Expanding the methodological toolbox of HRM researchers:
The added value of latent bathtub models and optimal matching analysis.

van der Laken, P.A.1, Bakk, Z.2, Giagkoulas, V.3, van Leeuwen, L.3, Bongenaar, E.M.3

1 Human Resource Studies, Tilburg University, Tilburg, the Netherlands
2 Methodology & Statistics, Leiden University, Leiden, the Netherlands
3 HR Analytics, Shell International, The Hague, the Netherlands

Contact information:
Paul van der Laken
p.a.vdlaken@tilburguniversity.edu
Dpt. Human Resource Studies, Tilburg University

Forthcoming in Human Resources Management (Special Issue on “Workforce Analytics”)

Abstract:

Researchers frequently rely on general linear models (GLM) to investigate the impact of human resource management (HRM) decisions. However, the structure of organizations and recent technological advancements in the measurement of HRM processes cause contemporary HR data to be hierarchical and/or longitudinal. At the same time, the growing interest in effects at different levels of analysis and over prolonged periods of time further drives the need for HRM researchers to differentiate from traditional methodology. While multi-level techniques have become more common, this paper proposes two additional methods which may complement the current methodological toolbox of HRM researchers. Latent bathtub models can accurately describe the multi-level mechanisms occurring in organizations, even if the outcome resides at the higher level of analysis. Optimal matching analysis can be useful to unveil longitudinal patterns in HR data, particularly in contexts where HRM processes are measured on a continuous basis. Illustrating the methods’ applicability to research on employee engagement, this paper demonstrates that the HRM community – both research and practice – can benefit from a more diversified methodological toolbox, drawing on techniques from in- and outside the direct field to improve the decision-making process.

Keywords: HR analytics, evidence-based HRM, bathtub models, latent variable models, optimal matching analysis, multi-level, longitudinal
Introduction

Human resource management (HRM) emerged as a function in the early 20th century to effectively manage and rationalize the employment relationship (Ulrich & Dulebohn, 2015). Nowadays, HRM is increasingly becoming a ‘science’ that aims to enhance the decisions organizations make regarding their human capital (Boudreau & Ramstad, 2005; 2007; Rasmussen & Ulrich, 2015; Ulrich & Dulebohn, 2015). In creating a basis of evidence for such decisions, HRM scholars have primarily relied on general linear models (GLM) such as linear regression. However, the data gathered and compiled in the contemporary HR function is increasingly of hierarchical and longitudinal nature, causing the current methodological toolbox of HRM researchers to fall short (Angrave, Charlwood, Kirkpatrick, Lawrence, & Stuart, 2016; Bersin, 2015).

Methods other than GLM may better account for the complex effects in these new forms of HR data. On the one hand, organizational entities are hierarchical structures which causes the effects of HRM to occur at and across different levels of analysis simultaneously (Hitt, Beamish, Jackson, & Mathieu, 2007; Wright & Nishii, 2007). On the other hand, as measurement happens on a more continuous basis, HR data structures often consist of many observations nested within subjects over prolonged periods of time (Angrave et al., 2016; Bersin, 2015). Acknowledging the above, scholars have been increasingly moving from GLM applied at a single level of analysis towards multi-level techniques (Boselie, Dietz, & Boon, 2005; Sanders, Cogin, & Bainbridge, 2014; Snape & Redman, 2010). However, the most commonly applied multi-level methods do not work well when examining bottom-up effects, linking individual phenomena to organizational-level outcomes, and they can become overly complex when examining multiple, potentially categorical, variables simultaneously and over prolonged periods of time.

This article proposes two statistical methods that are rarely applied to HRM research questions, despite having added value over and above more traditional methodology. First, bathtub models are proposed as a way to account for multi-level models where the outcome resides at the higher level of analysis. Outperforming the traditional approaches of aggregation and disaggregation (Bennink, 2014), bathtub modeling can add value to HRM research on, among others, group composition or bottom-up effects. Second, optimal matching analysis (OMA) is advocated for its ability to detect longitudinal patterns. It can reduce large volumes of both categorical and ordinal data into a smaller set of underlying trajectories. Although relatively unknown in the general HRM field, it has been a valuable tool for career pattern analysis (Dlouhy & Biemann, 2015). This paper aims to demonstrate the added value of each method to the HRM methodological toolbox by discussing their applicability to research on employee
engagement. After discussing each method’s strengths and weaknesses separately, the paper concludes with overview of potential future applications and synergies.

**Bathtub modeling and engagement**

Over the past two decades, the influence of HRM on organizational performance has received much scholarly attention (e.g. Becker & Gerhart, 1996; Paauwe, Guest, & Wright, 2013) and a major part of the impact of HRM policies and practices has been demonstrated to be indirect via the behavior of employees (Christian, Garza, & Slaughter, 2011; Harter, Schmidt, & Hayes, 2002; Jiang, Lepak, & Baer, 2012; Kehoe & Wright, 2013; Subramony, 2009). However, an investigation of this mediational process is complex as it involves measurements at various levels of analysis. The design of HRM policies as well as the implementation of HRM practices commonly occur at either an organizational, functional, departmental or group level whereas the behaviors and cognitions they seek to influence are located at the level of the individual employee (Bowen & Ostroff, 2004; Snape & Redman, 2010; Wright & Boswell, 2002; Wright & Nishii, 2007). Although multi-level techniques provide an avenue to test models with such hierarchical structures (e.g. Snijders & Bosker, 1999), they are developed for models where the outcome variable lies at the lower level of analysis. Problems arise when the outcome occurs at a macro-level (e.g., organizational performance) but the predictors reside at a micro-level (e.g., employee behaviors).

Scholars have creatively circumvented modeling such micro-macro processes. For example, studies relating HRM implementation and employee engagement to organizational performance have used three different approaches. Nevertheless, all have their downsides. First, the micro-level scores can be aggregated to the macro-level. As such, studies have investigated HRM, engagement and performance at the level of the organization or the work group (e.g. Harter et al., 2002; Jiang et al., 2012; Subramony, 2009; Whitman, Van Rooy, & Viswesvaran, 2010). However, after aggregation to the macro-level, the data loses all information on individual variations (Bennink, Croon, & Vermunt, 2013). To illustrate: an aggregated team score of average engagement can be interpreted either as all employees in the team being averagely engaged, or as the team being a mix of highly engaged and highly disengaged employees. Moreover, aggregation has considerable consequences because the reduction of the sample size – to the size of the macro-level sample – significantly decreases the power of the statistical test (Bennink et al., 2013; Krull & MacKinnon, 1999). As a second approach, scholars have restricted the analysis to micro-level variables. For example, studies have examined employees’ individual perceptions of HRM practices and its influence on employees’ engagement and individual performance scores (e.g. Christian et al., 2011;
Halbesleben, 2010). However, this approach introduces perspective bias as it looks at the impact of the perceptions employees have of the HRM strategies, policies and practices. Moreover, no conclusions can be drawn regarding the impact on the actual organizational performance as the models have to rely on micro-level outcomes, such as individual performance evaluations. Finally, disaggregation has been a third approach, in which the macro-level scores (i.e., HRM and organizational performance) are assigned to each micro-level case (i.e., employee). Luckily, this method is not frequently found in published academic studies as disaggregating scores violates the assumption of independent error terms (Keith, 2005), causing biased standard error estimates, overly liberal tests (Krull & MacKinnon, 1999) and artificially high power (Bennink et al., 2013).

Bathtub models provide a solution that overcomes the problems specific to the aforementioned approaches by correctly modeling the multi-level processes at hand. They offer an opportunity to investigate the relationship between macro-level variables through micro-level mechanisms. Using a latent variable model, a bathtub model raises micro-level responses to the macro-level while taking into account both within-group variance and sampling variability. Subsequently, the latent variable can be used as a regular predictor in the macro-level model. Applied to employee engagement, bathtub modeling provides an opportunity to examine how HRM policies and practices influence organizational performance through the behaviors of employees. This analysis can be conducted without ignoring the possibility that employee engagement is personal and individually determined (Macey & Schneider, 2008), and without inflating statistical power (Bennink et al., 2013, Krull & MacKinnon, 1999). The following illustrates the modeling process step-by-step. First the data requirements are described, followed by the two parts of the model and their respective interpretation. Afterwards, several limitations to bathtub modeling are presented.

Data requirements

Before conducting a bathtub model, the sample size and the data format need to be considered. The power of a bathtub model largely depends on the sample size at all levels of analysis. Hence, researchers should not only gather data of multiple groups but they should also ensure high response rates within the groups. Bennink (2014) demonstrates that a sample of as small as 40 groups with 10 respondents per group can already result in nearly unbiased parameter estimates in case of a bottom-up, micro-macro effect (e.g., engagement on organizational performance). In order to detect more complex effects like interactions or indirect effects, larger sample sizes at the macro-level are required. Simulations demonstrate that 200 groups with 10 respondents per group should be enough to detect most effects.
(Bennink, 2014), suggesting that increasing the macro-level sample size should be the focus for researchers seeking to examine small and/or complex effects.

After data collection, the bathtub modeling requires the data to be structured according to its multi-level nature. Depending on the statistical software used, the model can be estimated on a dataset with either a long or a wide format. For long datasets, each row would represent a micro-level case (e.g., employee) and one of the columns would identify the macro-level unit this micro-level case belongs to (e.g., team or work group). Subsequently, a bathtub model can be applied using a multi-level regression approach. For wide datasets, the format is more peculiar: each row needs to represent a macro-level unit whereas, for each micro-level case, each measurement should be stored in a separate column. This format is commonly referred to as the persons-as-variable approach and it does not work in all software packages (e.g., Mplus).

**Bathtub model**

A bathtub model consists of two parts: a measurement model and a structural model. The measurement model describes the relationship between the observed micro-level (individual) variable(s) and the macro-level latent variable(s) based on them. Because this part of the model is primarily concerned with raising the individual-level scores to the group-level, it is often referred to as the within-group part. Next, the structural model describes the relationship between the variables at the macro-level. By relating the latent variable, derived from the measurement part, to the other macro-level variables of interest, the bathtub model takes in account the individual variation that occurs at the micro-level. The below elaborates on both models in more detail.

The measurement part of a bathtub model uses a multi-level latent variable model to raise the observed individual data to the level of the group. Here, the format of the individual, micro-level data is quite important. The latent variable approach was initially proposed by Croon and Van Veldhoven (2007), who demonstrated how treating individual scores as exchangeable indicators for a continuous latent variable makes it possible to predict a group-level outcome with lower-level independent variables. It works by estimating an unobserved continuous score at the group-level based on the observed individual data. Because the resulting latent group score reflects the underlying individual data and its variance, it takes into account the measurement and sampling error that occurs at the micro-level. However, this approach was developed specifically for raising continuous variables to the group-level using a normally distributed latent variable and, while very useful, the approach has the limitation that it cannot treat categorical variables adequately.
Bennink, Croon, and Vermunt (2015) therefore propose an extension of this latent variable approach, which makes it possible to raise categorical variables to the group-level using a generalized latent variable modeling framework (Skrondal & Rabe-Hesketh, 2004). This extended model allows micro-level discrete variables to be raised to the macro-level using a categorical latent variable, called a latent class variable. Using this approach, the unobserved heterogeneity at group-level can be estimated using a latent class variable that clusters together groups that are more similar to each other. In this way, individual data can be raised to the group-level by creating clusters of macro-level groups (i.e., latent classes) based on the similarity of their micro-level scores. Although it is common practice to have as many latent classes at the macro-level as there are discrete categories at the micro-level (Bennink, 2014), the optimal number of classes may also be estimated based on fit measures like BIC, AIC or $\chi^2$. Moreover, the strength of the measurement model can be deducted using the entropy ($R^2$), with values above .70 reflecting a strong model where classes are adequately distinguished (Vermunt, 2010). This latent class approach is especially useful in the social sciences. For example, in HRM research, the effects of categorical variables are often of interest (e.g., gender and educational level) whereas employee behaviors are often measured with Likert-type items, which represent categorical, ordinal measures.

Once the measurement model is specified, the structural part of the model can be estimated. As the individual scores have been elevated in the measurement model, the structural model occurs completely at the macro-level and this is where the actual hypothesis testing takes place. Because it occurs on a single level, the interpretation of the structural model is comparable to that of a regular regression model, with direct, indirect and/or interaction effects depending on the specified model. The measurement scale of the dependent variable determines the type of regression model that applies (i.e., logistic, linear or ANOVA).

To illustrate the above, imagine a study examining the effect of a specific HRM practice, hours of leadership training received by managers, on the performance of teams. The researchers might want to examine whether a part of this effect is indirect, for instance, via the engagement of employees. The bathtub model that corresponds with such a study is presented in figure 1. Both the leadership training and the team performance occur at a macro-level and, therefore, their relationship can be estimated directly using a linear regression model at the group-level. In contrast, the engagement data is located at the level of the individual employee and thus requires elevation to the group-level before its involvement can be examined. For this elevation, researchers can choose either a continuous latent variable model or a latent class model. A continuous latent variable would imply that group-level engagement scores run from highly disengaged up to highly engaged groups, following a normal distribution. A linear regression
model could then be used to predict team performance, based on the leadership training and the continuous latent engagement scores. Alternatively, a latent class model could be applied. This would result in various macro-level classes and, subsequently, for each team, the probability that it belongs to one of the classes is estimated. For instance, one class could consist of teams where most individuals display an average level of engagement, whereas another class could be composed of teams with a few highly engaged and several disengaged team members. In contrast to an aggregation approach, the latent class model is able to differentiate between these two groups based on their different team dynamics. Afterwards, an ANCOVA model can be run at the macro-level, where team performance is explained by the leadership training and the discrete latent group engagement scores. Finally, based on the output test statistics, the study's hypotheses can be confirmed or rejected.

![Diagram of the mediation model](image)

**Figure 1:** A bathtub model of the mediating effect of employee engagement in the relationship between leadership training and team performance.

**Limitations of bathtub modeling**

Micro-macro analysis in general and bathtub modeling in particular are innovative and well-performing approaches to examine how micro-level mechanisms influence the relationship between macro-level variables. However, there are several limitations.
For example, it is unclear how sample size influence the performance of bathtub models. Although Bennink (2014) demonstrates that the method generally outperforms the traditional aggregation and disaggregation approaches, the more complex bathtub models with indirect or interaction effects can provide equally inaccurate estimates when applied to small macro-level samples. While a sample of 200 groups is usually sufficient to provide accurate results for these more complex models (Bennink, 2014), the minimum sample size to establish accurate results is unknown and seems to differ from one model to the next. Furthermore, to date, no research has been conducted regarding power in bathtub models with continuous latent variable.

Additionally, it is unknown how missing values impact the performance of bathtub models. The standard approach to handle missing values in multi-level research is listwise deletion but, to avoid information loss, multiple imputation is currently undergoing heavy development for the classical multi-level models examining top-down effects (Van Buuren, 2011). However, the impact of missing values and the optimal way to handle them has yet to receive empirical attention in multi-level research with bottom-up effects, such as bathtub models.

A further consideration lies in the person-as-variable approach, which considers the employees within a group as interchangeable. This approach implies that each individual within a group contributes equally to the estimation of the latent score for that group. For constructs like engagement, this seems theoretically sound as each employee’s engagement can be considered equally important to the team’s latent score. However, there may be HRM research questions in which the data of certain employees can be considered more important to or representative of a group’s latent score. For example, when accounting for differences in employment type (e.g., full-/part-time or temporary/fixed contracts) or when scores follow specific distributions within subgroups of employees (e.g., forced distributed performance evaluations).

Although bathtub modeling clearly outperforms traditional aggregation and disaggregation approaches (Bennink, 2014; Bennink et al., 2013, 2015; Croon & Van Veldhoven, 2007), multi-level structural equation models (SEM) may function as an alternative. However, similar to the traditional micro-macro analysis (Croon & Van Veldhoven, 2007), a multi-level SEM only works with continuous data and is unable to handle categorical predictors. Additionally, multi-level SEM on continuous data frequently provides more biased estimates of the bottom-up effects than the latent bathtub model advocated by this paper (Onrust, 2015). Nevertheless, for models where the outcome resides at the micro-level, multi-level SEM can be preferable (Lüdtke et al., 2008).
On a final note, bathtub models are very flexible and they run in most statistical software developed for latent variable modeling, including Mplus (Muthén & Muthén, 1998-2016) and Latent GOLD (Vermunt & Magidson, 2013). They can be conducted in R (R Core Team, 2016) as well, using the lavaan package (Rosseel, 2012), but this requires some technical expertise. Similar to SEM, latent bathtub models can be extended to include multiple variables: both continuous and discrete latent variables, with single or with multiple response variables at the micro-level (Bennink et al., 2015). More detailed statistical descriptions of the model can be found in Croon and van Veldhoven (2007) and Bennink and colleagues (2013, 2014) whereas a syntax for bathtub implementation in Mplus is provided by Bennink (2014).

**Optimal matching analysis and employee engagement**

The rapid development of HR technology has initiated a trend towards the continuous measurement of personnel behaviors and cognitions. Mobile applications, social network mapping, sociometric badges, wearables and continuous employee feedback systems are rendering more complex and longitudinal HR data (Angrave et al., 2016; Bersin, 2015). Employee engagement is one of the constructs which organizations are increasingly measuring on such an ongoing basis, potentially because research demonstrates it is less stable than previously assumed. Although the work engagement employees experience is often regarded as a stable state of mind with a dispositional element to it (Macey & Schneider, 2008; Schaufeli, Salanova, González-Romá, & Bakker, 2002), studies demonstrate that the explained variance among consecutive yearly measures ranges from a high 74% to a low 31% (Mauno, Kinnunen, & Ruokolainen, 2007; Schaufeli, Bakker, & Salanova, 2006; Seppälä et al., 2009). It seems that there is also a temporary, transient element to the construct, with employees reporting weekly and even daily fluctuations in their levels of engagement (Bakker & Bal, 2010; Llorens, Schaufeli, Bakker, & Salanova, 2007; Sonnentag, 2003).

In line with the above, researchers call for prolonged periods of observation with more frequent measurements in order to investigate engagement patterns. For example, Harter and colleagues (2002) conclude their meta-analysis of the Gallup engagement data calling for “longitudinal designs that study changes in employee satisfaction–engagement, the causes of such changes, and the resulting usefulness to the business future research” (p.276). Similarly, despite their relatively stable results, Seppälä and colleagues (2009) argue that “longer follow-up with several measurement points would also allow investigation of the developmental trajectories of work engagement; utilizing a person-oriented approach
would yield a more specific understanding of stability/change in work engagement than the conventional methods of the variable-centered approach” (p.478).

HRM researchers have been using methodology other than GLM to examine longitudinal patterns for quite some time (Sanders et al., 2014), but optimal matching analysis (OMA) remains relatively unknown in the field. OMA is a quantitative method originating from the natural sciences, where it has been especially useful in detecting temporal patterns. The method works by assessing the similarity among longitudinal sequences, after which an unsupervised learning algorithm can be used to group the sequences based on their similarity. The result is a categorical variable representing the patterns hidden in the longitudinal data, which can be an insightful classifier on its own whereas it may additionally function as a predictor or outcome in further analysis.

OMA has several unique characteristics that make it an addition to the current methodological toolbox of HRM researchers. Similar to other longitudinal methods, OMA requires only a small sample, and simulations demonstrate that results are still 95% accurate for samples as small as 50 employees (Dlouhy & Biemann, 2015). In contrast to other methods, OMA is a person-oriented method, meaning that the object under observation – the employee or team – is the focus of the analysis, rather than the variance in a specific independent variable (Abbot, 1988). Hence, OMA can examine patterns on multiple variables simultaneously, which is a more complex assignment for other longitudinal methods such as multi-level and latent growth models (Curran, Obeidat, & Losardo, 2010). For example, OMA could analyze employees’ patterns on multiple dimensions of engagement (e.g., vigor, dedication and absorption; Schaufeli et al., 2002) or on engagement data coupled with other constructs. Additionally, OMA functions particularly well with the new forms of HR data. OMA’s classification results improve with the number of nested observations (Dlouhy & Biemann, 2015) whereas other longitudinal methods would soon require higher order polynomials to model the observations over prolonged periods of time (Curran et al., 2010). Finally, OMA is a relatively easy method to implement and interpret. To illustrate these advantages, the following section elaborates on a hypothetical study of weekly employee engagement data, like those gathered by a mobile application. First, the data requirements are described and afterwards each of the model steps is explained. Finally, several limitations and alternatives are presented.

**Data requirements**

OMA works by comparing cases based on their temporal sequences. These sequences consist of a string of elements, each reflecting the temporal state of the case at a specific moment in time. At this stage of the analysis, OMA handles these elements as categorical labels, not recognizing any ordinal nature among
them. Although this seems a disadvantage, it allows OMA to discover patterns in data that does not necessarily have an underlying order. For example, employees’ trajectories across functions, locations or organizational units (i.e., nominal variables) can be examined simultaneously with their engagement or performance levels (i.e., ordinal variables). While an order among elements can be assigned at a later stage of the analysis, at this time, it suffices to label each state with a unique element (e.g., working in Finance, with high engagement and high performance = ‘A’). Once the elements are determined, OMA requires the input dataset to be transformed into a wide format, where each row represent a case and each column a measurement occasion, so that for each case a sequence of elements arises.

Missing values are common in longitudinal research but, compared to other methods, OMA handles them relatively easy. As long as no more than 30% of the elements within a sequence are missing, replacing missing values by an additional element (e.g., X or ?) will result in model performance that is nearly equal to that of a complete dataset (3% decreased classification accuracy in Dlouhy & Biemann, 2015). Although late joiners, attrition or other factors may cause sequences to have different lengths, OMA results will accurately account for this as long as the length of the shortest included sequence is at least 70% of that of the longest sequence. Nonetheless, specific sequences may contain more than 30% missing values. These sequences can either be deleted or, if missing values gather around the start or end of the sequences, the timeframe of observation can be shortened, though both approaches may introduce bias in the results. Alternatively, researchers can decide to impute the missing values. Despite potential collinearity between sequential and missing elements, gap closure by recursive imputation has been demonstrated to provide accurate estimates of the missing data points (Halpin, 2012).

Penalty costs

OMA determines the similarity between sequences based on the operations needed to align them – to make them similar. There are two types of operations that can be used to align sequences: indel and substitutions. Indel is short for the deletion or the insertion of an element somewhere in a temporal sequence whereas substitutions refer to the replacement of an element somewhere in a sequence by another element in that same place. Both operations need to be assigned a penalty cost to reflect the dissimilarity the operation corrected, and especially the ratio between their respective costs is important (Aisenbrey & Fasang, 2010; Biemann & Datta, 2014). By default, OMA uses a standard indel-substitution cost ratio of 1:2, meaning that the deletion and insertion of an element comes at the same penalty cost of a substitution.
This standard ratio implies that OMA views all underlying states as equally dissimilar. However, as discussed earlier, HRM research often uses ordinal variables which may cause certain states to be more similar to each other because of their place in the underlying order. In order to reflect this lesser dissimilarity, the penalty cost of substitutions between those states can be decreased. Such a decrease can be based solely on theoretical assumptions but, alternatively, the observed transitions in the actual data may function as a basis. In this latter case, the rationale is that the transition frequency between states provides information about the similarity of these states (Biemann & Datta, 2014).

To illustrate the above, assume three five-week sequences: E-E-E-E-E, representing a consistently engaged employee; N-N-E-E-E, representing an employee who went from neutral to engaged; and D-D-E-E-E, representing an employee who was disengaged for two weeks. Using the standard cost ratio (i.e., 1:2), either sequence can be changed into the other at a penalty cost of 4: either by two element deletions and two element insertions or by two substitutions. Alternatively, researchers could set theory-based substitution costs, penalizing transitions between states adjacent in the underlying order at a lower rate. For example, 1.75 for the adjacent levels of engagement whereas the standard 2 for more ‘distant’ levels. Subsequently, the penalty costs between the first and last sequence in the above example would still equal 4 whereas those between the first two sequences and the last two sequences would now amount to 3.5. Data-based substitution costs would have the same effect if transitions between distant ordinal states occur less frequent.

Although setting custom substitution costs thus seems sensible, the approach has two downsides. First, substitution costs are not sensitive to the direction of a transition. Hence, Aisenberg and Fasang (2010) argue that custom costs should only be used “if there is either a theoretical justification for the assumption that the costs are the same independent of the direction of the movement [...] or if one of the directions is impossible” (p.430). Phenomena like positive and negative spirals (Fredrickson & Joiner, 2002) may, for instance, cause transitions towards the extremes of the engagement continuum to be more frequent than the other way around. Second, data-based substitution costs reflect the within-sequence variability of elements and, therefore, the chosen timespan of elements has a strong influence. Applied to the engagement example, this becomes evident. Assuming engagement fluctuates over time, element transitions would occur relatively frequent when longer-spaced timespans, like yearly measurements are used. In contrast, element repetition would occur frequently in case of hourly observations. Both have consequences for the data-based substitution costs and, while neither necessarily deteriorates results, researchers should consider that an inter-dependency exists.
Clustering the sequences

Once the penalty costs are determined, the optimal matching algorithm can assess the (dis)similarity of each dyad of sequences by aligning them. Although there may be several ways to align two sequences, the algorithm seeks the one with the least penalty costs. As illustrated above, assigning custom substitution costs would thus make the algorithm more prone to use substitution. The process of sequence alignment is repeated for all dyads in the dataset and the resulting penalty costs are stored in a Euclidian distance matrix, referred to as the dissimilarity matrix.

Next, a classification algorithm can be applied to the dissimilarity matrix. Any unsupervised learning algorithm that handles Euclidian distance matrices can be used. However, Dlouhy and Biemann (2015) “do not recommend using k-means, median, centroid and single linkage clustering for OMA at all” (p.171). Out of the eight techniques they tested, Ward’s minimum variance method consistently performed best. Irrespective of the chosen algorithm, the result is a categorical variable where cases are assigned to a category based on the underlying patterns in their sequence. In the engagement example, employees will be grouped based on the patterns that have occurred in their weekly engagement levels. One can expect to find, for example, clusters of consistently disengaged, neutral and engaged employees. Similarly, other clusters may include employees whose engagement has been steadily rising or falling, whose engagement demonstrates certain cyclical patterns, or whose engagement fluctuates randomly. Depending on the assigned penalty costs and the number of clusters, employees with missing values will be added either to the regular clusters or to clusters with specific, recurring patterns of missing values.

OMA’s output can be valuable to researchers in four ways. First, OMA’s relatively simple implementation and interpretation makes it an effective tool to get descriptive insights in longitudinal data and the patterns stored within them. Second, OMA can be used for identification purposes. For example, using the cluster output, researchers can effectively identify which employees fall in a category of consistent disengagement and reach out with follow-up interviews or supportive interventions. Third, the output can be used as an independent variable in subsequent analysis to study the consequences of following a specific pattern. For example, the displayed engagement patterns could function as predictor in attrition models or in studies on the effects of emotions. Fourth and final, the cluster output can function as dependent variable for subsequent analysis to examine pattern occurrence. For example, individual differences or HRM practices may explain why certain employees are more likely to display particular engagement patterns.
Limitations of OMA

A general shortcoming of OMA is that it aims to summarize a database filled with potentially very complex sequential patterns into a (handful of) categorical variable(s). While this has proven useful in certain research fields – including research on DNA, life courses and careers – it has yet to be tested whether longitudinal measures of employee behaviors and cognitions can be similarly reduced to a set of patterns.

A second challenge relates to setting the substitution costs right. Several studies illustrate how this ratio should be matched to the specific requirements of the research question and the analysis (Hollister, 2009; Lesnard & Kan, 2011). Although standard and custom cost ratios have been argued to lead to similar conclusions (Biemann & Datta, 2014), one might question whether this holds when employee experiences are the focus of research as these are subjected to a wide variety of personal and institutional factors. Moreover, there are rightful concerns regarding the symmetrical nature of the substitution costs (Aisenbrey & Fasang, 2010) which implies that transitions between states are considered similar, irrespective of their direction. There seems to be no simple solution to the aforementioned issues, apart from some best practices in the cost setting procedure (e.g., Gauthier, Widmer, Bucher, & Notredame, 2009).

A third limitation lies in the descriptive nature of OMA. The method can reduce large information volumes into a smaller, workable sets of underlying patterns and complementary analysis may provide insights into why these patterns occur and what they result in. However, researchers seeking to test why, how and when patterns occur and trajectories develop may turn to other methods. Here, multi-level and latent growth models can be used to examine the rate of pattern development as well as its causes. Additionally, hazard and Markov models may uncover why and when transitions between states happen. Moreover, time series analysis could be applied to investigate reoccurring patterns and forecast the future state of employees.

Finally, the only software that currently provides a means for automated implementation of OMA is R (R Core Team, 2016). The TraMineR package (Gabaldinho, Ritschard, Müller, & Studer, 2011) contains functions that automate the process to a large extent and only minor specification and customized programming is required. Additionally, the package includes several visualization functions that facilitate the interpretation of the model’s output. However, getting accustomed to the R language and syntax can be effortful.
Discussion

Previous research has heavily relied on GLM to investigate HRM processes and their potential impact on performance. This paper proposes two statistical modeling techniques that, despite their novelty to the field, can be valuable additions to the methodological toolbox of HRM researchers and practitioners. Particularly in light of the growing need to justify, prioritize, and improve decision-making (Boudreau & Ramstad, 2007; Rasmussen & Ulrich, 2015; Ulrich & Dulebohn, 2015) and the new forms of HR data that arise due to technological developments (Angrave et al., 2016; Bersin, 2015). Using latent variables, bathtub models are put forward as the solution to examine multi-level mechanisms with outcomes at the team or organizational level without decreasing the sample size or neglecting the variation inherent in employees’ responses to HRM activities. Optimal matching analysis is proposed as particularly useful to examine the longitudinal patterns that occur in repeated observations over a prolonged timeframe. Research on employee engagement was used to illustrate how each method functions and how they add value over and above the current methods used in HRM research.

Although both bathtub modeling and OMA both elevate micro-level data to a macro-level, the two methods strongly vary in their purpose, in their complexity and in the expertise required to implement them. The application of OMA does not require deep statistical or conceptual knowledge and its non-parametric nature along with the pattern visualizations facilitate an easy interpretation. However, this simplicity is also reflected in the primarily descriptive insights the method provides. In contrast, the underlying equations as well as the output of bathtub models are harder to explain to laymen such as business and HR professionals (see Bennink et al., 2013, 2014; Croon & Van Veldhoven, 2007), which may increase the difficulty researchers experience in translating the latent variable model’s results into actionable insights for decision-makers.

Nevertheless, both techniques can add value to HRM research on a variety of themes, either applied separately or in synergy with each other and traditional methodology. Recruitment and selection is one field of potential future application. OMA has been frequently applied on career patterns (e.g., Blair-Loy, 1999) and, similarly, the method could be valuable for selection purposes by clustering applicants based on their prior work experiences. For example, applicants’ historic job positions can be coded into unique states based on the associated management responsibilities or the required level of technical expertise. This information could be extracted directly from applicants’ résumés but the digital job market is becoming an ever richer alternative source of data (e.g., LinkedIn, Xing, ResearchGate). The resulting clusters can facilitate decision-making in the selection process alongside the applicants’
assessment center scores, interviewer ratings and other recruitment data. A latent bathtub model could use such recruitment data to examine the effectiveness of recruitment, selection and/or socialization practices. Furthermore, bathtub models could also investigate whether certain (combinations of) applicant profiles improve the effectiveness of teams.

The methods may additionally be valuable with regard to workforce planning, facility management and flexible working arrangements. Recent work by Lesnard and Kan (2011) demonstrates how a two-stage OMA can be used to cluster the daily work schedules of employees and subsequently use these clusters to unveil the patterns employees display in their weekly schedules. Using the data collected by sociometric badges and ‘smart’ workplaces, HRM scholars and practitioners could use OMA to uncover patterns in the use of office spaces over time and across locations. This may have direct practical value in terms of the cost reductions related to facility management, but may also be insightful for the design of flexible work arrangements. Moreover, certain configurations of work schedules within teams may have a detrimental impact on team effectiveness. This could be examined using the cluster output of OMA in subsequent bathtub models, raising the work schedules to a team-level and relating them to team performance.

Furthermore, talent management research may benefit from OMA and bathtubs. In line with literature on the Paretian performance distribution (O’Boyle & Aguinis, 2012), contemporary organizations often focus their attention on a small group of employees labeled with high leadership potential (so-called HYPO’s). OMA could model the history of job positions that distinguish such HYPO’s from other employees. Moreover, it could be examined whether receiving HYPO-status influences the developmental opportunities employees receive in the period that follows. Such analyses may provide valuable insights for the design of talent management policies and practices. With or without the OMA clusters, bathtub models could examine how talent management policies affect individual employees and, in turn, organizational performance. Moreover, bathtub models could examine which HRM practices stimulate the development of employees in general, and HYPO’s in specific, and whether this development contributes to the achievement of business goals. Finally, irrespective of HRM implementation, a latent variable model could be used to examine whether the presence of HYPO’s in a team influences team effectiveness.

Finally, the flexibility of latent variable models makes it a valuable tool to examine a wide variety of contemporary HRM themes. For example, latent approaches can model unobserved heterogeneity between respondents, which is valuable in HRM research on team diversity in terms of, for example, location, tenure, age, gender or cultural background. A more abstract example lies in the investigation of
team heterogeneity in terms of individual psychological contracts (e.g. Bakk, Tekle, & Vermunt, 2013). In the same way, the data of e-mail traffic or sociometric badges could be used to examine heterogeneity in terms of the personal networks employees have. Subsequently, the impact of such heterogeneity on the development, retention or performance of individuals, teams and organizations can be assessed. Combined with OMA, this latent heterogeneity of teams could be monitored over longer periods to assess whether improvement or other patterns occur, potentially following changes in HRM policies. As a final application, latent variable models can be used to grasp constructs that are otherwise hard to measure. For instance, a latent performance score could be estimated using multiple indicators of employee’s behavior or job output. This latent score has the potential to be a more accurately reflection of employees’ actual performance than the separate indicators or their combined average (Murphy, 2008).

In conclusion, bathtub models and OMA can add value on many HRM themes, allowing an investigation of research questions that were previously hard to examine. The methods can stimulate the quality of decision-making either by offering a different analytical approach to the issue at hand, or by working in synergy with each other and more traditional HRM methodology. Bathtub modeling and OMA are just two examples of methods that are common in other research fields and may benefit the HRM community. We hope that, by demonstrating how these specific methods add value in analyzing the new forms of HR data, HRM researchers and practitioners become more open to methodological developments in- and outside the field of HRM. In the end, alternative methodology can offer a different perspective and facilitate improved, more objective decision-making regarding human capital.

References


