Werf, and Smilde (2009); and Açar, Lawaetz, Rasmussen, and Bro (2013).

5. Applications in the social and behavioural sciences

As shown above, three-way analysis is not commonly used within the social and behavioural sciences, although there are many kinds of projects that would benefit from its power and the insights derived from three-way techniques. If we want to explore the use of three-way analysis in social and behavioural scientific research in more depth, a crucial question is: "Is descriptive analysis, and three-way analysis in particular, actually useful in the social and behavioural sciences?". Despite the relatively modest growth in three-way applications, I maintain that this is definitely the case, but to my mind this applies to all descriptive analytical techniques.

5.1. Three-way data types in social and behavioural research

To answer the question of the usefulness of three-way methods in the social and behavioural sciences in detail, we must first establish if the data collected within this discipline actually call for three-way techniques. In order to eliminate any doubt on this point, I here provide a non-exhaustive overview of major types of three-way data encountered in the social and behavioural sciences; for applications see, e.g., Kroonenberg (2008).

- **Three-way profile data**
  Persons have scores on a number of variables measured under several conditions or on several occasions.

Example: It is recorded whether children seek proximity to strangers, observe them with distrust, etc. when child, parent, and stranger are in the same room; when the child is alone with the stranger; and when the stranger enters after the child has been left alone in the room.

- **Repeated measures data**
  A number of participants are measured several times on the same variables.
  Example: Each year in a sample of primary school children their reading ability is assessed on several different measures.

- **Three-way rating data**
  A number of subjects judge various emotions in different situations leading to a kind of stimulus-response data.
  Example 1: Children with asthma indicate on several scales how they would react in different situations, for example playing alone in the street, walking alone in the woods, playing with their friends in the school grounds, going on a cycling trip. For each of these situations the children are asked to imagine how afraid they would be to have an asthmatic attack, how lonely they would feel, or how exhilarated they would be.
  Example 2: Individuals are presented with a number of concepts pertaining to their day-to-day life which have to be judged with a set of scales describing the characteristics of these concepts. A well-known representative of this kind of data is Osgood's semantic differential technique. Therapists presented a woman with a multiple personality with concepts regarding her daily life (state of mind, mental illness, her doctor, her spouse, her job, etc.) and asked her to rate the concepts on scales such as good versus bad, calm versus vehement, strong versus weak (see Osgood and Luria 1954).

- **Sets of correlation or covariance matrices**
  Example: The correlations or covariances between subtests of an intelligence test are computed for samples of children at different ages, or from different countries. The question is whether the intelligence test has the same structure over time and/or in different languages.

- **Sets of similarity matrices**
  Example: Children are asked to indicate to what extent they prefer peanut butter sandwiches over jam sandwiches, cheese, kippers, truffles, etc. An additional three-way profile data set with the characteristics of these food stuffs can be useful for explicitly linking the judged preferences between the food stuffs to their perceived characteristics.

- **Results of fMRI studies**
  Example: The results of brain scans measured in various locations in the brain, under different conditions, in different people.

- **Three-way interactions**
  Example 1: Three-way interactions resulting from analysis of variance. If the three (or more) factors in the `anova` have only few categories, analysing the three-way interaction with three-way component methods is hardly profitable. This changes, however, if there are many levels. In a study on the perceived reality of television films for children (the response variable) there were eight categories of films, five age groups, and eleven perceived reality aspects. This leads to an $8 \times 5 \times 11$ three-way interaction, and next to

other effects, it is this interaction that was the focus of the study (Kroonenberg and Van der Voort 1987). A technical discussion can be found in Van Eeuwijk and Kroonenberg (1998) and an example from nursing in Gilbert, Sutherland, and Kroonenberg (2000).
Example 2: Large three-way interaction tables also occur in the case of large contingency tables with three variables with many categories. As an example, suppose that every month the behaviours between mothers and their babies are observed using a number of behavioural categories; see, e.g., the study by Van den Boom and Hoeksma (1994), analysed by Carlier and Kroonenberg (1996).

- **Large binary three-way data sets**
  Example: In a study 54 persons were asked to indicate with yes or no whether they displayed any of 15 hostile behaviors in 23 frustrating situations. This resulted in a binary 23 x 15 x 54 data array. Such an array is clearly too large to analyse for many classical techniques, but three-way binary component analysis, called three-way hierarchical classes analysis, is specifically designed to handle this type of large binary contingency tables (Ceulemans, Van Mechelen, and Leenen 2003; Ceulemans and Van Mechelen 2005).

The last two data types deserve further mention because they point to areas where three-way analysis within various disciplines has been underused. Categorical and binary data are fairly common in the social and behavioural sciences, as both criterion variables and predictor variables. For designs with one or more continuous variables and many levels in the categorical predictors, analysis of variance is used in many cases, and more than two factors with many levels quickly lead to large interaction tables. Similarly, large three-way contingency tables have large three-way interaction tables. Indicating that these interactions are significant hardly answers the research questions about the major underlying patterns.

Bro and Jakobsen (2002) developed the GEMANOVA procedure for handling higher-dimensional interactions with multiway methods; one example is Romano, Næs, and Brockhoff (2015). Three-way interactions for contingency tables have been tackled with three-way correspondence analysis by Carlier and Kroonenberg (1996, 1998). In their book Beh and Lombardo (2015, Chapter 11) provide a contemporary overview of several variants of multiway correspondence analysis.

From the above it should be clear that research in the social and behavioural sciences produces many different kinds of data which can seriously benefit from three-way and multiway analysis. At the same time it is clear that large-scale descriptive analyses are not everybody's cup of tea, and that there will always be interpretational issues, because understanding the patterns in three different component spaces and the way they are linked is not something that comes easily. On the other hand, the techniques allow for answers to complicated questions about interactions, although the answers will not necessarily be straightforward.

6. Three-way and multiway software

During my academic career I have been working on several programs for three-way analysis plus one four-way program. These programs have been used in virtually all my published applied papers. In this section I will set out some considerations that went into the development of these programs and the interface binding them together in a single integrated
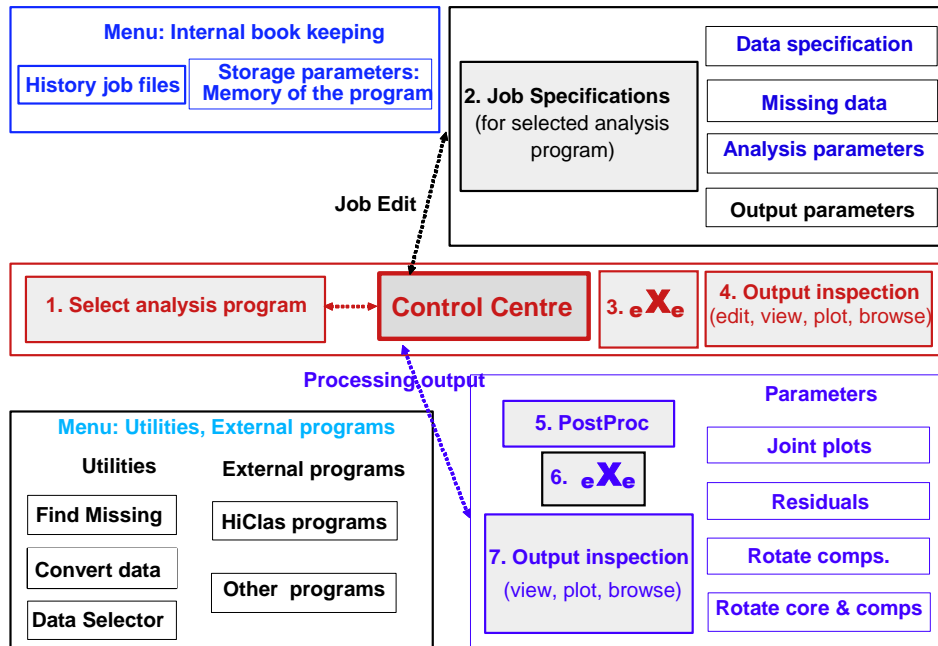
TWPack structure

Figure 5:   **TWPack** content structure.

program suite. I do not claim that all aspects mentioned here will always be relevant in program development but I expect that the kinds of issues, problems and solutions I came across in this context will occur in similar situations. Some of the problems and my solutions may at times seem a bit outdated, but that follows from my early start in the business.

6.1. Experiences in developing three-way software

It all started in 1975 when I wrote a Fortran program for a variant of the Tucker3 model, in which components are computed for only two of the three ways. This was followed by a program for the Tucker3 model itself (see Figure 2). Unsurprisingly, these two programs were later merged into one, and obviously the PARAFAC or CP model had to be included as well. Later I programmed additional models such as the simultaneous component model (Timmerman and Kiers 2003), three-mode correspondance analysis (Carlier and Kroonenberg 1996) and a few others. To be able to handle all these programs some sort of integrated environment was necessary; furthermore, such an environment is not complete without all kinds of utilities for preprocessing, postprocessing, graphing the results, evaluating residuals, etc.

Luckily I decided early on that I wanted to have a graphical interface independent of the analysis programs themselves. Due to various circumstances the first of such interfaces (**IF3**) was written in Pascal (Kroonenberg 1996). Later, two different versions were written in Delphi, the most recent one called **TWPack**; a manual (Kroonenberg and De Roo 2010). Information about the availability of the program suite can be found on the website of The Three-Mode Company (Kroonenberg 2016). Separating the analysis programs from the interface was one of the most influential and defining decisions for my practical work on three-way analysis, as

it enabled the individual programs to also function as stand-alone programs with a complete query-and-answer input system as well as a file input system. This in turn allowed for their direct inclusion in simulation programs and all kinds of different interfaces.

The interface was designed in such a way that firstly, menu systems create the necessary parameter files for the analysis programs; see Figure 5 for the structure of the interface. The interface then calls the desired analysis program, which reads data files and the created parameter files for the job at hand. The analysis program produces output files containing plain ascii and browser-readable html. In addition, the program itself creates a parameter file for postprocessing the output, as well as a command file with instructions for the graphics program **gnuplot** (Williams and Kelley 2016). Postprocessing consists of rotating the components and/or the core array, analysing and displaying residuals, as well as creating further graphics. A browser internal to the interface displays the output in a form which is reminiscent of the standard SPSS, as it also has a sidebar for navigation. The navigability is a great plus given the large amount of output.

For all of this to work an interface needs to have a *memory*. In the SPSS statistical package the memory consists of the system file (`*.sav`), which contains the data and their definitions such as variable names, value names, format, etc. as well as information about missing data, which variables should be analysed as a set, etc. SPSS saves the information about previous analyses in a syntax file (`*.sps`), thus making rerunning analyses possible. By default the SPSS syntax is also listed in the output and can be copied into the syntax file. The format of the SPSS system file is a proprietary one, but can be read by several other statistical packages such SAS and EQS.

For each data set and its analysis the **TWPack** interface has one single memory or *job file* containing a basic description of the data, the locations of the relevant parameter files, and the details of the previous analysis, all of which should reside in the data directory. A data file has its own companions in a label file and a missing data file (if missing data are present). Job files are editable by a basic ascii editor. In order to keep track of what has been done to or with the data, it is necessary to have a separate directory for each data set. The great disadvantage of this system is that after a series of analyses one has a large number of files floating around in the data directory (see Figure 5). However, all files related to a particular job have the same job name as part of their file name.

Unlike the **TWPack** interface, SPSS (and probably many other statistical programs) does not do reverse engineering, i.e., given a syntax or command file the program cannot set the menu system to correspond with the syntax file. For really complicated syntax files reverse engineering would be a great boon for users, because they do not necessarily have sufficient knowledge of the intricacies of the command language to easily make the same modifications they can make via the menu system.

An attractive aspect of separating interface and analysis programs is that if operation systems change the analysis programs one can just rbe ecompiled, and if other interface building software becomes available or the current one becomes obsolete, one can build and rebuild this separately. Ideally one should be able to build interfaces in interpretational software such as R or MATLAB and still make use of the compiled analysis programs for fast executions, but such interfaces have not been built yet. The idea of separating the two types of programs is even extended in such a way that via relatively minor changes in the interface, external analysis programs with their own built-in interfaces can be called from within **TWPack**.

6.2. Three-way and multiway software in the world at large

Naturally, there are several other collections of programs which are geared to performing three-way and multiway analysis, of which the MATLAB **N-way Toolbox** is probably the widest range of programs.

MATLAB toolboxes

- Andersson and Bro (2000): **N-way Toolbox**
  http://www.models.life.ku.dk/nwaytoolbox);
  see also the Eigenvector Research's **PLS Toolbox**
  http://www.eigenvector.com/software/pls_toolbox.htm.

- Bader and Kolda (2006): **Tensor Toolbox**
  http://www.sandia.gov/~tgkolda/TensorToolbox/.

- Comon, Luciani, and de Almeida (2009): **Tensor Package**
  http://www.gipsa-lab.grenoble-inp.fr/ pierre.comon/TensorPackage/
  tensorPackage.html

- De Lathauwer et al.: **TensorLab** Toolbox (Vervliet, Debals, Sorber, Van Barel, and De Lathauwer 2016)
  http://www.tensorlab.net/.

R packages

- Leibovici (2010): **PTAk**
  https://CRAN.R-project.org/package=PTAk.

- Giordani, Kiers, and Del Ferraro (2014): **ThreeWay**
  https://CRAN.R-project.org/package=ThreeWay.

- Of course Jan de Leeuw (together with Patrick Mair) has an R package with three-way methods: **smacof**, but this is primarily concerned with multidimensional scaling (De Leeuw and Mair 2009)
  https://CRAN.R-project.org/package=smacof.

Undoubtedly several other researchers have developed relevant software in this area. A full comparison of all these tool boxes and packages would be a worthwhile undertaking for a later paper.

7. Assignments for three-way researchers

Part of the blame for the limited use of three-way methods, especially in the social and behavioural sciences, lies with the psychometricians themselves. As lecturers, I think we have neglected to stop and consider that the use of statistical methods is dictated in part by what we teach and how we present these methods to our colleagues and students. In the future

Figure 6: Power to three-way analysis.

we will have to take into account that academics are interested in their own fields, not in our methods. Thus, if we want to show them that we have something to offer, the burden is on us to demonstrate this. In particular, we have to show that three-way methods have an added value compared to the standard statistical toolkit and can deliver results that cannot be obtained any other way. In other words, we have to argue that three-way analysis is like Heineken (see Figure 6).

7.1. Statistical packages: Calculator and adviser in one

In concrete terms I would like to encourage methodologists (including myself) to concentrate, more than they have done so far, on: (1) showing the added value of the techniques in applied articles in non-methodological journals; (2) presenting relevant analyses in collaboration with researchers working in the field, at the latter's own, discipline-oriented, conferences; (3) training young academics specialising in these subjects in the use of three-way techniques, by organising courses that offer plenty of opportunity to practise with their own data; (4) producing guides for performing analyses, not just in print but also through intelligent and self-explanatory computer programs with easily accessible input and output.

Generally, statistical programs for three-way analysis are not half as simple to use as for instance SPSS with its extensive graphical interface. Quite apart from this, I would like programs, including my own, to follow Leland Wilkinson's example in his **SkyTree Adviser** statistical suite. He describes his system as follows: "**SkyTree Adviser** is an expert system for machine learning. It not only chooses appropriate models, but it also evaluates them with unique algorithms for detecting miss-specifications, outliers, and other anomalies (Wilkinson 2014)." Thus, after performing a basic analysis the program will evaluate the outcomes and advise the user on the status of the analysis. In a sense, it talks to the researcher. It will also print off a paragraph in which the results are described and the technical implications of the
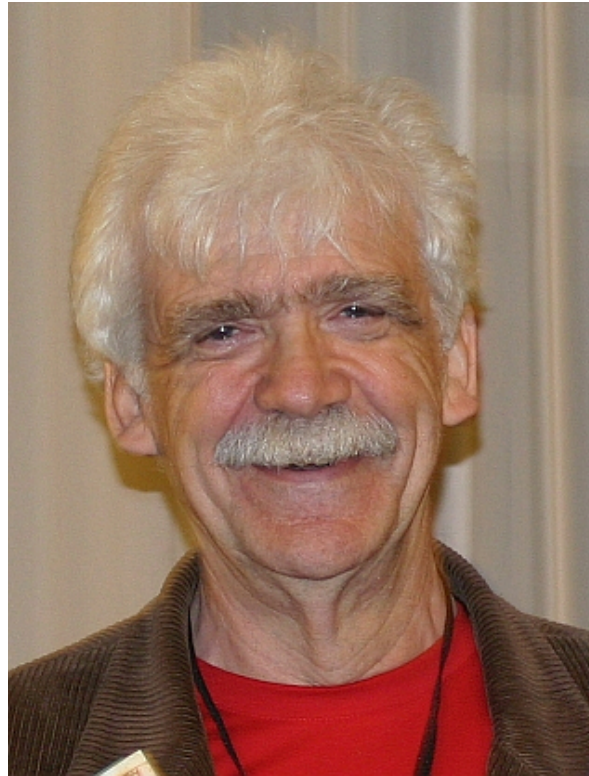
Figure 7: Jan de Leeuw, 2011, Rennes, France.

results are explained. In this way, the program functions as an 'analyser' and personal adviser in one. Various multivariate statistics textbooks give examples of how the results of a certain analysis can be interpreted, but they always deal with the authors', not the researcher's data. It would be ideal if programs for statistical analysis explain the meaning of the results provided in their output. Until now, user-friendliness of computer programs has been almost exclusively discussed in terms of input rather than output and interpretation. Creating three-way analysis programs with user-friendly output seems like a perfect, rewarding project for a retired professor.

8. Conclusion: Returning to the beginnings

Thus, from the humble beginnings of my master's thesis and the related program, over the years I have developed an integrated system for various, but of course not all possible, programs for three-way analysis. Jan de Leeuw stood at the cradle of my career in three-way analysis and as a good parent he released me into adulthood by pointing me the way to go without imposing his ideas on how it should be done. In other words, an ideal academic parent for an independent-spirited young researcher. My only regret is that he spent so much of his life far, far away in California so that we really did not have much contact after my formative years. I think the time has come to confront my ideas about software expressed above with Jan's long-term experience in developing software in R, and find out how the two approaches can be most fruitfully combined.

References

Açar E, Lawaetz AJ, Rasmussen MA, Bro R (2013). "Structure-Revealing Data Fusion Model with Applications in Metabolomics." In *EMBC'13: Proceedings of 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 6023–6026. IEEE, New York.

Açar E, Yener B (2009). "Unsupervised Multiway Data Analysis: A Literature Survey." *IEEE Transactions on Knowledge and Data Engineering*, **21**, 6–20. `doi:10.1109/tkde.2008.112`.

Andersson CA, Bro R (2000). "The *N*-Way Toolbox for MATLAB." *Chemometrics and Intelligent Laboratory Systems*, **52**, 1–4. `doi:10.1016/s0169-7439(00)00071-x`.

Bader BW, Kolda TG (2006). "Algorithm 862: MATLAB Tensor Classes for Fast Algorithm Prototyping." *ACM Transactions on Mathematical Software*, **32**, 635–653. `doi:10.1145/1186785.1186794`.

Basford KE, Kroonenberg PM, Cooper M (1996). "Three-Mode Analytical Methods for Crop Improvement Programs." In M Cooper, GL Hammer (eds.), *Plant Adaptation and Crop Improvement*, pp. 291–305. CAB International, Wallingford.

Beh EJ, Lombardo R (2015). *Correspondence Analysis. Theory, Practice and New Strategies*. John Wiley & Sons, Chicester.

Bro R (1998). *Multi-Way Analysis in the Food Industry. Models, Algorithms, and Applications*. Ph.D. thesis, University of Amsterdam. URL `http://www.models.kvl.dk/users/rasmus/brothesis.pdf`.

Bro R (2006). "Review on Multiway Analysis in Chemistry – 2000–2005." *Critical Reviews in Analytical Chemistry*, **36**, 279–293. `doi:10.1080/10408340600969965`.

Bro R, Jakobsen M (2002). "Exploring Complex Interactions in Designed Data Using GEMANOVA. Color Changes in Fresh Beef during Storage." *Journal of Chemometrics*, **16**, 294–304. `doi:10.1002/cem.722`.

Bro R, Kiers HAL (1997). "TRICAP: Three-Way Methods in Chemistry and Psychology, Lake Chelan, WA, USA, 4–9 May 1997." *Journal of Chemometrics*, **11**, 463–466. `doi:10.1002/(sici)1099-128x(199709/10)11:5<463::aid-cem494>3.3.co;2-8`.

Carlier A, Kroonenberg PM (1996). "Decompositions and Biplots in Three-Way Correspondence Analysis." *Psychometrika*, **61**, 355–373. `doi:10.1007/bf02294344`.

Carlier A, Kroonenberg PM (1998). "The Case of the French Cantons: An Application of Three-Way Correspondence Analysis." In J Blasius, M Greenacre (eds.), *Visualization of Categorical Data*, pp. 253–275. Academic Press, New York.

Carroll JD, Chang JJ (1970). "Analysis of Individual Differences in Multidimensional Scaling via an N-Way Generalization of "Eckart-Young" Decomposition." *Psychometrika*, **35**, 283–319. `doi:10.1007/bf02310791`.

Ceulemans E, Van Mechelen I (2005). "Hierarchical Classes Models for Three-Way Three-Mode Binary Data: Interrelations and Model Selection." *Psychometrika*, **70**, 461–480. `doi:10.1007/s11336-003-1067-3`.

Ceulemans E, Van Mechelen I, Leenen I (2003). "Tucker3 Hierarchical Classes Analysis." *Psychometrika*, **68**, 413–433. `doi:10.1007/bf02294735`.

Cichocki A, Mandic D, De Lathauwer L, Zhou G, Zhao Q, Caiafa C, Phan HA (2015). "Tensor Decompositions for Signal Processing Applications: From Two-Way to Multiway Component Analysis." *IEEE Signal Processing Magazine*, **32**, 145–163. `doi:10.1109/msp.2013.2297439`.

Cichocki A, Zdunek R, Phan AH, Amari SI (2009). *Nonnegative Matrix and Tensor Factorizations. Applications to Exploratory Multi-Way Data Analysis and Blind Source Separation.* John Wiley & Sons, Chicester.

Comon P, Jutten C (eds.) (2010). *Handbook of Blind Source Separation.* Academic Press, Oxford.

Comon P, Luciani X, de Almeida ALF (2009). "Tensor Decompositions, Alternating Least Squares and Other Tales." *Journal of Chemometrics*, **23**, 393–405. `doi:10.1002/cem.1236`.

Coppi R, Bolasco S (eds.) (1989). *Multiway Data Analysis.* North Holland, Amsterdam.

De Lathauwer L (2014). "Tensor Decompositions: Golden Tools for Data Mining." Lecture presented at the 25th Anniversary VOC. Rolduc, The Netherlands. 7 November 2014.

De Leeuw J (1975). "The TUCKALS2 Algorithm – Version 2." Unpublished, URL `http://www.stat.ucla.edu/~deleeuw/janspubs/1975/notes/deleeuw_U_75c.pdf`.

De Leeuw J (2016a). "Bibliography of Jan De Leeuw." Accessed 2016-06-01, URL `http://gifi.stat.ucla.edu/`.

De Leeuw J (2016b). "Curriculum Vitae of Jan De Leeuw." Accessed 2016-06-01, URL `http://gifi.stat.ucla.edu/cv-year.pdf`.

De Leeuw J (2016c). "Pronounce My Last Name." Accessed 2016-06-01, URL `http://gifi.stat.ucla.edu/leeuw.mov`.

De Leeuw J, Mair P (2009). "Multidimensional Scaling Using Majorization: SMACOF in R." *Journal of Statistical Software*, **31**(3), 1–30. `doi:10.18637/jss.v031.i03`.

Dorans NJ (2004). "A Conversation with Ledyard R Tucker." *Technical report*, ETS. Available from `https://www.ets.org/Media/Research/pdf/tucker.pdf`.

Geladi P (1994). "TRIC, The First Conference on Three-Way Methods in Chemistry, Epe, Netherlands, 22–25 August, 1993." *Journal of Chemometrics*, **8**, 97–98. `doi:10.1002/cem.1180080110`.

Gilbert DA, Sutherland M, Kroonenberg PM (2000). "Exploring Subject-Related Interactions in Repeated Measures Data Using Three-Mode Principal Components Analysis." *Nursing Research*, **49**, 57–61. `doi:10.1097/00006199-200001000-00009`.

Giordani P, Kiers HAL, Del Ferraro MA (2014). "Three-Way Component Analysis Using the R Package **ThreeWay**." *Journal of Statistical Software*, **57**(7), 1–23. `doi:10.18637/jss.v057.i07`.

Hand DJ (2009). "Modern Statistics: The Myth and the Magic." *Journal of the Royal Statistical Society A*, **172**, 287–306. `doi:10.1111/j.1467-985x.2009.00583.x`.

Harshman RA (1970). "Foundations of the PARAFAC Procedure: Models and Conditions for an "Explanatory" Multi-Modal Factor Analysis." *UCLA Working Papers in Phonetics*, **16**, 1–84.

Harshman RA (2001). "An Index Formalism That Generalizes the Capabilities of Matrix Notation and Algebra to N-Way Arrays." *Journal of Chemometrics*, **15**, 689–714. `doi:10.1002/cem.665`.

Harshman RA, Lundy ME (1984a). "Data Preprocessing and the Extended PARAFAC Model." In Law *et al.* (1984), pp. 216–284.

Harshman RA, Lundy ME (1984b). "The PARAFAC Model for Three-Way Factor Analysis and Multidimensional Scaling." In Law *et al.* (1984), pp. 122–215.

Harshman RA, Lundy ME (1994). "PARAFAC: Parallel Factor Analysis." *Computational Statistics & Data Analysis*, **18**, 39–72. `doi:10.1016/0167-9473(94)90132-5`.

Hitchcock FL (1927a). "Multiple Invariants and Generalized Rank of a p-Way Matrix or Tensor." *Journal of Mathematics and Physics*, **7**, 39–79. `doi:10.1002/sapm19287139`.

Hitchcock FL (1927b). "The Expression of a Tensor or a Polyadic as a Sum of Products." *Journal of Mathematics and Physics*, **6**, 164–189. `doi:10.1002/sapm192761164`.

Kiers HAL, Harshman RA (1997). "Relating Two Proposed Methods for Speedup of Algorithms for Fitting Two- And Three-Way Principal Component and Related Multilinear Models." *Chemometrics and Intelligent Laboratory Systems*, **36**, 31–40. `doi:10.1016/s0169-7439(96)00074-3`.

Kiers HAL, Ten Berge JMF, Bro R (1999). "PARAFAC2 – Part I. A Direct Fitting Algorithm For the PARAFAC2 Model." *Journal of Chemometrics*, **13**, 275–294. `doi:10.1002/(sici)1099-128x(199905/08)13:3/4<275::aid-cem543>3.0.co;2-b`.

Kolda TG, Bader BW (2009). "Tensor Decompositions and Applications." *SIAM Review*, **51**, 455–500. `doi:10.1137/07070111x`.

Kroonenberg PM (1983). *Three-Mode Principal Component Analysis: Theory and Applications.* DSWO Press, Leiden. Book available from `http://www.leidenuniv.nl/fsw/three-mode/bibliogr/kroonenbergpm_thesis/kroonenberg1983.pdf`; Errata, 1989; available from the author.

Kroonenberg PM (1984). "Three-Mode Principal Component Analysis: Illustrated with an Example from Attachment Theory." In Law *et al.* (1984), pp. 64–103.

Kroonenberg PM (1996). "**3WayPack** User's Manual (Version 2)." *Program manual*, The Three-Mode Company, Leiden University, Leiden. Available from `http://three-mode.leidenuniv.nl/`.

Kroonenberg PM (2005). "Three-Mode Component and Scaling Models." In BS Everitt, D Howell (eds.), *Encyclopedia of Statistics in Behavioral Sciences*, pp. 2032–2044. John Wiley & Sons, Chichester.

Kroonenberg PM (2008). *Applied Multiway Data Analysis*. John Wiley & Sons, Hoboken.

Kroonenberg PM (2014a). "History of Multiway Component Analysis and Three-Way Correspondence Analysis." In J Blasius, MJ Greenacre (eds.), *Visualization and Verbalization of Data*, pp. 77–94. Chapman & Hall/CRC, New York.

Kroonenberg PM (2014b). "De Driewegwereld Kent Geen Grenzen [The Three-Way World Knows No Boundaries]." *Afscheidscollege. [Farewell lecture]*, Department of Education, Leiden University, Leiden. Available in English and Dutch from `http://three-mode.leidenuniv.nl/`.

Kroonenberg PM (2016). "The Three-Mode Company." Accessed 2016-06-01, URL `http://three-mode.leidenuniv.nl/`.

Kroonenberg PM, De Leeuw J (1977). "TUCKALS2: A Principal Component Analysis of Three-Mode Data." *Research Bulletin RB 001-'77*, Department of Data Theory, Leiden University, Leiden.

Kroonenberg PM, De Leeuw J (1978). "TUCKALS2: Een Hoofdassenanalyse Voor Drieweggegevens. [TUCKALS2: A Principal Component Analysis for Three-Mode Data]." *Methoden en Data Nieuwsbrief*, **3**, 30–53.

Kroonenberg PM, De Leeuw J (1980). "Principal Component Analysis of Three-Mode Data by Means of Alternating Least Squares Algorithms." *Psychometrika*, **45**, 69–97. `doi:10.1007/bf02293599`.

Kroonenberg PM, De Roo Y (2010). "**3WayPack**: A Program Suite for Three-Way Analysis." *Program manual*, The Three-Mode Company, Leiden University, Leiden. Available from `http://three-mode.leidenuniv.nl/`.

Kroonenberg PM, Van der Voort THA (1987). "Multiplicatieve Decompositie Van Interacties Bij Oordelen over De Werkelijkheidswaarde Van Televisiefilms [Multiplicative Decomposition of Interactions from Judgements about the Preceived Reality of Television Films]." *Kwantitatieve Methoden*, **8**(23), 117–144.

Kroonenberg PM, Van Ginkel JR (2012). "Combination Rules for Multiple Imputation in Three-Way Analysis Illustrated with Chromatography Data." *Current Analytical Chemistry*, **8**, 224–235. `doi:10.2174/157341112800392544`.

Law HG, Snyder Jr CW, Hattie JA, McDonald RP (eds.) (1984). *Research Methods for Multimode Data Analysis*. Praeger, New York.

Leibovici DG (2010). "Spatio-Temporal Multiway Decompositions Using Principal Tensor Analysis on k-Modes: The R Package **PTAk**." *Journal of Statistical Software*, **34**(10), 1–34. `doi:10.18637/jss.v034.i10`.

Muñoz de la Peña A, Goicoechea HC, Escandar GM, Olivieri AC (2015). *Fundamentals and Analytical Applications of Multiway Calibration*. Elsevier, Amsterdam.

Olivieri AC, Escandar GM (2014). *Practical Three-Way Calibration.* Elsevier, Waltham.

Osgood CE, Luria Z (1954). "A Blind Analysis of a Case of Multiple Personality." *Journal of Abnormal and Social Psychology*, **49**, 579–591. `doi:10.1037/h0054362`.

Romano R, Næs T, Brockhoff PB (2015). "Combining Analysis of Variance and Three-Way Factor Analysis Methods for Studying Additive and Multiplicative Effects in Sensory Panel Data." *Journal of Chemometrics*, **29**, 29–37. `doi:10.1002/cem.2659`.

Rubingh CM, Bijlsma S, Jellema RH, Overkamp KM, Van der Werf MJ, Smilde AK (2009). "Analyzing Longitudinal Microbial Metabolomics Data." *Journal of Proteome Research*, **8**, 4319–4327. `doi:10.1021/pr900126e`.

Smilde AK, Bro R, Geladi P (2004). *Multi-Way Analysis: Applications in the Chemical Sciences.* John Wiley & Sons, Chichester.

Timmerman ME, Kiers HAL (2003). "Four Simultaneous Component Models for the Analysis of Multivariate Time Series from More than One Subject to Model Intraindividual and Interindividual Differences." *Psychometrika*, **68**, 105–121. `doi:10.1007/bf02296656`.

Torgerson WS (1958). *Theory and Methods of Scaling.* John Wiley & Sons, New York.

Tucker LR (1963). "Implications of Factor Analysis of Three-Way Matrices for Measurement of Change." In CW Harris (ed.), *Problems in Measuring Change*, pp. 122–137. University of Wisconsin Press, Madison.

Tucker LR (1966). "Some Mathematical Notes on Three-Mode Factor Analysis." *Psychometrika*, **31**, 279–311. `doi:10.1007/bf02289464`.

Tucker LR (1972). "Relations between Multidimensional Scaling and Three-Mode Factor Analysis." *Psychometrika*, **37**, 3–27. `doi:10.1007/bf02291410`.

Van den Berg F, Ceulemans E (2003). "TRICAP 2003: Three-Way Methods in Chemistry and Psychology, University of Kentucky, Lexington, KY, USA, 22–27 June 2003." *Journal of Chemometrics*, **17**, 618–619. `doi:10.1002/cem.823`.

Van den Boom DC, Hoeksma JB (1994). "The Effect of Infant Irritability on Mother-Infant Interaction: A Growth Curve Analysis." *Developmental Psychology*, **30**, 581–590. `doi:10.1037/0012-1649.30.4.581`. For the data see `http://www.leidenuniv.nl/fsw/three-mode/data/irritablebabiesinfo.htm`.

Van Eeuwijk E, Kroonenberg PM (1998). "Multiplicative Models for Interaction in Three-Way ANOVA, with Applications to Plant Breeding." *Biometrics*, **54**(4), 1315–1333. `doi:10.2307/2533660`.

Vervliet N, Debals O, Sorber L, Van Barel M, De Lathauwer L (2016). "**TensorLab** 3.0." *Program manual*, ESAT, Katholieke Universiteit Leuven, Leuven, Belgium. Available from `http://www.tensorlab.net/`.

Wilkinson L (2014). "Adviser: Learning How to Get A Second Opinion on Your Analysis When It's Important to Get It Right." Strata 2014 Conference Abstract, Accessed 06-01-2016, URL `http://conferences.oreilly.com/strata/strata2014/public/schedule/detail/32843`.

Williams T, Kelley C (2016). "**gnuplot** 5.0.2: An Interactive Plotting Program." URL http://gnuplot.info/.

Yener B, Açar E, Aguis P, Bennett K, Vandenberg SL, Plopper GE (2008). "Multiway Modeling and Analysis in Stem Cell Systems Biology." *BMC Systems Biology*, **2**(63). doi:10.1186/1752-0509-2-63.

Affiliation:

Pieter M. Kroonenberg
Faculty of Social and Behavioural Sciences
Leiden University
2333 AK Leiden, The Netherlands
E-mail: kroonenb@fsw.leidenuniv.nl
and
The Three-Mode Company
URL: http://three-mode.leidenuniv.nl/