PSYCHOMETRIKA—VOL. 81, NO. 2, 411–433
JUNE 2016
DOI: 10.1007/s11336-014-9433-x

# KSC-N: CLUSTERING OF HIERARCHICAL TIME PROFILE DATA

JOKE HEYLEN, IVEN VAN MECHELEN, PHILIPPE VERDUYN, AND EVA CEULEMANS

UNIVERSITY OF LEUVEN

Quite a few studies in the behavioral sciences result in hierarchical time profile data, with a number of time profiles being measured for each person under study. Associated research questions often focus on individual differences in profile repertoire, that is, differences between persons in the number and the nature of profile shapes that show up for each person. In this paper, we introduce a new method, called KSC-N, that parsimoniously captures such differences while neatly disentangling variability in shape and amplitude. KSC-N induces a few person clusters from the data and derives for each person cluster the types of profile shape that occur most for the persons in that cluster. An algorithm for fitting KSC-N is proposed and evaluated in a simulation study. Finally, the new method is applied to emotional intensity profile data.

Key words: hierarchical data, time profiles, shape and amplitude variability, individual differences, clustering, KSC.

## 1. Introduction

In many research domains, there is an increasing interest in how specific variables change across time, and in how this differs across persons or situations. For example, in emotion research, many authors argue that emotions are processes that unfold over time and thus can only be fully understood when their temporal properties are studied (see e.g., Kuppens, Stouten, & Mesquita, 2009). One such property is emotional intensity (Frijda, 2007). Inter- or intraindividual differences in the intensity course of particular emotional episodes are studied by asking persons to recollect and draw experienced emotion intensity profiles (Sonnemans & Frijda, 1994; Verduyn, Van Mechelen, & Frederix, 2012; Verduyn, Van Mechelen, Tuerlinckx, Meers, & Van Coillie, 2009). As illustrated in Figure 1, the resulting intensity profiles can vary hugely with respect to shape (e.g., steepness of onset, skewness) and amplitude (i.e., the height of the profile). For instance, this figure shows (a) three early blooming profiles with a steep onset and a peak in the beginning of the episode, followed by a gradual return to baseline, and (b) a late blooming profile that steadily accumulates across time and reaches its peak in the second half of the episode only (see Heylen, Verduyn, Van Mechelen, & Ceulemans, 2014). Moreover, it is notable that the intensity of the early blooming emotional responses differs, in that the corresponding profiles have different amplitudes.

Frijda, a leading scholar in the field of emotion research, advanced charting and explaining shape and amplitude variability in such intensity profiles as a central task for emotion researchers (Frijda, 2007). Studying intensity profiles is necessary from both a theoretical and a clinical perspective. Indeed, many psychopathologies such as depression, post-traumatic stress disorder, and borderline personality disorder are characterized by disturbances in emotion dynamics (Ehlers & Clark, 2000; Lieb, Zanarini, Schmahl, Linehan, & Bohus, 2004).

To model shape and amplitude variability in time profiles, researchers often rely on functional principal component analysis (fPCA; Ramsay & Silverman, 2005; Verduyn et al., 2009).

Correspondence should be made to Joke Heylen, Research Group of Methodology of Educational Sciences, University of Leuven, Tiensestraat 102, 3000 Leuven, Belgium. Email: Joke.Heylen@ppw.kuleuven.be

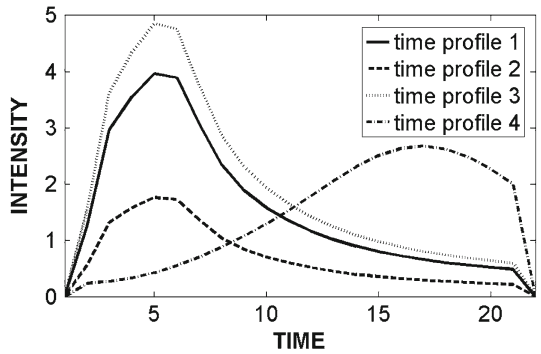


FIGURE 1.
Intensity profiles of four hypothetical emotional episodes (see Section 2 for more details).

This method approximates the observed time profiles by summing the average time profile and a linear combination of reference profiles (i.e., the extracted components). However, a major drawback of this approach is that shape and amplitude variability are entangled, which hampers the interpretation of the reference profiles.

A recently developed clustering technique, called K-spectral centroid clustering (KSC, Yang & Leskovec, 2011), is more appealing, as it allows to pull apart shape and amplitude differences. KSC assigns the observed time profiles to one of $K$ profile clusters. Each profile cluster is associated with a reference profile that captures a distinct profile shape that is present in the data (e.g., an early blooming profile). Moreover, all time profiles receive an amplitude coefficient which indicates the extent to which the reference profile needs to be inflated or deflated to optimally approximate the observed time profile. The usefulness of this technique for understanding the variability in emotional intensity profiles was demonstrated by Heylen et al. (2014).

KSC targets standard two-way two-mode data. Yet, data structures may be more intricate. For instance, in the emotional intensity example, researchers often collect multiple intensity profiles for each person. The resulting data have a hierarchical structure with intensity profiles (level 1) being nested in persons (level 2). Such hierarchical data are very rich and allow one to address research questions regarding individual differences in profile repertoire. In particular, hierarchical data allow to examine whether persons differ with respect to the number and/or nature of intensity shapes that show up in their emotional life, and whether such differences are related to measures of psychopathology and psychological well-being. Regarding number, some persons may adapt their emotional responses to the eliciting event, whereas others always react in the same way. Regarding nature, some people's emotional life may be dominated by short-lived emotional outbursts, whereas others may have difficulty shutting down an emotional response once it is elicited (Davidson, 1998).

To directly model such individual differences in profile use, we propose a novel KSC method for nested data, called KSC-N. This method simultaneously assigns the level 2 observations to a few clusters and fits a separate KSC model to each of those clusters. Thus, in the case of emotional intensity profiles, KSC-N induces a few person clusters from the data and derives per person cluster the types of profile shapes that occur most for the persons involved.

The remainder of this paper is organized as follows: In the next section, the KSC model for two-way two-mode data (Yang & Leskovec, 2011) is recapitulated, and the new model for nested data is introduced. In Section 3, we discuss the KSC-N loss function and an algorithm for estimating the model parameters. Next, a model selection procedure is proposed. Section 4 describes a simulation study to evaluate the performance of this algorithm. In Section 5, we apply

KSC-N to emotional intensity profiles. Finally, in Section 6, we provide some directions for future research.

## 2. Model

### 2.1. KSC

KSC was developed for $N$ (profiles) by $J$ (time points) data, where the $n$th profile is denoted by $\mathbf{x}_n$. In this paper, each profile corresponds to a particular emotional episode in the life of one of the persons under study. For now, we assume that each person provides one profile, reflecting a specific emotion type (e.g., joy, gratitude, anger, sadness) or situation. Of course, the duration of emotional episodes may vary considerably, from seconds to hours. To avoid that differences in duration rather than shape would drive the analysis results, duration differences are controlled for by stretching all profiles to equal length. Consequently, the time points should be interpreted in a relative way only. Note that in other fields of application (see Section 6), the time points may be more comparable, in that they reflect a fixed time scale measured in seconds, days, or years. Throughout this subsection we will make use of the hypothetical emotional intensity data set in Figure 1, which consists of four time profiles pertaining to 22 time points.

KSC groups the $N$ observed time profiles into $K$ clusters. Each cluster is associated with a reference profile, which reflects the prototypical shape of the time profiles that belong to that cluster. Moreover, amplitude differences within each cluster are modeled by assigning an amplitude coefficient to each time profile, which indicates the extent to which the associated reference profile needs to be inflated or deflated to optimally approximate the observed time profile.

More formally, the observed time profiles $\mathbf{x}_n$ are modeled as

$$\mathbf{x}_n = \sum_{k=1}^{K} p_{nk} f_n \mathbf{b}_k' + \mathbf{e}_n, \tag{1}$$

with $p_{nk}$ a binary partition score indicating to which of the $K$ clusters the $n$th profile belongs (where each profile belongs to a single cluster only), $f_n$ the amplitude coefficient for the $n$th profile, $\mathbf{b}_k$ the reference profile of cluster $k$, and $\mathbf{e}_n$ the residuals for the $n$th profile. To identify the obtained solutions, the reference profiles $\mathbf{b}_k$ are scaled to a norm of one. Note that Yang and Leskovec (2011) also align time profiles and reference profiles to remove differences in time shift. We discard this KSC-feature in this paper as potential time shifts between emotional intensity profiles are meaningful shape differences (i.e., indicating reactivity differences) that should be captured as such. This issue is further considered in Section 6.

To illustrate the interpretation of a KSC model, we will model the hypothetical emotional intensity data in Figure 1. The four time profiles can be perfectly reconstructed by means of a KSC model with two clusters. The two reference profiles are displayed in Figure 2. The clusters can be interpreted as 'early blooming' and 'late blooming', respectively. We can easily reconstruct the observed time profiles by combining these reference profiles with the partition and amplitude coefficients of the time profiles, shown in Table 1. For instance, time profile 1 is an early bloomer as it belongs to the first cluster; moreover, this reference profile needs to be multiplied by 9, according to the corresponding amplitude coefficient.

### 2.2. KSC-N

KSC-N assumes that multiple time profiles are gathered for $I$ persons. Indeed, in most studies on emotional intensity profiles, each person typically provides more than one profile, pertaining

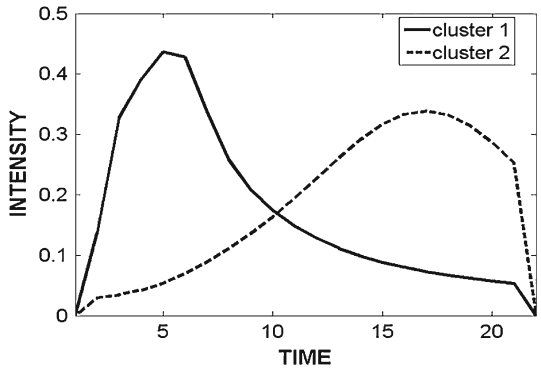


FIGURE 2.
Reference profiles of the KSC model with two clusters for the hypothetical data in Figure 1.

TABLE 1.
Partition scores $p_{nk}$ and amplitude coefficients $f_n$ of the KSC model with two clusters for the hypothetical data in Figure 1.

| Time profiles | Partition scores | | Amplitude coefficients |
|---|---|---|---|
| | Cluster 1 | Cluster 2 | |
| Profile 1 | 1 | 0 | 9 |
| Profile 2 | 1 | 0 | 4 |
| Profile 3 | 1 | 0 | 11 |
| Profile 4 | 0 | 1 | 8 |

to different emotion types and/or different situations. The number of person-specific profiles $N_i$ ($i = 1 \ldots I$) may vary across persons, and the total number of time profiles equals $N = \sum_{i=1}^{I} N_i$. The $n$th profile of the $i$th person is denoted by $\mathbf{x}_{n_i}$. The hypothetical data set in Figure 3, which consists of twelve time profiles nested into three persons, are used as a guiding example in the following paragraphs.

KSC-N simultaneously groups the persons into a few person clusters and models the time profiles within each person cluster with a separate KSC model. This implies that the profile clusters are nested in the person clusters and that the number of profile clusters may differ across person clusters. Thus, the $n$th time profile $\mathbf{x}_{n_i}$ of person $i$ is modeled as

$$\mathbf{x}_{n_i} = \sum_{c=1}^{C} p_{ic}^{\text{pers}} \sum_{k_c=1}^{K_c} p_{n_i k_c}^{\text{prof}} f_{n_i} \mathbf{b}'_{k_c} + \mathbf{e}_{n_i}, \tag{2}$$

with $C$ the number of person clusters, $p_{ic}^{\text{pers}}$ a binary partition score that indicates whether the $i$th person belongs to the $c$th person cluster, $K_c$ the number of profile clusters within person cluster $c$, $p_{n_i k_c}^{\text{prof}}$ a binary partition score that indicates whether the $n_i$th profile belongs to the $k_c$th profile cluster, $\mathbf{b}_{k_c}$ the $k_c$th reference profile of person cluster $c$, and $\mathbf{e}_{n_i}$ the residuals for the $n_i$th profile.

Table 2 and Figure 4 show the parameters of a KSC-N model that perfectly reconstructs the data in Figure 3, by means of two person clusters that are respectively associated with one and two profile clusters. From Table 2, it can be read that person C belongs to the first person cluster; Figure 4 reveals that this person always experiences a late blooming profile. Persons A and B are assigned to the second person cluster and report early blooming intensity profiles which have

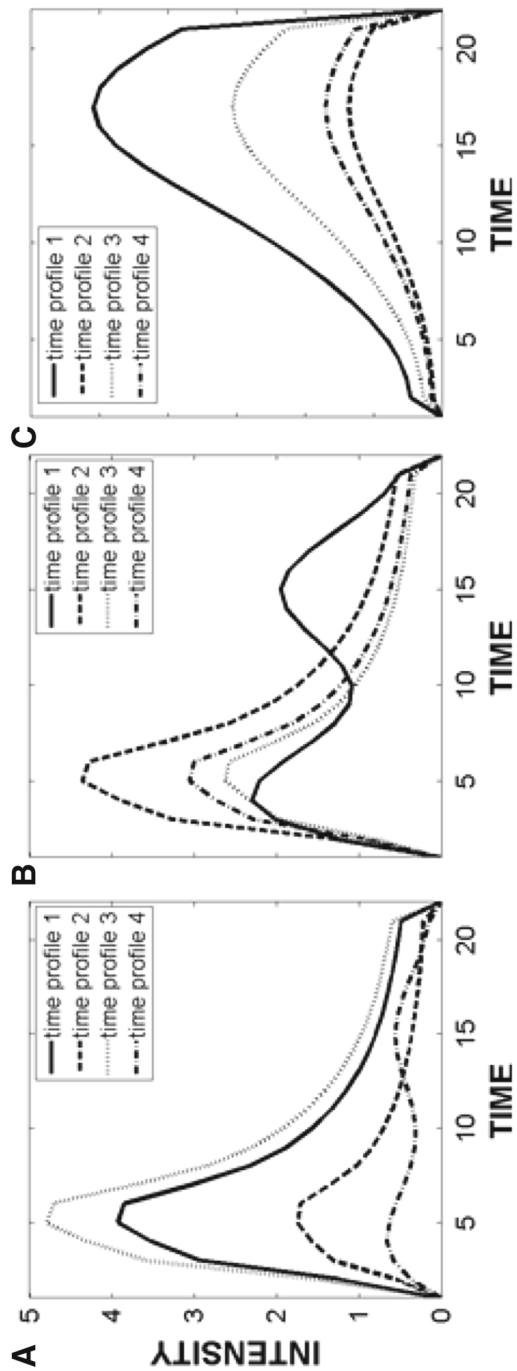


FIGURE 3.
Hypothetical intensity profiles of four different emotional episodes for three different persons (a, b, and c).

TABLE 2.
Partition scores $p_{ic}^{\text{pers}}$ and $p_{n_i k_c}^{\text{prof}}$ and amplitude coefficients $f_{n_i}$ of the KSC-N model with two person clusters, associated with one and two profile clusters, respectively, for the hypothetical data in Figure 3.

| Person | Time profiles | Person partition scores | | Profile partition scores | | | Amplitude coefficients |
|---|---|---|---|---|---|---|---|
| | | Cluster 1 | Cluster 2 | Cluster 1 | Cluster 2 | | |
| | | | | Profile cluster 1 | Profile cluster 1 | Profile cluster 2 | |
| Person A | Profile 1 | **0** | **1** | | 1 | 0 | 9 |
| | Profile 2 | | | | 1 | 0 | 4 |
| | Profile 3 | | | | 1 | 0 | 11 |
| | Profile 4 | | | | 0 | 1 | 2 |
| Person B | Profile 1 | **0** | **1** | | 0 | 1 | 7 |
| | Profile 2 | | | | 1 | 0 | 10 |
| | Profile 3 | | | | 1 | 0 | 6 |
| | Profile 4 | | | | 1 | 0 | 7 |
| Person C | Profile 1 | **1** | **0** | 1 | | | 15 |
| | Profile 2 | | | 1 | | | 4 |
| | Profile 3 | | | 1 | | | 9 |
| | Profile 4 | | | 1 | | | 5 |

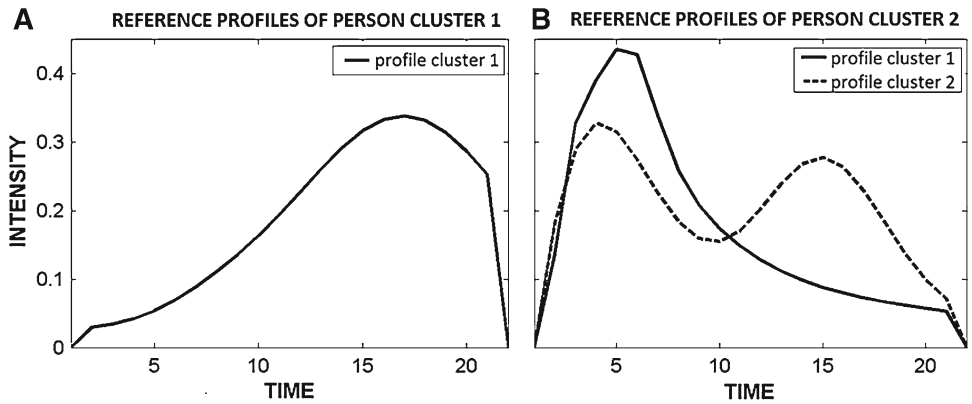


FIGURE 4.
Reference profiles of the KSC-N model with two person clusters, associated with one and two profile clusters respectively for the hypothetical data in Figure 3.

either a single peak or multiple peaks. The KSC-N model thus directly reveals that all profiles of person C have an identical shape, whereas persons A and B vary between two different shapes.

Of course, one could in principle model hierarchical time profile data with KSC. Indeed, a KSC-N model is a restricted KSC model with $\sum_{c=1}^{C} K_c$ clusters, in that the time profiles of a particular person can only be assigned to the profile clusters of the corresponding person cluster. In contrast, with KSC all profile clusters would be freely available for all persons. Hence, a KSC model with $\sum_{c=1}^{C} K_c$ clusters will always fit a given data set at least as well as a KSC-N model with the same number of profile clusters. However, because all profile types are available for each person, the obtained partition of the time profiles may be less clear-cut. Moreover, in practice,

one will often retain a more simple KSC model (i.e., with less clusters), because the reference profiles of the person clusters may differ subtly only, implying that a KSC model with 'average' reference profiles will fit reasonably well, or because some reference profile may show up for a few persons only, making it more difficult to trace.

### 2.3. Relation to Existing Models

In the literature, time profiles are often analyzed by model-based techniques, such as growth curve and trajectory models and multilevel extensions thereof (e.g., Jones & Nagin, 2007; Palardy & Vermunt, 2010; Wang & Bodner, 2007). These models differ in three important respects from the KSC and KSC-N approach: Firstly, whereas growth curve and trajectory models are probabilistic and thus imply distributional assumptions, KSC and KSC-N are deterministic techniques. Secondly, in growth curve and trajectory models, the reference profile of each cluster is modeled as a polynomial function of time, summarizing the reference profiles in terms of intercepts and slopes. KSC and KSC-N model the reference profiles through a parameter for each single time point, which is much more flexible, especially when the number of time points grows large. Thirdly, the intercepts and slopes of growth curve and trajectory models together determine both the predicted shape and amplitude of the time profiles, implying that these models entangle shape and amplitude again. In KSC and KSC-N, shape and amplitude are clearly distinguished and captured in separate parameter sets.

## 3. Data Analysis

### 3.1. Loss Function

For a given number of person clusters $C$ and associated numbers of profile types $K_c$, the aim of a KSC-N analysis is to find the person and profile partitions, amplitude coefficients $f_{n_i}$, and reference profiles $\mathbf{b}_{k_c}$ that minimize the following least squares loss function:

$$L_{\text{KSC-N}} = \sum_{i=1}^{I} \sum_{n_i=1}^{N_i} \left\| \mathbf{x}_{n_i} - \sum_{c=1}^{C} p_{ic}^{\text{pers}} \sum_{k_c=1}^{K_c} p_{n_i k_c}^{\text{prof}} f_{n_i} \mathbf{b}'_{k_c} \right\|^2 . \tag{3}$$

### 3.2. Algorithm

To minimize loss function (3), we propose to use an alternating least squares (ALS) algorithm that consists of five steps:

1. Initialize the partition of the persons: Randomly assign the $I$ persons to one of the $C$ person clusters. Each cluster has equal probability of being assigned to. No empty clusters are allowed.
2. Within each person cluster, estimate the profile partition and the reference profiles: For each person cluster, perform a KSC analysis on the time profiles of the persons that belong to the cluster, by conducting the following steps:
    a. Initialize the partition of the time profiles: Randomly assign the selected time profiles to one of the $K_c$ profile clusters, where each profile cluster is equally likely. Each profile cluster must contain at least one time profile of each person in the person cluster under study.

b. Estimate the reference profile $\mathbf{b}_{k_c}$ for each profile cluster: For each profile cluster $k_c = 1, \ldots, K_c$ vertically concatenate the $N_{k_c}$ profiles assigned to this cluster in a matrix $\mathbf{X}_{k_c}$. To make sure that each time profile has the same impact on the shape that is revealed, each time profile is separately scaled so that its norm equals one, yielding the rescaled matrix $\mathbf{X}^*_{k_c}$; note that, without scaling, profiles with larger norms will dominate the obtained shape. Conduct an eigenvalue decomposition on $\mathbf{X}^{*\prime}_{k_c}\mathbf{X}^*_{k_c} - (N_{k_c}\mathrm{ID}(J))$, where ID indicates an identity matrix. The eigenvector that corresponds to the largest eigenvalue is used as the estimate of the reference profile. Note that one may wonder whether it would be worthwhile to retain more than a single eigenvector per profile cluster and combine them in a weighted sum to approximate the data. Such an approach comes very close to performing a fPCA per cluster, however, and thus suffers from the same limitation in that shape and amplitude differences are entangled.

c. Estimate the partition of the time profiles: For each time profile and each profile cluster $k_c$, compute the optimal amplitude coefficient $f_{n_i}^{(k_c)} = \mathbf{x}_{n_i}\mathbf{b}_{k_c}$ and calculate the profile clustering criterion $\mathrm{L}_{n_i k_c} = \left|\left|\mathbf{x}_{n_i} - f_{n_i}^{(k_c)}\mathbf{b}'_{k_c}\right|\right|^2$ which quantifies the extent to which the time profile fits within the profile type. Assign each profile to the profile cluster for which $\mathrm{L}_{n_i,k_c}$ is minimal. Check for each person in the person cluster under study, whether all profile clusters contain at least one time profile of that person. If this is not the case, move the time profile of that person that fits its current profile cluster the least to the profile cluster without profiles of that person[1].

d. Estimate the reference profile $\mathbf{b}_{k_c}$ for each profile cluster, by executing Step b. Repeat Steps c and d until convergence is reached, that is, until the decrease in the KSC loss function $L_{\mathrm{KSC},c}$ for person cluster $c$ is smaller than $10^{-6}$.

$$L_{\mathrm{KSC,c}} = \sum_{i=1}^{I}\sum_{n_i=1}^{N_i} p_{ic}^{\mathrm{pers}} \left\| \mathbf{x}_{n_i} - \sum_{k_c=1}^{K_c} p_{n_i k_c}^{\mathrm{prof}} f_{n_i}\mathbf{b}'_{k_c} \right\|^2 \tag{4}$$

3. Estimate the partition of the persons: For each person $i$ and each person cluster $c$, optimally assign all time profiles to the $k_c$ profile clusters and compute the associated amplitude coefficients (see step 2c). Next, calculate the person clustering criterion $\mathrm{L}_{ic} = \sum_{n_i=1}^{N_i}\left\|\mathbf{x}_{n_i} - \sum_{k_c=1}^{K_c} p_{n_i k_c}^{\mathrm{prof}} f_{n_i}\mathbf{b}'_{k_c}\right\|^2$, which quantifies the extent to which the person fits within the person cluster. Assign the person to the person cluster for which this criterion is minimal. Check whether each of the person clusters contains at least one person. If this is not the case, move the person that fits its current cluster the least to the empty cluster.[2]
4. Within each person cluster, estimate the profile partition and the reference profiles, by executing Step 2.
5. Repeat Steps 3 and 4 until the algorithm has converged that is until the decrease in loss function $L_{\mathrm{KSC-N}}$ (3) is smaller than $10^{-6}$.

As generally holds for ALS algorithms (see e.g., Ceulemans, Van Mechelen, & Leenen, 2007; Steinley, 2003), the KSC-N algorithm may end in a local minimum. Therefore, we propose to use a multistart procedure in which the algorithm is run several times using different random initializations of the person partition. For the sake of efficiency, we suggest to execute steps 1

[1] Before moving this time profile, we check whether this time profile is the only profile in its current profile cluster. If this is the case we move on to the time profile that fits its current cluster the second least, and so on.

[2] Again, taking into account the sizes of the person clusters (see footnote 1).

and 2 of the algorithm for a large number of random initializations (e.g., 100 and preferably more, if computation time allows) and calculate the resulting loss function. Next, the 10 % most promising initializations (i.e., with lowest loss function values) are selected and run through the entire algorithm (see e.g., Wilderjans & Ceulemans, 2013). The solution with the lowest loss function value is retained as the final solution. Moreover, we recommend a similar multistart procedure for the KSC algorithm in Step 2, where a large number of random initializations (e.g., at least 100) are run through steps a to d.

### *3.3. Model Selection*

For most data sets, the optimal number of person clusters $C$ and/or profile types $K_c$ is unknown. To tackle the resulting model selection problem, one may estimate KSC-N solutions with 1 up to $C^{\max}$ person clusters and 1 up to $K^{\max}$ profile types, and retain the model with the optimal balance between fit (loss function value) and complexity (number of person clusters and profile types). However, the number of models from which to choose grows large rapidly. For instance, if both $C^{\max}$ and $K^{\max}$ equal five, a total of $\sum_{C=1}^{C\max} \frac{(K^{\max}+C-1)!}{(K^{\max}-1)!C!} = 251$ solutions have to be estimated. To limit computation time, we propose to use a stepwise approach, based on the work of De Roover, Ceulemans, Timmerman, Nezlek, and Onghena (2013); this procedure will be illustrated in Section 5.

The first step of the procedure consists of choosing among KSC-N solutions with an equal number of profile types $K_c$ ($K_c = K$) in each person cluster, implying that only $C^{\max}K^{\max}$ solutions have to be fitted. Specifically, we propose to plot for each number of person clusters $C = 1, \ldots, C^{\max}$ the least squares loss value of the model with $C$ person clusters and $K$ profile clusters against the number of profile clusters $K = 1, \ldots, K^{\max}$. The optimal number of person clusters $C^{\text{optimal}}$ is determined by examining the decrease in the loss value (over the different numbers of profile clusters $K$) when adding an extra person cluster and by retaining the number of person clusters $C$ after which this decrease levels off. Next, we focus on the scree line for $C^{\text{optimal}}$ person clusters and select the number of profile clusters after which the loss function values no longer decreases considerably as $K^{\text{optimal}}$.

Next, we determine the optimal number of profile types $K_c^{\text{optimal}}$ for each person cluster. To this end, we run KSC-N analyses with 1 to $K^{\max}$ profile clusters while keeping the person partition $p_{ic}^{\text{pers}}$ of the model that was retained in the first step fixed. Subsequently, we plot for each person cluster the least squares loss values $L_{\text{KSC,c}}$ (4) against the number of profile clusters $K_c = 1, \ldots, K^{\max}$ and choose the number of profile clusters after which the loss function value levels off as $K_c^{\text{optimal}}$.

## 4. Simulation Study

### *4.1. Problem*

In this simulation study, the KSC-N algorithm is evaluated with respect to sensitivity to local minima and goodness of recovery, under optimal conditions. With optimal conditions, we mean that the data are generated according to a KSC-N model with specific numbers of person clusters $C$ and profile types $K_c$, and were analyzed given these numbers. We examined the influence of six data characteristics on the performance of the algorithm: (1) number of person clusters, (2) number of profile clusters per person cluster, (3) differences in person cluster size, (4) differences in profile cluster size, (5) degree of congruence between the reference profiles, and (6) amount of error on the data. These data characteristics are of substantive interest and are therefore often investigated in simulation studies on clustering algorithms. Based on the results of these studies, we expect that results will be worse when the data contain more person and/or time profile clusters

(e.g., Brusco & Cradit, 2005; Milligan, Soon, & Sokol, 1983; Timmerman, Ceulemans, Kiers, & Vichi, 2010). Regarding differences in person and profile cluster size, our hypothesis is that the algorithm will perform best when person and profile clusters are of equal size (e.g., Brusco & Cradit, 2001; Milligan et al., 1983; Steinley, 2003). The degree of congruence between reference profiles was manipulated to study whether performance drops when the reference profiles of the different profile clusters are more similar (e.g., De Roover, Ceulemans, Timmerman, Vansteelandt, Stouten, & Onghena, 2012). Regarding amount of error, we expect that performance will be worse when the data contain more error (e.g., Brusco & Cradit, 2005).

Finally, to compare the output of KSC and KSC-N analyses, we also ran KSC analyses with $\sum_{c=1}^{C} K_c$ clusters. As stated earlier, these KSC analyses are less restrictive than the KSC-N ones, in that all profile clusters are freely available for all persons. Therefore, we hypothesize that KSC will be more susceptible to noise and will recover the profile partition and reference profiles worse than KSC-N.

### 4.2. *Design and Procedure*

In the simulation study, we fixed the following data characteristics: The number of time points $J$ was set to 20, the number of persons $I$ to 40, and the number of profiles per person $N_i$ to 15. These values were based on the empirical data that we will analyze in Section 5.

The six data characteristics introduced earlier were manipulated in a full factorial design:

1. The number of person clusters $C$ at 2 levels: 2 and 4;
2. The number of profile clusters per person cluster $K_c$ at 2 levels: either 2 profile clusters for each person cluster or 2 profile clusters for half of the person clusters and 3 profile clusters for the remaining half;
3. The differences in person cluster size at three levels (Milligan et al., 1983): equal (equal number of persons in each person cluster), unequal with majority (60 % of persons in one person cluster, the other ones equally distributed over the remaining person clusters), unequal with minority (10 % of persons in one person cluster, the other ones equally distributed over the remaining person clusters);
4. The differences in profile cluster size at three levels (Milligan et al., 1983): equal (for each person, each profile cluster contains an equal number of profiles), unequal with majority (for each person, 60 % of the profiles belong to one profile cluster, the remaining profiles are equally distributed over the remaining profile clusters), unequal with minority (for each person, 10 % of profiles belong to one profile cluster, the remaining profiles are equally distributed over the remaining profile clusters);
5. the degree of congruence between the reference profiles of all profile clusters, measured in terms of the Tucker congruence coefficient φ, at two levels: low (φ lower than 0.70; Haven & ten Berge, 1977) and high congruence (φ between 0.70 and 0.90);
6. the amount of error $e$, the expected proportion $\frac{\|\mathbf{E}^2\|}{\|\mathbf{X}^2\|}$ of the sum of the squared residuals in the residual matrix $\mathbf{E}$ and the sum of the squared observations in the data matrix $\mathbf{X}$, at two levels: 0.10 and 0.30.

For each cell of the design, five data matrices were generated, by constructing time profiles as follows:

$$\begin{aligned}\mathbf{x}_{n_i} &= \mathbf{t}_{n_i} + \mathbf{e}_{n_i} \\ &= f_{n_i}^{\text{true}} \mathbf{b}_{k_c}^{\text{true}\prime} + \mathbf{e}_{n_i},\end{aligned}$$

where $\mathbf{t}_{n_i}$ denotes the true time profile, $f_{n_i}^{\text{true}}$ denotes the true amplitude coefficient of time profile $n_i$ and $\mathbf{b}_{k_c}^{\text{true}}$ denotes the true reference profile of the profile cluster to which time profile $n_i$ is allocated.

The amplitude coefficients $f_{n_i}^{\text{true}}$ were randomly sampled from the normal distribution $N(50,100)$. The reference profiles $\mathbf{b}_{k_c}^{\text{true}}$ were generated as a weighted combination of three functions, with the weights being randomly drawn from the same uniform distribution and next rescaled so that they sum to 100. The functions were either a beta,[3] a lognormal, or a normal probability density function, where each of these three types has an equal probability of being chosen. The parameters of the beta functions were uniformly sampled between 1 and 10.5, the mean of the lognormals between ln(0) and ln($J$), and the parameters of the normals as well as the standard deviation of the lognormals between 0 and $J$. Next, the congruence between all pairs of resulting reference profiles was computed, and it was checked whether all congruence values lie within the specified bounds, i.e., between 0 and .70 in the low congruence conditions and between .70 and .90 for the high congruence conditions[4]; if not, new reference profiles were generated. Subsequently, the residuals $e_{n_i j}$ were generated such that the resulting data profiles $\mathbf{x}_{n_i}$ are non-negative, as this will always be the case for the emotional intensity profiles that we focus on in this paper. To this end, each residual $e_{n_i j}$ was sampled from a truncated normal distribution: $e_{n_i j} \sim N(\mu, \sigma_{n_i j}^2) \left| \mathrm{e}_{n_i j} \geq -t_{n_i j} \right.$, where $\mu$ equals 0 and $\sigma_{n_i j}^2$ is chosen such that $\sum_{i=1}^{I} \sum_{n_i=1}^{N_i} \sum_{j=1}^{J} E\left(e_{n_i j}^2\right) = \|\mathbf{T}\|^2 \frac{e}{1-e}$, with $\mathbf{T}$ the true data matrix, consisting of all true time profiles. Furthermore, the partition scores $p_{ic}^{\text{pers}}$ and $p_{n_i k_c}^{\text{prof}}$ are generated by randomly assigning the correct (i.e., according to the design) numbers of persons and profiles to the clusters.

In total, 2 (number of person clusters) × 2 (number of profile clusters per person cluster) × 3 (differences in person cluster size) × 3 (differences in profile cluster size) × 2 (congruence between reference profiles) × 2 (amount of error) × 5 (replicates) = 720 simulated data sets were generated. Each data set was analyzed with KSC-N, using the correct number of clusters $C$ and profile types per cluster $K_c$. 100 random initializations of the partition scores $p_{ic}^{\text{pers}}$ and $p_{n_i k_c}^{\text{prof}}$ were used, from which the most promising ten were selected (see Section 3). Finally, each data set was also analyzed with KSC, using $K = \sum_{c=1}^{C} K_c$ clusters and 100 random initializations of the profile partition scores $p_{nk}$.

### *4.3. Results*

*4.3.1. Sensitivity to Local Minima* To examine how sensitive the KSC-N algorithm is to local minima, we should compare the loss function value of each retained solution to that of the corresponding global minimum. If the latter value is lower than the former, the algorithm ended in a local minimum for sure. However, the global minima are unknown because the simulated data are perturbed with error (see e.g., De Roover et al., 2012). Therefore, we used the solution that is obtained when we seed the algorithm with the correct $p_{ic}^{\text{pers}}$ and $p_{n_i k_c}^{\text{prof}}$ values as surrogate for the optimal solution.

We conclude that the algorithm returned a local minimum for sure for 65 out of the 720 data sets (9 %). Out of these 65 data sets, 52 (80 %) had two profile clusters for half of the person clusters and 3 profile clusters for the remaining half, 48 (74 %) had four person clusters, and 46 (71 %) had an error amount of 0.30. Additionally, we examined attraction rates, by computing how many of the ten most promising random starts led to a solution of which the loss function differs less than $10^{-6}$ from that of the obtained solution. On average, this was the case for 4.8 of the starts. Moreover, we performed an analysis of variance with the attraction rate as dependent variable and the manipulated data characteristics as independent ones. Only considering the effects for which the partial eta-squared values $\eta_p^2$ exceed 20 %, we found sizeable main effects of the number of person clusters ($\eta_p^2 = .39$), the number of profile clusters per person cluster ($\eta_p^2 = .43$), the

[3] As the beta distribution is only defined on the interval [0 1], we evaluate the probability density function in $J$ points equally distributed between 0 and 1. These $J$ points form the time points of the reference profile.

[4] On average, the mean congruence amounts to.38 for low congruence data sets and.81 for high congruence data sets.

degree of congruence between reference profiles $\left(\eta_p^2 = .24\right)$, and the amount of error on the data ($\eta_p^2 = .47$). These effects imply that less random runs end in the obtained solution, when the underlying data are more complex, the reference profiles are more congruent, and the data contain more error (see Figure 5).

*4.3.2. Goodness of Recovery* The goodness of recovery of the obtained solutions will be evaluated with respect to (a) the person clustering, (b) the profile clustering, and (c) the reference profiles.

(a) Recovery of the person clustering

To study the goodness of recovery of the person clustering, we compute the Adjusted Rand Index (ARI; Hubert & Arabie, 1985) between the true and estimated person partition. This person ARI value equals zero when both partitions do not resemble each other more than expected by chance and amounts to one when they are identical.

The overall mean person ARI equals .98. Furthermore, for 674 out of the 720 data sets (94 %), a person ARI value of one is found. This means that the algorithm recovers the person clustering very well. Out of the 46 data sets for which the person ARI was lower than 1, all data sets had an amount of error of .30, and 41 (89 %) had highly congruent reference profiles. To examine the person clustering in these 46 data sets more in detail, we also computed the proportion of persons that were correctly clustered, after permuting the estimated person and profile clusters so that their reference profiles are maximally congruent with the true reference profiles (see below). Figure 6 shows the distribution of the resulting proportions, which is very negatively skewed.

(b) Recovery of the profile clustering

Since the profiles of a person cannot be allocated to the correct profile cluster if the person itself was clustered wrongly, we only take the profiles of the correctly clustered persons into account when computing a profile ARI between the true and estimated profile clustering. Thus, for the 674 data sets with a person ARI value of one, all 600 time profiles are considered. For the remaining 46 data sets, only the clustering of the time profiles of the correctly assigned persons is inspected.

We found an overall mean profile ARI of .90. To investigate the extent to which the manipulated data characteristics influence the profile ARI, we performed an ANOVA with the profile ARI as dependent variable. We found sizeable main effects of the degree of congruence between reference profiles $\left(\eta_p^2 = .51\right)$ and the amount of error on the data $\left(\eta_p^2 = .61\right)$, and an interaction effect of both these data characteristics $\left(\eta_p^2 = .45\right)$. These effects imply that the profile clustering is recovered worse when the data contain more error or more congruent reference profiles, and that the congruence effect is stronger in case the error level is higher (see Figure 7).

To evaluate the quality of the profile clustering that results from the KSC analysis of the data sets, we again computed ARI values, once taking all 600 time profiles into account and once considering only the time profiles of the persons that were correctly clustered in the KSC-N analysis. In the former case, the overall mean KSC profile ARI equals .76, in the latter .77. Thus, as hypothesized, the KSC-N analyses succeed better in recovering the true profile partitioning, due to the restriction that the time profiles of specific persons can only be assigned to the profile types of the person cluster to which a person belongs. To examine whether this performance difference between KSC-N and KSC depends on the manipulated data characteristics, we conducted an ANOVA with the difference between the KSC and KSC-N ARI (for the time profiles of the correctly clustered persons) as the dependent variable. Sizeable main effects of the degree of congruence between reference profiles $\left(\eta_p^2 = .47\right)$ and the amount of error $\left(\eta_p^2 = .57\right)$ were

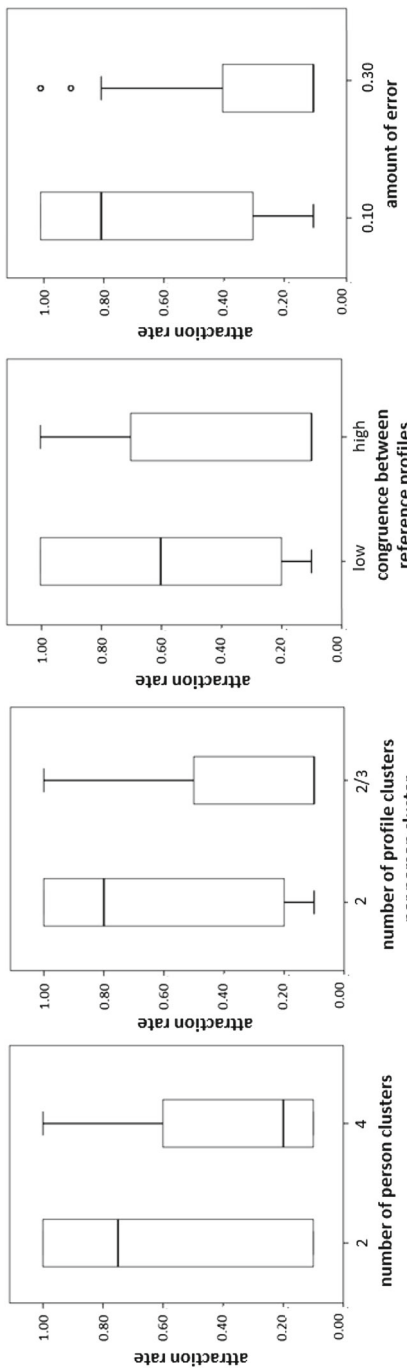


FIGURE 5.
Boxplots of the attraction rates as a function of the number of person clusters, the number of profile clusters per person cluster, the congruence between reference profiles, and the amount of error on the data in the simulation study.

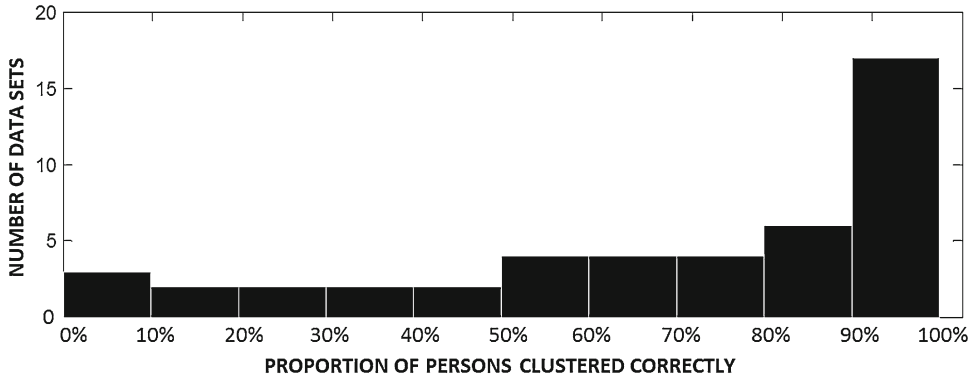


FIGURE 6.
Histogram of the proportion correctly clustered persons for data sets of the simulation study with an ARI value lower than 1.

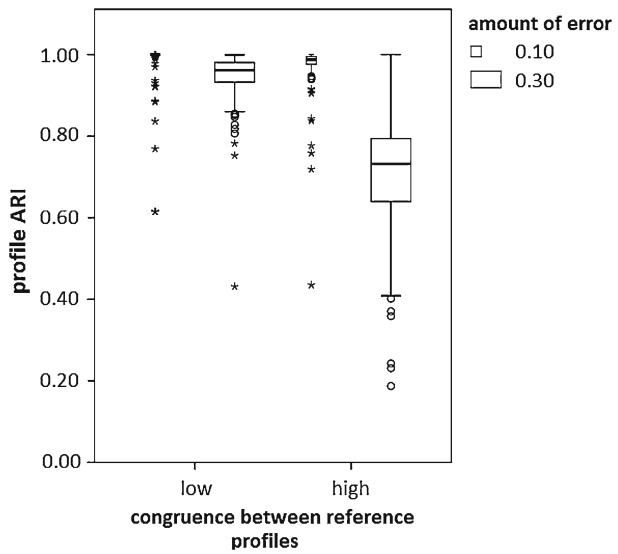


FIGURE 7.
Boxplots of the profile ARI as a function of the congruence between reference profiles and the amount of error on the data in the simulation study.

revealed and, additionally, an interaction effect of both characteristics $\left(\eta_p^2 = .23\right)$, which imply that KSC-N especially outperforms KSC when the data contain highly congruent reference profiles and a large amount of error (see Figure 8).

(c) Recovery of the reference profiles

To investigate how well the true reference profiles are recovered, we computed the Tucker congruence coefficient φ between each estimated reference profile $\mathbf{b}_{k_c}^{\mathrm{M}}$ and the corresponding true reference profile $\mathbf{b}_{k_c}^{\mathrm{T}}$ and averaged the resulting values across all profile clusters, yielding a goodnesss of reference profile (GORP) statistic: $\mathrm{GORP} = \frac{\sum_{c=1}^{C}\sum_{k_c=1}^{K_c}\varphi(\mathbf{b}_{k_c}^{\mathrm{M}},\mathbf{b}_{k_c}^{\mathrm{T}})}{\sum_{c=1}^{C}K_c}$. To take the permutational freedom of the person and profile clustering into account, we evaluated all possible permutations

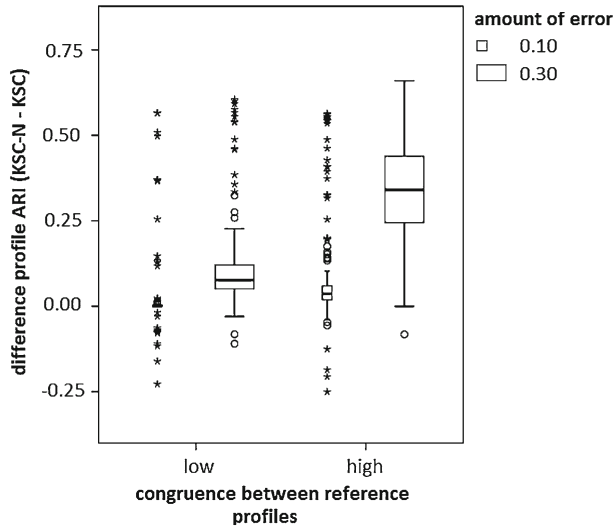


FIGURE 8.
Boxplots of the difference in profile ARI as a function of the congruence between reference profiles and the amount of error on the data in the simulation study.

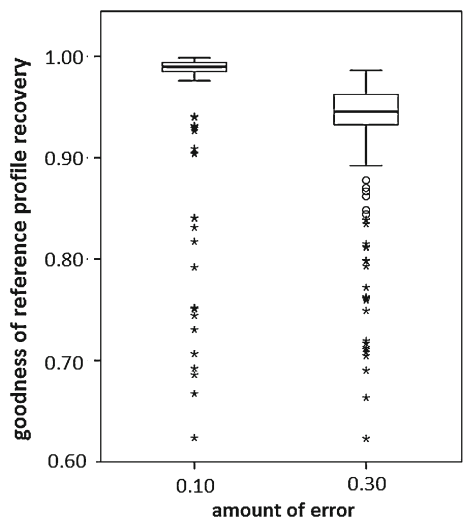


FIGURE 9.
Boxplots of the Goodness of reference profile (GORP) as a function of the amount of error on the data sets of the simulation study.

and retained the one that maximizes this GORP value. Note that the GORP statistic takes values between zero (no recovery at all) and one (perfect recovery).

The overall mean GORP value is .96, which indicates that the estimated and true reference profiles are very similar. We conducted an ANOVA with the GORP value as dependent variable and found a sizeable main effect of the amount of error on the data $\left(\eta_p^2 = .23\right)$. Specifically, Figure 9 shows that the GORP value is lower when the data are perturbed with a higher amount of error.

Finally, we also calculated the GORP values for the KSC analyses and found an overall mean GORP of .84, which is substantially lower than the overall mean for the KSC-N analyses. An ANOVA with the difference between the KSC-N and KSC GORP as dependent variable revealed no sizeable main or interaction effects.

### *4.4. Discussion*

For the simulated data under study, using 100 random initializations of $p_{ic}^{\text{pers}}$ and $p_{n_i k_c}^{\text{prof}}$ seems sufficient to avoid ending in a local minimum for 91 % of the data sets, given that the correct number of person clusters and profile clusters is used. Local minima typically occurred in the more difficult conditions of our design, for which attraction rates are also lower, in that less random starts yield the same solution. Therefore, when analyzing empirical data for which the optimal number of person and profile clusters is often unknown (see Section 3.3), one may adopt the following strategy: In the model selection phase, using 100 starts from which the ten best ones are retained seems sufficient, in that the obtained solution may not be the best possible one, but the loss function values, clustering, and reference profiles will closely resemble those of the best possible solution, as is also the case in the simulation study. Afterwards, once the optimal number of person and profile clusters has been determined, it may be wise to repeat the estimation of this solution using as many random starts as is computationally feasible to reduce the probability of obtaining a local minimum. Regarding recovery, performance deteriorates for data sets with a high amount of error or with highly congruent profiles. These findings are in line with our expectations. Regarding the relative performance of KSC-N and KSC in recovering the profile clusters and reference profiles, we conclude that KSC-N outperforms KSC, as expected, especially in data sets with a high amount of error and/or a high degree of congruence between reference profiles.

## 5. Application

A major challenge for emotion researchers is to systematically describe and explain shape and amplitude variability in emotional intensity profiles (Frijda, 2007). Heylen et al. (2014) demonstrated that KSC clustering is a useful method to identify the prototypical shapes of emotional intensity profiles. In particular, whereas some profiles resemble an early blooming profile (i.e., emotions which reach their peak immediately following stimulus exposure), other profiles take a late blooming shape (i.e., emotions that only reach their peak after a period of intensity accumulation). Moreover, whereas early blooming profiles were associated with the use of adaptive regulation strategies, such as reappraisal, late blooming profiles were related to maladaptive strategies, such as rumination. However, no insight was gained into possible individual differences in emotional profile repertoire.

Therefore, with this application, we aim to examine whether persons can be divided in a number of clusters, which are each associated with a distinct repertoire of emotional intensity profile shapes. A first possible difference between the repertoires pertains to the number of shapes. Whereas members of one person cluster may have multiple kinds of intensity profiles in their emotional repertoire, other clusters may be characterized by a rigid way of emotional responding. A second possible difference pertains to the nature of the shapes. Heylen et al. (2014) found that early and late blooming episodes are the two most prototypical intensity profile shapes, but it may be that new prototypical profiles will show up that are uniquely related to a particular person cluster.

In addition to charting such individual differences in the repertoire of emotional intensity profile shapes, we will examine whether these differences can be linked to measures of depression.

Many psychopathologies have been related to disturbed emotion dynamics; we focus on depression as it is highly prevalent and puts the strongest burden on society (World Health Organization, 2008). On the one hand, the depression literature is equivocal on elevated or reduced negative emotional reactivity in depressed persons (Thompson, Mata, Jaeggi, Buschkuehl, Jonides, & Gotlib, 2012), implying that it is unclear whether their emotion intensity profiles have an explosive start or not. On the other hand, researchers largely agree that negative emotions last longer for depressed people, as they show sustained processing of negative information (Siegle, Granholm, Ingram, & Matt, 2001), delayed amygdala recovery following exposure to negative stimuli (Siegle, Steinhauer, Thase, Stenger, & Carter, 2002), and resistance to change when experiencing negative affect (Kuppens, Allen, & Sheeber, 2010). Therefore, if the person clusters would differ in depression scores, we expect the cluster with elevated levels of depressive symptoms to have at least one reference profile that reflects sustained activity of negative emotion in that it lingers longer at peak intensity.

To study individual differences in emotional intensity profiles and their relation to depression, we collected data from a community sample of 69 persons who participated voluntarily. The sample consisted of 16 men and 53 women, with a mean age of 30.3 years (SD = 11.0). Participants reported on their daily experience of four different emotions—joy, gratitude, anger, sadness—during 14 consecutive days. In this application, we focus solely on negative emotions as these are more typically studied in the context of depression. Each evening, before going to bed, participants filled out an online questionnaire on the emotional episodes they experienced earlier that day, limited to three episodes per emotion type. Participants were asked to estimate the duration of these episodes by specifying the number of hours, minutes, and/or seconds the emotion had lasted. Subsequently, they were asked to draw as precisely as possible the intensity course of the emotional episode over time.[5] Similar to Heylen et al. (2014) and Verduyn et al. (2009, 2012), the intensity profiles were stored with a resolution of 448 pixels on the time-axis and 350 pixels on the intensity-axis, implying that the relative timing of the profiles was studied. During preprocessing, the number of time points was reduced to 28, by averaging blocks of 16 consecutive time points. These 28 time points were always preceded and followed by a time point with intensity value zero, which indicate the start and end of the episode. Additionally, on the first day of data collection, participants completed two depression questionnaires, namely the CES-D (Radloff, 1977) and DASS-21 (Lovibond & Lovibond, 1995).[6]

62 participants (i.e., 90 % of the participants) experienced one or more negative emotional episodes during the study period. We only included participants who reported at least five negative emotional episodes, as assigning persons to a cluster based upon less profiles would yield very unreliable results. This requirement was met for 43 participants (69 %) who contributed a total of 413 intensity profiles. The data therefore comprised 413 time profiles nested in 43 persons, with the number of profiles per person ranging from 5 to 22, and with a mean of 10 profiles per person.

To find an optimal KSC-N model for this data set, we used the model selection procedure described in Section 3.3. Specifically, we first analyzed the data with the KSC-N algorithm with $C$ varying from one to five and $K_c = K$ from one to four. For each $C$ and $K$ value, we used 100 random initializations of both the person and profile partition and retained the ten most promising ones. Figure 10a shows the loss function values of the obtained solutions. Regarding the optimal number of clusters $C$, we conclude that the largest gains in loss function value result from going from one to two and, less clearly, from two to three person clusters. Adding a fourth or fifth person cluster has a much smaller effect. Therefore, we decided to use two person clusters. Moreover,

[5] Participants also answered a number of appraisal and regulation questions regarding the emotional episode, but as we focus on individual differences these are not relevant for this study.

[6] Participants filled out additional trait questionnaires measuring fear and stress (DASS-21), self-esteem (JVGG), neuroticism and extraversion (NEO-FFI), life satisfaction (SWLS), and positive and negative affect (PANAS). Our focus, however, lies with depression

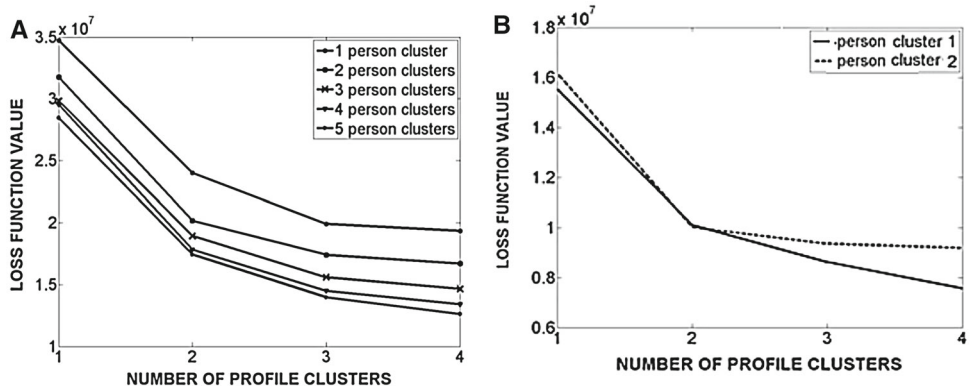


FIGURE 10.
(**a**) Loss function values $L_{\text{KSC}-\text{N}}$ for KSC-N solutions for the emotional intensity data with $C$ varying from one to five and $K_c = K$ varying from one to four. (**b**) Loss function values $L_{\text{KSC,c}}$ for KSC-N solutions for the emotional intensity data with two person clusters and with the number of profile clusters varying from one to four.

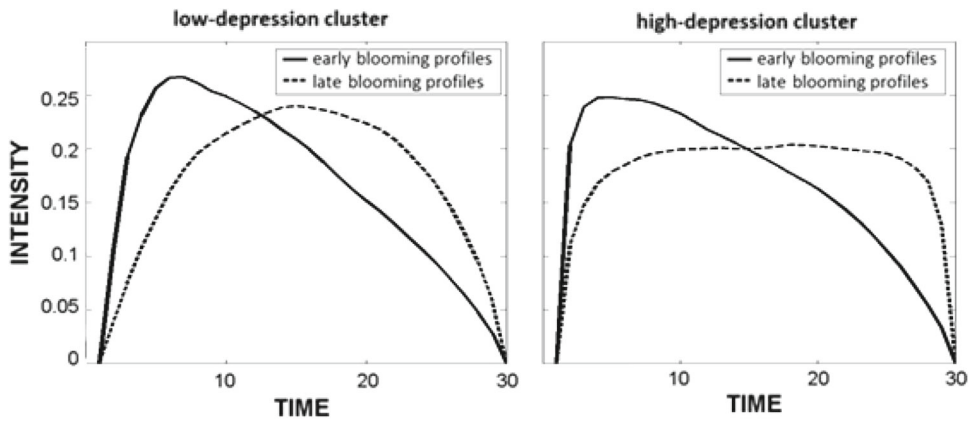


FIGURE 11.
Obtained reference profiles of the KSC-N model with two person clusters, each associated with two profile clusters for the emotional intensity data set.

as all scree lines display a clear elbow at two profile clusters, we retained the solution with two person clusters and two profile clusters for each person cluster as starting point for the next model selection step.

In this next step, we ran additional KSC-N analyses, again with 100 random initializations of the profile partition but keeping the person clustering obtained in the previous step fixed. In these analyses, the number of profile clusters was allowed to vary across the person clusters. Figure 10b plots the resulting loss function values $L_{\text{KSC},c}$ (4) against the number of profile clusters, for both person clusters. Based on this figure, retaining the model with two profile clusters per person cluster seems appropriate. Thus, we reran the corresponding KSC-N analysis, using 5,000 random starts in Step 1 and 500 in Step 2.a of the algorithm.

The two reference profiles of both person clusters are displayed in Figure 11. Out of the 43 persons, 22 are assigned to the first person cluster and 21 to the second person cluster. As the persons in the second cluster have significantly higher depression scores than those in the first [$F(1, 41) = 6.67$, $p = 0.014$ for CES-D and $F(1, 41) = 9.56$, $p = 0.004$ for DASS-21], we

will call the first person cluster the low-depression cluster and the second one the high-depression cluster. From Figure 11, we conclude that both person clusters contain a profile with a steep onset followed by a gradual return to baseline (i.e., early blooming profile). Moreover, both person clusters also contain a profile that reaches peak intensity later during the emotional episode (i.e., late blooming profile). The two person clusters differ in that (1) the onset of both profiles is steeper in the high-depression cluster and (2) the peak of the late blooming intensity profile lasts a very short time for the low-depression cluster, whereas this peak is sustained quite long in the high-depression person cluster.

From our analyses, we conclude that the obtained subgroups of participants do not differ with respect to the complexity of their emotional repertoire, as both subgroups are characterized by two possible shapes. Regarding the nature of the reference profiles, our results are in line with the literature on depression in that we find evidence for sustained activity of negative emotions in relatively more depressed people. Apart from that, with respect to emotional reactivity, clear differences were obtained, with more depressed people having a more steep onset of both the early blooming and the late blooming profiles.

To compare the KSC-N results with those that are obtained with KSC, we also applied KSC to the data set. We used 100 random initializations of the profile clustering, from which the most promising ten were selected. Based upon the scree plot in Figure 12a, we retained a solution with three clusters. The three reference profiles are displayed in Figure 12b and reveal an early blooming profile, a late blooming profile with short peak intensity, and a late blooming profile with long peak intensity. Thus, KSC succeeds in uncovering the major differences in profile shapes, but, as predicted in Section 2.3, more subtle differences, such as the difference in onset steepness between the two early blooming profiles of the KSC-N solution, remain hidden. In addition, unlike KSC, KSC-N includes distinct person-specific parameters (i.e., the person clustering) that may represent individual differences in an insightful way.

## 6. Discussion

We introduced the KSC-N method for clustering hierarchical time profile data. This method allows to easily capture individual differences in the number and nature of the profile shapes that occur across different episodes. In this paper, we focused on individual differences in emotional intensity profiles and found that the revealed differences were meaningfully related to depression, which illustrates that our method can be useful for studying emotion dynamics.

Of course, emotion research is only one of the possible fields of application of KSC-N. As another field, one can think of individual differences in the development trajectories of a variety of language-related skills. To study such differences from the perspective of specific language impairment (SLI) and dyslexia (see, e.g., Catts, Adlof, Hogan, & Weismer, 2005), one could administer multiple tests for measuring oral language and narrative skills at several points in time throughout kindergarten and primary school, to SLI children with and without a literacy delay and to typically developing children, and subsequently apply KSC-N to the resulting time profiles. Based on the results of, for instance, (Vandewalle, Boets, Boons, Ghesquière and Zink 2012), we expect that the analysis would reveal a first subgroup consisting of typically developing children for whom all tests show a steadily increasing development trajectory. A second subgroup might mainly consist of SLI children without a literacy deficit; these children can be expected to resemble the first subgroup on most tests but to show a different profile shape for text listening comprehension tests (with which they will struggle initially, whereas they will learn to handle them quite well later on). Finally, one may obtain a third subgroup comprising most SLI children with a literacy deficit, who can be expected to display a single developmental trajectory shape, in that they will only slightly improve on all tests.

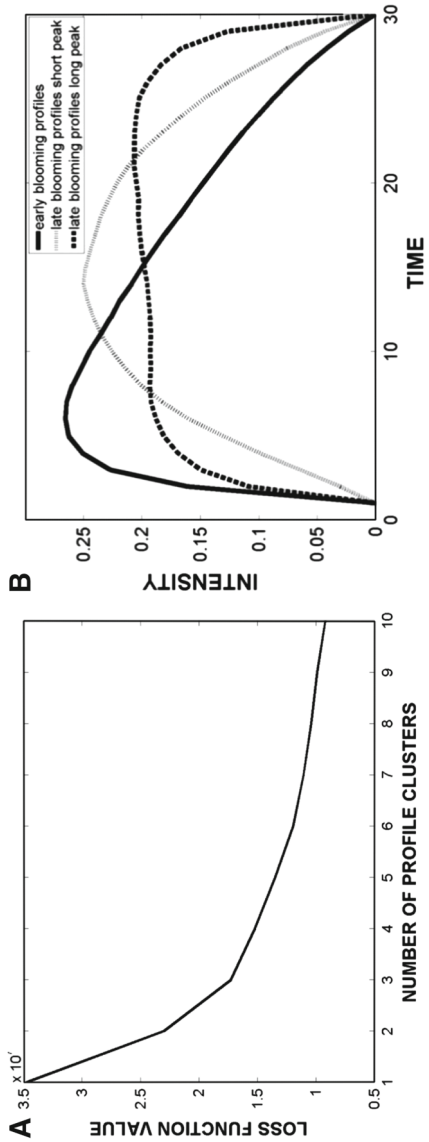


FIGURE 12.
(**a**) Loss function values $L_{\text{KSC}}$ for KSC solutions of the emotional intensity data with $K$ varying from 1 to 10. (**b**) Reference profiles of the KSC solution of the emotional intensity data with three clusters.

We see various directions for future research. First, KSC-N models the time profiles of each person cluster independently when extracting the profile types that describe the corresponding persons best. Although this allows to maximally trace possible differences across the person clusters, a drawback of this modeling choice is that it may hamper to detect whether or not some of the profile types occur for all persons and can therefore be considered common to all person clusters, whereas other profile types are distinctive in that they are only found in particular person clusters. To explicitly capture that some reference profiles are common and others distinctive, one could build on the work of De Roover, Timmerman, Mesquita, and Ceulemans (2013) and develop a KSC-N variant in which some of the profile types are shared by all person clusters.

Second, our analysis focuses on capturing shape differences in the time profiles. However, it can sometimes happen that some of the time profiles are shifted versions of one another (e.g., two profiles could be almost identical in shape, but one of them could set off one or more time points later than the other). In the context of emotion research, such shifts are meaningful and should be captured as shape differences. Yet, in other fields, researchers may not care about possible time shifts, implying that they have to be removed from the data. One possible way to do so would be to continuously align the time profiles and reference profiles throughout the KSC-N algorithm. Such a shifting option was implemented by Yang and Leskovec (2011) for the KSC approach.

Third, when using KSC-N, the reference profiles of the profile clusters are obtained by conducting an eigenvalue decomposition on the time profiles within each cluster and retaining the first eigenvector. As this first eigenvector is closest in least squares sense to all profiles that belong to the cluster under study, the obtained reference profile may be fairly strongly distorted by profiles with an outlying shape. Therefore, in future research, it may make sense to look for alternative estimation strategies that are less sensitive to possible outliers in the data. A promising approach in this regard might be to determine the reference profiles and amplitude coefficients in a robust way, by replacing the eigenvalue decompositions by robust PCA's (Hubert, Rousseeuw, & Vanden Branden, 2005).

Finally, it would be worthwhile to develop a stochastic variant of KSC-N (and KSC for that matter), using a mixture approach. In comparison to the deterministic clustering framework that underlies KSC-N, mixture modeling would entail a number of advantages such as estimates of classification uncertainty and confidence intervals for the reference profiles. Furthermore, in such a stochastic variant, one may opt to treat the amplitude scores as random effects. This implies that the amplitude scores no longer have to be estimated for each profile separately, which may help to avoid inconsistent estimates of the reference profiles (see Haberman, 1997, for a discussion of this topic in the context of item response modeling).

## Acknowledgments

This research was supported by Grant GOA/15/003 from the Research Fund of the University of Leuven. Philippe Verduyn is a Post-Doctoral Fellow of the Research Foundation—Flanders (FWO). The research leading to the results reported in this paper was supported in part by the Interuniversity Attraction Poles program financed by the Belgian government (IAP/P7/06).

## References

Brusco, M. J., & Cradit, J. D. (2001). A variable selection heuristic for K-means clustering. *Psychometrika*, *66*, 249–270.

Brusco, M. J., & Cradit, J. D. (2005). ConPar: A method for identifying groups of concordant subject proximity matrices for subsequent multidimensional scaling analyses. *Journal of Mathematical Psychology*, *49*, 142–154.

Catts, H. W., Adlof, S. M., Hogan, T., & Weismer, S. E. (2005). Are specific language impairment and dyslexia distinct disorders? *Journal of Speech, Language, and Hearing Research: JSLHR*, *48*, 1378.

Ceulemans, E., Van Mechelen, I., & Leenen, I. (2007). The local minima problem in hierarchical classes analysis: An evaluation of a simulated annealing algorithm and various multistart procedures. *Psychometrika*, *72*, 377–391.

Davidson, R. J. (1998). Affective style and affective disorders: Perspectives from affective neuroscience. *Cognition and Emotion*, *12*, 307–330.
De Roover, K., Ceulemans, E., Timmerman, M. E., Nezlek, J. B., & Onghena, P. (2013). Modeling differences in the dimensionality of multiblock data by means of clusterwise simultaneous component analysis. *Psychometrika*, *78*, 648–668.
De Roover, K., Ceulemans, E., Timmerman, M. E., Vansteelandt, K., Stouten, J., & Onghena, P. (2012). Clusterwise simultaneous component analysis for analyzing structural differences in multivariate multiblock data. *Psychological Methods*, *17*, 100–119.
De Roover, K., Timmerman, M. E., Mesquita, B., & Ceulemans, E. (2013). Common and cluster-specific simultaneous component analysis. *PLoS ONE*, *8*(e62280), 1–14.
Ehlers, A., & Clark, D. M. (2000). A cognitive model of posttraumatic stress disorder. *Behaviour Research and Therapy*, *38*, 319–345.
Frijda, N. H. (2007). *The laws of emotion*. Mahwah, NJ: Lawrence Erlbaum Associates Inc.
Haberman, S. J. (1997). Maximum likelihood estimates in exponential response models. *The Annals of Statistics*, *5*, 815–841.
Haven, S., & ten Berge, J. M. F. (1977). *Tucker's coefficient of congruence as a measure of factorial invariance: An empirical study*. Heymans Bulletin 290 EX, unpublished report by the Department of Psychology, University of Groningen.
Heylen, J., Verduyn, P., Van Mechelen, I., & Ceulemans, E. (2014). Variability in anger intensity profiles: Structure and predictive basis. Cognition & Emotion, 1–10. [Epub ahead of print]
Hubert, L., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*, 193–218.
Hubert, M., Rousseeuw, P. J., & Vanden Branden, K. (2005). ROBPCA: A new approach to robust principal component analysis. *Technometrics*, *47*, 64–79.
Jones, B. L., & Nagin, D. S. (2007). Advances in group-based trajectory modeling and an SAS procedure for estimating them. *Sociological Methods and Research*, *35*, 542–571.
Kuppens, P., Allen, N. B., & Sheeber, L. B. (2010). Emotional inertia and psychological maladjustment. *Psychological Science*, *21*, 984–991.
Kuppens, P., Stouten, J., & Mesquita, B. (2009). Individual differences in emotion components and dynamics: Introduction to the special issue. *Cognition and Emotion*, *23*, 1249–1258.
Lieb, K., Zanarini, M. C., Schmahl, C., Linehan, M. M., & Bohus, M. (2004). Borderline personality disorder. *The Lancet*, *364*, 453–461.
Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy*, *33*(3), 335–343.
Milligan, G. W., Soon, S. C., & Sokol, L. M. (1983). The effect of cluster size, dimensionality, and the number of clusters on recovery of true cluster structure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *5*, 40–47.
Palardy, G. J., & Vermunt, J. K. (2010). Multilevel growth mixture models for classifying groups. *Journal of Educational and Behavioral Statistics*, *35*, 532–565.
Radloff, L. S. (1977). The CES-D scale a self-report depression scale for research in the general population. *Applied Psychological Measurement*, *1*(3), 385–401.
Ramsay, J. O., & Silverman, B. W. (2005). *Functional data analysis*. New York, NY: Springer.
Siegle, G. J., Granholm, E., Ingram, R. E., & Matt, G. E. (2001). Pupillary and reaction time measures of sustained processing of negative information in depression. *Biological Psychiatry*, *49*, 624–636.
Siegle, G. J., Steinhauer, S. R., Thase, M. E., Stenger, V. A., & Carter, C. S. (2002). Can't shake that feeling: Event-related fMRI assessment of sustained amygdala activity in response to emotional information in depressed individuals. *Biological Psychiatry*, *51*, 693–707.
Sonnemans, J., & Frijda, N. H. (1994). The structure of subjective emotional intensity. *Cognition and Emotion*, *8*, 329–350.
Steinley, D. (2003). Local optima in K-means clustering: What you don't know may hurt you. *Psychological Methods*, *8*, 294–304.
Thompson, R. J., Mata, J., Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Gotlib, I. H. (2012). The everyday emotional experience of adults with major depressive disorder: Examining emotional instability, inertia, and reactivity. *Journal of Abnormal Psychology*, *121*, 819.
Timmerman, M. E., Ceulemans, E., Kiers, H. A. L., & Vichi, M. (2010). Factorial and reduced K-means reconsidered. *Computational Statistics and Data Analysis*, *54*, 1858–1871.
Vandewalle, E., Boets, B., Boons, T., Ghesquière, P., & Zink, I. (2012). Oral language and narrative skills in children with specific language impairment with and without literacy delay: A three-year longitudinal study. *Research in Developmental Disabilities*, *33*(6), 1857–1870.
Verduyn, P., Van Mechelen, I., & Frederix, E. (2012). Determinants of the shape of emotion intensity profiles. *Cognition and Emotion*, *26*, 1486–1495.
Verduyn, P., Van Mechelen, I., Tuerlinckx, F., Meers, K., & Van Coillie, H. (2009). Intensity profiles of emotional experience over time. *Cognition and Emotion*, *23*, 1427–1443.
Wang, M., & Bodner, T. E. (2007). Growth mixture modeling identifying and predicting unobserved subpopulations with longitudinal data. *Organizational Research Methods*, *10*, 635–656.
Wilderjans, T. F., & Ceulemans, E. (2013). Clusterwise Parafac to identify heterogeneity in three-way data. *Chemometrics and Intelligent Laboratory Systems*, *129*, 87–97.

World Health Organization (2008). *Integrating mental health into primary care: A global perspective*. World Health Organization.
Yang, J., & Leskovec, J. (2011). Patterns of temporal variation in online media. *Proceedings of the fourth ACM international conference on Web search and data mining*, pp. 177–186.

*Manuscript Received: 21 MAR 2014*
*Published Online Date: 10 DEC 2014*