Insight Into Individual Differences in Emotion Dynamics With Clustering

Anja F. Ernst[1] 🄳, Marieke E. Timmerman[1],
Bertus F. Jeronimus[1], and Casper J. Albers[1]

Abstract

Studying emotion dynamics through time series models is becoming increasingly popular in the social sciences. Across individuals, dynamics can be rather heterogeneous. To enable comparisons and generalizations of dynamics across groups of individuals, one needs sophisticated tools that express the essential similarities and differences. A way to proceed is to identify subgroups of people who are characterized by qualitatively similar emotion dynamics through dynamic clustering. So far, these methods assume equal generating processes for individuals per cluster. To avoid this overly restrictive assumption, we outline a probabilistic clustering approach based on a mixture model that clusters on individuals' vector autoregressive coefficients. We evaluate the performance of the method and compare it with a nonprobabilistic method in a simulation study. The usefulness of the methods is illustrated using 366 ecological momentary assessment time series with external measures of depression and anxiety.

In recent years, psychology has profited increasingly from the inclusion of temporal dynamics into its theories and assessment, for instance, with regard to the development of emotions and psychopathology over time and across places (Bos, Schoevers, & aan het Rot, 2015; Hamaker, Grasman, & Kamphuis, 2016; Jebb, Tay, Wang, & Huang, 2015; Koval, Kuppens, Allen, & Sheeber, 2012). This development and the emergence of ambulatory assessment (AA) techniques have propelled a focus on ideographic models. These models center on the dynamics of a single individual to fully express individual characteristics (Molenaar, 2004). This provides rich insight into the dynamic variation within an individual. The ideographic approach contrasts to the traditional, nomothetic approach, where general predictions about the population are derived based on the variation between individuals. Despite their popularity ideographic approaches also have disadvantages, especially being ineffective when studying large samples and lacking generalizability. Fitting a new model for every individual is unwieldy in a large sample. Furthermore, ideographic models often describe the dynamics of a single individual in great detail, but results cannot be generalized to other individuals. This is a severe limitation because psychological researchers typically strive to generalize or at least to assess how an individual compares with others, as is done with nomothetic approaches.

To introduce ideographic approaches for psychological assessment practices, notions of reliability and validity have to be considered. Reliability and validity have been firmly established for nomothetic methods but their application to idiographic methods received comparatively little consideration. Bridging ideographic and nomothetic approaches may enable the derivation of reliable, valid, and standardized assessment methods through normative and generalizable statistical models (Beltz, Wright, Sprague, & Molenaar, 2016; Wright & Zimmermann, 2019). For instance, the interpretation of intraindividual dynamics might be facilitated by standardized tools and comparisons with normative group data. Thus, the recent shift toward ideographic methods propels the need for integrative efforts of the intraperson and interperson perspectives.

To generalize dynamic processes researchers must draw inferences across levels of analysis, such as from the within-person level to the between-person level and vice versa. Often researchers assume that the between-person and within-person processes are equal, which would imply that studying the process at one level suffices to generalize to the other level. This so-called ergodicity assumption fails to

[1]University of Groningen, Groningen, Netherlands

Corresponding Author:
Anja F. Ernst, Faculty of Behavioural and Social Sciences, Department Psychometrics and Statistics, University of Groningen, Grote Kruisstraat 2/1, Groningen, 9712 TS. Netherlands.
Email: a.f.ernst@rug.nl

hold for most psychological processes, meaning the analysis at one level will yield different results than the analysis at another level (Molenaar, 2004). Some researchers presume that ergodicity would be strictly required to generalize across levels and, in the ubiquitous presence of nonergodicity for psychological processes, abandon generalizations altogether. This is a misconception (Adolf & Fried, 2019), because generalizing within-person dynamics to the between-person level is possible if one can achieve conditional ergodicity (Molenaar, 2004). Conditional ergodicity is met if within-person and between-person structures are equivalent after controlling for the unique factors of both structures, such as autoregression, time trends, or the existence of distinct subgroups (Voelkle, Brose, Schmiedek, & Lindenberger, 2014).

Similar to the concept of ergodicity, researchers investigating the items of multiple individuals over time would like to achieve measurement invariance (MI) across levels, with an absence of measurement differences (also called item bias) across individuals and time. Consequently, MI is a less restrictive condition than ergodicity. It requires only the measurement process to be equivalent and allows for heterogeneity with respect to the latent variables and their interrelations over individuals and time. The threat of unexplained heterogeneity, however, applies similarly to MI as it does to ergodicity; in a heterogeneous sample the intraperson and interperson differences may be too pronounced to achieve MI between levels (Adolf, Schuurman, Borkenau, Borsboom, & Dolan, 2014).

The quest into conditional inferences across levels of analysis requires researchers to identify the sources of heterogeneity at both levels of analysis. A possible reason for nonequivalence between levels could be the existence of subgroups at the between-individual level (Voelkle et al., 2014). Often empirical findings and the provided theoretical underpinnings thereof suggest the existence of subgroups of individuals that exhibit quantitative similarities in their dynamics. These subgroups are assumed to be characterized by individuals that resemble one another in their dynamism, while the subgroups mutually differ in their dynamics. It has been theorized, for example, that different age groups exhibit distinct affective responses to stress (Scott, Sliwinski, & Fields, 2013). Similarly, the dynamic relationships between somatic symptoms and positive and negative affect have been shown to differ in magnitude as well as in direction across individuals (Schenk, Bos, Slaets, de Jonge, & Rosmalen, 2017). For instance, the mutual dynamic relationships of positive and negative affect has been found to differ markedly between patients with Parkinson's disease (Shifren, Hooker, Wood, & Nesselroade, 1997). Shifren et al. (1997) note, "Thus, despite selecting a sample that was homogeneous with regard to disease type, functional abilities, and cognitive impairment, the results show heterogeneity for the structure and variation of mood."

(p. 336). Such findings suggest that differences in emotion dynamics cannot simply be accounted for by static covariates.

Some statistical approaches that bridge the ideographic and nomothetic approaches, like multilevel analysis, assume similar patterns of effects across the whole sample. While ideographic models circumvent the problem by allowing for heterogeneity in the structures of the models, this property often leads to overfitting. For example, Van der Krieke et al. (2017) used multilevel analysis and ideographic time series models to describe within-person and between-person variability. Results revealed large differences in dynamics between individuals. Neither of the two approaches was able to integrate information of individual dynamics with information on similar dynamics from other individuals, and to give an indication of why individuals' dynamics differed markedly.

Another approach to bridge ideographic and nomothetic perspectives uses likelihood ratio tests to assess reasons for nonequivalence between structures in multisubject time series (Voelkle et al., 2014). While this procedure allows to test whether equivalence holds across possible sources of heterogeneity, the researcher has to specify the sources of variance a priori. In many cases, the factors that cause nonequivalence between structures are not known. The quest to bridge ideographic and nomothetic approaches has thus motivated the search for meaningful subgroups to stratify the heterogeneous sample into homogenous subgroups.

To assess the dimensions of within-person and between-person differences, we propose an exploratory procedure that identifies subgroups of individuals who differ qualitatively in their within-individual dynamics: a dynamic clustering method. Clustering provides insight into the systematic patterns that cause nonequivalence and allows to draw inferences about both, commonalities and differences of the two levels. Identifying clusters in time series can enable us to distil the signature dynamics of different psychological processes from intensive longitudinal data. This will allow a more precise description of individual developments. Specifically, we propose a clustering method based on the most frequently employed time series model in the social sciences, the multivariate vector autoregressive (VAR) model of order p (cf. Lütkepohl, 2005). VAR parameters enable the description of multivariate dynamic relations in a format familiar to researchers in the field. In contrast to time series clustering procedures proposed so far, we will allow for within-cluster variation by employing a Gaussian finite mixture model to cluster individuals' VAR coefficients, thereby arriving at a probabilistic clustering.

We investigate the properties of our probabilistic clustering model in a simulation study and illustrate its application on an empirical example; the method's performance is contrasted to a nonprobabilistic alternating least squares VAR(1)

Table 1. Overview of Model-Based Time Series Clustering Methods for Psychological Data.

Paper	Probabilistic	Filtering method	Adaptive method	Variation within clusters
McNicholas and Murphy (2010)	+	−	+	−
Anderlucci and Viroli (2015)	+	−	+	−
Viroli (2011)	+	−	+	−
Michael and Melnykov (2016)	+	−	+	−
Ernst, Albers, Jeronimus, and Timmerman (2019)	+	−	+	−
Krone, Albers, Kuppens, and Timmerman (2018)	−	+	−	−
Bulteel et al. (2016)	−	+	+	−
Gates, Lane, Varangis, Giovanello, and Guskiewicz (2017)	−	+	−	±

Note. + Denotes a property is present, − a property is absent, and ± a property is present to a limited extent.

model (Bulteel, Tuerlinckx, Brose, & Ceulemans, 2016). R code for estimating the clustering methods is provided on the projects OSF page at (https://osf.io/sv2tg/?view_only=d 57284da4e5f413c8727e6f98a4291fa). Before we outline our dynamic clustering model, we begin with a short overview of clustering methods for longitudinal data.

Clustering Methods for Time Series Designs

Contemporary clustering methods are most commonly divided into two groups: *nonprobabilistic*[1] and *probabilistic* approaches. The former operates by minimization of a criterion, for instance, the sum of squares of distances, resulting in a nonprobabilistic assignment of individuals into clusters. The latter involves a statistical model, such as a mixture model that assumes data to arise from a mixture of distributions where each component corresponds to a cluster (McLachlan & Basford, 1988; McLachlan & Peel, 2004). At the heart of time series clustering methods lays the challenge of accounting for the dependence structure that is characteristic to dynamic data. Time series clustering approaches in other fields focus mainly on classification and use feature-based distance measures or compare the individual time points directly (for an interdisciplinary review, see Liao, 2005). The present study is focused on approaches that account for the temporal dependence through a statistical time series model. These approaches hold most promise for psychological research, because they provide a description of the cluster-wise dynamics in terms of interpretable model parameters which allows for generalizability of results. In these models, each cluster consists of time series described by similar model parameters.

Jacques and Preda (2014) contrast *adaptive* methods and *filtering* methods to cluster longitudinal data. Adaptive models combine the two crucial steps—accounting for dependency between time points and forming clusters—in a single estimation step, while filtering methods carry out both steps separately. Below, we review shortly the most prominent model-based time series clustering methods for

psychological AA data. A classification of these methods is listed in Table 1.

Recently, McNicholas and Murphy (2010), Viroli (2011), and Anderlucci and Viroli (2015) proposed *probabilistic, adaptive* approaches to time series clustering, where each cluster is modelled by an identical generalized autoregressive process. The covariance structure of these models depends on the length of the time series, rendering these models unsuited for time series with a high number of observations and/or unequal length. Recently, probabilistic, adaptive approaches have been proposed for long and unequal time series for the univariate (Michael & Melnykov, 2016) and multivariate case (Ernst et al., 2019).

Adaptive clustering methods offer many advantages over filtering methods. For instance, some are able to describe different clusters through different time series models (for details, see Ernst et al., 2019). These models, however, are very complex and their iterative estimation procedure makes them slow when studying large data sets and scenarios where various numbers of potential clusters and/or potential time series models are considered. As a result, the number of potential models that will have to be compared increases exponentially with the number of time series models and the number of clusters that are considered in model estimation. The use of filtering circumvents scalability problems. Most important, though these adaptive models lack random effects at the individual level. Therefore, these models give no indication of the within-cluster deviation beyond a single white noise error term. So, while elegant holistic models exist to cluster psychological time series, these models assume identical VAR coefficients for individuals of the same cluster. Therefore, the representation does not give indication of the individual deviations from the autoregressive cluster parameters which are highly plausible to exist in psychological AA data.

Bulteel et al. (2016) put a *nonprobabilistic, adaptive* clustering method forward that clusters multivariate, time series by representing clusters through identical VAR model parameters. That is, the method operates via an alternating

least squares method, minimizing the within-group sum of squares based on the VAR parameters of the cluster an individual is assigned to. This results in a *nonprobabilistic* clustering of individuals. This method does not account for individual variation within clusters.

S-GIMME, a *nonprobabilistic, filtering* approach, estimates temporal and contemporaneous associations between variables at either the population-, subgroup- (i.e., cluster-), or individual-level (Gates et al., 2017; Lane, Gates, Pike, Beltz, & Wright, 2019). Only effects that significantly improve the modification index at the respective level are estimated. S-GIMME is exceptional in that it accounts for individual differences by estimating individual-level effects. The estimation of associations at various levels distinguishes S-GIMME from the other time series clustering approaches discussed here that estimate the associations between all variables at the cluster-level. While providing a partitioning that takes individual differences into account, S-GIMME provides only limited indication of the within-cluster variation, especially with regard to the associations that are estimated at the population- or subgroup-level.

Krone et al. (2018) used a *nonprobabilistic, filtering* approach to cluster emotion time series. They employed the autoregressive Bayesian dynamic model as a filter and clustered the resulting parameters with a *k*-means algorithm.

To relax the rather strict assumption of equal dynamic processes within clusters that has been made (for at least some associations) by the dynamic clustering methods so far, we propose a *probabilistic, filtering* method. Unlike the methods reviewed above, this will account for and give indication of within-cluster deviation for all estimated dynamic associations.

Proposing a Probabilistic Clustering Procedure Using Finite Mixture Models

The probabilistic clustering method proposed in this article accounts for the variation in both, the individual dynamics and the individual mean levels, by clustering on both VAR slopes and intercepts. The method proceeds in a two-step fashion: (a) summarizing individuals' multivariate dynamics by use of the VAR(p) model and (b) probabilistic clustering of model parameters of Step 1, based on the Gaussian finite mixture model (McLachlan & Basford, 1988; McLachlan & Peel, 2004).

Summarize Individual Time Series. The VAR model of order p, VAR(p), for an individual j ($j = 1, \ldots, j$) is given by the following equation:

$$y_{j,t} = c_j + \left(\sum_{a=1}^{p} \Phi_{j,a} y_{j,t-a} \right) + u_{j,t} \tag{1}$$

$$u_{j,t} \sim i.d.d.N\left(0, \Sigma_{u_j}\right)$$

with $y_{j,t}$ the $m \times 1$ vector of observed variables for individual j at time point t ($t = 1, \ldots, T_j$), c_j the $m \times 1$ vector of VAR intercepts, $\Phi_{j,a}$ the $m \times m$ matrix of VAR slopes that describe the dynamic influence of the state of the m variables at time lag a on the state of the variables at time point t of individual j and $u_{j,t}$ the dynamic errors at time point t. The VAR model assumes equal distance between time points and relies on the assumption of stationary time series (Lütkepohl, 2005). For Gaussian time series, stationarity implies that the mean and variance–covariance of a series will stay constant.

For each individual's time series, the most parsimonious lag order (p_j) balanced with goodness of fit can be determined by model selection criteria such as the Bayesian information criterion (BIC; Schwarz, 1978). Based on this, the highest (plausible) lag to capture all the individual process in the sample can be selected (i.e., $p = max(p_j)$). We propose to subsequently estimate a VAR model of lag p for all individuals.[2] Because mixture model approaches to clustering are known to show disappointing performance in high-dimensional space (Bouveyron & Brunet-Saumard, 2014), and to avoid overfitting, overparametrization through an unnecessarily high lag order should be avoided. Similarly, only a moderate number of variables, say between two and seven, should be clustered using the approach proposed here.

Probabilistic Clustering. Once VAR intercepts and slopes are obtained for every individual the distribution of VAR coefficients can be estimated through a finite mixture model. Specifying probability distributions per cluster rather than using fixed values, we relax the assumption of identical parameters for individuals within a cluster as in the method by Bulteel et al. (2016). The probability density function of obtaining x_j from the population containing k component densities can be expressed through:

$$f\left(x_j; \theta\right) = \sum_{i=1}^{k} \pi_i f_i\left(x_j; \theta\right)$$

where $x_j = vec(c_j, \Phi_{j,1}, \ldots, \Phi_{j,p})$, a $m^2 \times p + m$ vector, θ represents the parameters of the k component densities, components have mixing proportions $\pi = (\pi_1, \ldots, \pi_k)^t$ where $\sum_{i=1}^{k} \pi_i = 1$ and each component density, f_i, is assumed multivariate Gaussian.[3] Hence, the parameters in θ are the mean vectors μ_i ($i = 1, \ldots, k$) and variance-covariance matrices $\Sigma_1, \ldots, \Sigma_k$. The vector μ_i contains of cluster i the mean VAR intercepts and the mean VAR slopes. The (co)variances of the individual parameters within cluster i around mean μ_i is indicated by covariance matrix Σ_i. The parameters of the complete model can be collected in the vector,

$$\Theta = \left(\pi^{\tau}, \theta^{\tau} \right)$$

The parameters are estimated via maximum likelihood. After $\widehat{\Theta}$ has been obtained, a maximum likelihood estimate of Θ, estimates for the posterior probabilities of cluster membership can be formed for each participant, yielding a probabilistic clustering of the n individuals. The probability that participant j with VAR(p) coefficients $\boldsymbol{x}_j$ belongs to cluster i is estimated by,

$$pr\left(participant\ j \in i | \boldsymbol{x}_j; \widehat{\Theta} \right) = \frac{\hat{\pi}_i f_i\left(\boldsymbol{x}_j; \hat{\theta} \right)}{\sum_{l=1}^{k} \hat{\pi}_l f_l(\boldsymbol{x}_j; \hat{\theta})}$$

with component densities f_i assumed to have the $N(\mu_i,\ \Sigma_k)$ distribution. A crisp partitioning of every of the n individuals into exactly one of the k clusters is achieved by assigning every participant to the cluster to which they have the highest posterior probability of belonging. The probabilistic clustering model proposed in this article will employ the expectation–maximization (EM) algorithm (Dempster, Laird, & Rubin, 1977) to achieve maximum likelihood estimate $\widehat{\Theta}$.

In the context of a finite mixture model, the selection of the number of clusters can be conceived of as a comparison between statistical models, where the different statistical models represent different choices for the number of clusters. Similarly, models with different constraints about the structure of the Σ_i's can be specified. This may be useful as such constraints can facilitate model estimation (cf. Fraley & Raftery, 2002; McNicholas, 2016; Steinley & Brusco, 2011). To this end, the number of clusters and the most parsimonious parametrization of the covariance matrix can be determined with the same probability model that is used to estimate the cluster parameters. To select the best fitting model, we employ the BIC (Schwarz, 1978).[4]

Autoregressive Representation of Time Series as Filter

The clustering based on filtered VAR coefficients as we propose here offers the following advantages: (a) Condensing the observed time series into model parameters before clustering reduces the high dimensionality that is characteristic for multivariate time series data. This problem, the curse of dimensionality (Bellman, 1957), can lead to highly biased estimates during the clustering process. VAR coefficients offer a reduction of dimensionality in which meaningful features are extracted from the data by use of the VAR model; (b) Clustering on VAR coefficients allows for comparison of time series of differing length; (c) VAR coefficients enable the description of multivariate dynamic relations in a format familiar to researchers in the field. That is, VAR coefficients

represent the dependence structure of a time series, and thus summarize the dynamics of individuals in a meaningful way.

For instance, emotional inertia, the propensity of affective states to resist to change, is often considered a key component of psychological well-being. This temporal dependency gives insight into a crucial and often neglected component of emotional well-being (Koval & Kuppens, 2012). High emotional inertia may be indicative of impaired emotion regulation, as a lack of emotion regulation may result in an inability to recover from negative emotions. It might also be indicative of a heightened preoccupation that results in emotional insensitivity and decreased involvement with the environment (Brose, Schmiedek, Koval, & Kuppens, 2015). Emotional inertia is traditionally operationalized as the autoregressive coefficient of an emotion (Koval et al., 2012; Krone et al., 2018). Because VAR coefficients are multivariate in nature, they provide insight into the dynamics of individual emotions, as well as the dynamics between the various emotions.

Sequential procedures to clustering, as outlined here, can be jeopardized when the reduction technique of the first step selects a summary of the data that is not discriminant of the different groups (Bouveyron & Brunet-Saumard, 2014). Our filtering method can be expected to be appropriate as long as VAR coefficients provide a proper indication of the underlying dynamics. An exception of this condition occurs, for instance, if the processes under study are not time-invariant, but meaningful changes in individuals' dynamic development occur during the course of the study, for instance, due to an intervention (see, e.g., Bringmann et al., 2017).

Also in other fields has the distance between time series parameters been widely applied as filter in nonprobabilistic clustering algorithms, especially the Euclidean distance between autoregressive coefficients, for instance, in combination with k-means or fuzzy C-medoids (D'Urso, De Giovanni, & Massari, 2015; D'Urso, De Giovanni, Massari, & Di Lallo, 2013). These measures are popular because of the absolute convergence of the autoregressive sequences of processes belonging to the admissible ARIMA (autoregressive integrated moving average) class. Thus, the distance between vectors of autoregressive parameters is a measure of structural dissimilarity between two time series processes. The distance indicates the cluster to which the dynamical structure of a series is closest.

Employing a Gaussian mixture distribution, we assume the individual VAR coefficients to be normally distributed around the mean of their associated cluster. The within-cluster covariance matrix can be freely estimated, which allows flexibility with regard to the orientation and distribution of the different clusters. The distance between maximum likelihood estimates of autoregressive parameters of two time series originating from a processes with the same dynamical structure will be asymptotically multivariate

normally distributed for stationary and nonstationary time series of any lag order (Corduas & Piccollo, 2008). Provided that time series within a cluster are generated by an identical dynamical process, VAR coefficients will thus asymptotically distribute normally around their cluster mean. This implies that model misspecification with regard to the number of clusters and/or a small sample size might thus cause the different cluster components to be nonnormally distributed. The Gaussian mixture model, however, has been shown largely robust to deviations (e.g., deviations caused by upper and lower limits or even uniform distribution within clusters (Steinley & Brusco, 2011). Because we assume the existence of clusters that are generated by at least structurally similar dynamic processes, we assume the within-cluster distribution of VAR coefficients will be well-captured by a normal distribution.

In the following, we will investigate the performance of our model-based clustering method in a simulation study and show its usefulness on an empirical ecological momentary assessment data set. The method shall be compared with an adaptive, nonprobabilistic clustering method (Bulteel et al., 2016).

Simulation

Simulation 1: Equal Intercepts Between Clusters

To assess the performance of the proposed probabilistic clustering method, we will conduct the following simulation. The performance of the probabilistic and nonprobabilistic method (Bulteel et al., 2016) will be compared with regard to recovery and optimization. Recovery is evaluated as the degrees to which the cluster membership and the cluster-specific VAR slopes can be retrieved; optimization is determined as sensitivity to local minima. We will investigate to what extent these two performance aspects depend on six data characteristics. Four of these factors pertain to the partition of individuals: (a) the number of clusters, (b) the relative cluster sizes, (c) the distance between clusters, and (d) the variation within clusters. Two factors reflect efficiency: (e) the number of individuals and (f) the number of time points (T_j). Regarding these factors, we have the following hypotheses: the performance of the clustering methods will decrease with (a) a higher number of clusters, (b) larger differences in relative size of the clusters, (c) lower distance between clusters, (d) higher variation within clusters, (e) fewer individuals, and (f) fewer time points. Also, we expect the probabilistic clustering method proposed in this article to outperform the nonprobabilistic clustering method as soon as variation within clusters is introduced, and to show a smaller vulnerability to differences in relative cluster size. All simulation and analysis were carried out in R (R Core Team, 2017). R code for estimating the clustering methods is provided in the supplementary material (available online) and on the projects OSF

page at https://osf.io/sv2tg/?view_only=d57284da4e5f413c8727e6f98a4291fa.

Design and Procedure

For every individual, a time series was generated according to Equation (1). Herewith, the following factors were kept constant: the variance–covariance matrix of the white noise series Σ_{u_j} was set to the identity matrix and intercepts, c_j, were equal to zero for all clusters. The latter implies that we simulate under a relatively difficult condition, because differences between clusters are captured in the VAR coefficient matrix only. We restricted our simulation to VAR processes of order 1, VAR(1). The number of variables was set to $m = 4$. To generate data the six factors introduced above were varied in a completely crossed design.

1. The number of clusters k: 2 or 4.
2. The relative size of the clusters: *Equal distribution* of participants across clusters; *minority condition*, with one cluster containing 10% of participants and the rest evenly distributed across the remaining clusters; *majority condition*, with one cluster containing 60% of individuals and the remaining individuals evenly distributed over the other cluster(s).[5]
3. The distance between the different clusters as expressed through distance between VAR(1) regression coefficients matrices of the different clusters. For every cluster in every condition, we simulated a matrix that contained the mean VAR(1) regression coefficients of the cluster. For each variable, all autoregressive and cross-regressive effects were generated with autoregressive coefficients on the diagonal and cross-regressive associations on the off-diagonal. In all conditions, autoregressive weights in the VAR coefficient matrix $\mathbf{\Phi}_i$ were drawn from a uniform distribution on the interval [.7, .9]. Cross-regressive weights in the VAR coefficient matrix were generated as follows: *small differences*, where for every cluster-level VAR coefficient matrix half the cross-regressive weights were sampled from a uniform distribution on the interval [.3, .5] and the remaining half from the interval [0, −.2]; *medium differences*, where for every cluster-level VAR coefficient matrix all cross-regressive elements of the coefficient matrix were sampled from a uniform distribution from the interval [.3, .5] and half these coefficients were randomly selected to be multiplied by −1; *large differences*, where for every cluster-level VAR coefficient matrix all cross-regressive elements of the coefficient matrix were sampled from a uniform distribution from the interval [.2, .7] and half these coefficients
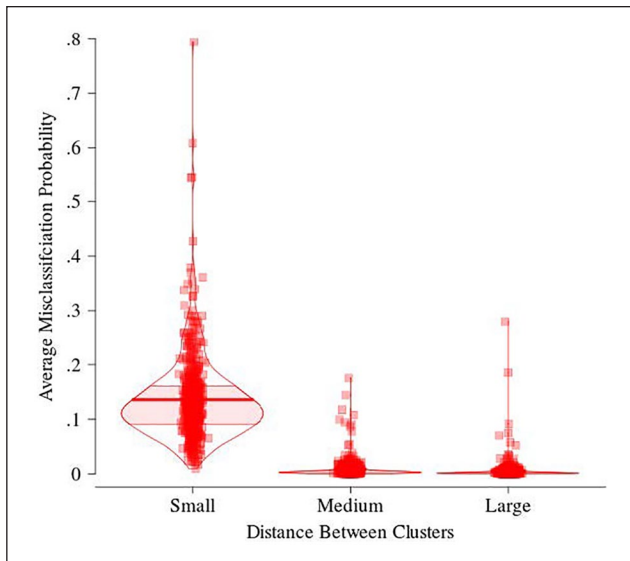
Figure 1. Distribution of average misclassification probability in the similar-true-parameters conditions across the three between-cluster distance conditions.
Note. Squares indicate the average misclassification probability within a data set, lines indicate the mean, bands the interquartile range. Average misclassification probability is based on overlap between components of the mixture distribution of individuals' true VAR(1) slopes.

were randomly selected to be multiplied by −1. To ensure that generated time series were stationary with eigenvalues of the regression coefficient matrices less than 1 in modulus, the VAR(1) slopes were multiplied with $.99/max(|\lambda_i|)$ in all three conditions; here $|\lambda_i|$ represents the vector of absolute values of the eigenvalues of the regression coefficient matrix Φ_i of the respective cluster i. Rescaling took place before the assignment of negative weights in the medium and large difference conditions. To offer an idea of the resulting distance between clusters for different conditions, the average misclassification probability of an individual within each simulated data set is provided in Figure 1. Misclassification probabilities were calculated based on overlap between components of the mixture distribution of individuals' true VAR(1) slopes according to Maitra and Melnykov (2010) using the MixSim R package (version 1.1.3; Melnykov, Chen, & Maitra, 2012). Overlap between components was only determined for the similar underlying VAR(1) coefficients conditions because only in these conditions true VAR(1) slopes were sampled from a mixture distribution. Naturally, a slight deviation will occur between individuals' estimated VAR(1) and true VAR(1) slopes due to sampling variation.

4. The variation between individuals of the same cluster (within-cluster distance):[6] *Identical underlying*

VAR(1) regression coefficients for individuals of the same cluster, where for each individual j in cluster i $\Phi_j = \Phi_i$, with Φ_i the matrix of regression slopes for cluster i. This data type is consistent with the implicit assumption of the nonprobabilistic clustering method proposed by Bulteel et al. (2016): exhibiting no variation between true parameters of individuals of the same cluster; With *similar underlying VAR(1) regression coefficients* for individuals of the same cluster. The matrix of regression slopes for individual j in cluster i, Φ_j, was taken from a multivariate normal distribution with mean vector $vec(\Phi_i)$ and variance–covariance matrix Σ_i, where $\Sigma_1 = \ldots = \Sigma_k = I_m \times .025$. Before selecting the resulting matrix of individual regression slopes, it was checked whether it was conform the assumption of stationary time series with all absolute values and all moduli of eigenvalues smaller than 1. The distribution of Φ_js within cluster i resembled thus a truncated multivariate normal distribution. Once a matrix Φ_j was selected for individual j, a time series of appropriate length was simulated with true VAR slopes equal to Φ_j according to equation (1).

5. The total number of individuals: 30, 60, or 120.
6. The number of time points per person: 51, 101, or 501.

For each of the resulting 324 conditions, we simulated 10 data sets. Subsequently, a probabilistic and a nonprobabilistic clustering analysis with the true number of clusters and correct lag order of $p = 1$ was performed on each data set with 1 rational start and 100 random starts. The nonprobabilistic clustering method had a rational initialization where the VAR(1) regression slopes were partitioned by use of hierarchical agglomerative clustering via Ward's criterion (Ward, 1963). The rational initialization of the EM algorithm used a partition reached by model-based hierarchical agglomerative clustering of individuals' regression slopes and intercepts. This method begins with clusters containing one individual each and merges the two clusters that are associated with the lowest decrease in classification likelihood for the mixture model (Banfield & Raftery, 1993).

The nonprobabilistic clustering method was estimated through an alternating least squares approach as proposed by Bulteel et al. (2016), using R. The code for estimating the method is provided in the supplementary material available online. The probabilistic clustering method was estimated by first determining VAR(1) coefficients using least squares estimation, subsequently clustering was carried out by use of the EM algorithm (Dempster et al., 1977) included in the mclust R package (version 5.3, Fraley & Raftery, 2002; Fraley, Raftery, Murphy, & Scrucca, 2012). The cluster parameters, Θ, and the posterior probabilities of cluster

Table 2. Mean (*SD*) ARI Values of the Two Clustering Methods Across Conditions.

Factor	Levels	Nonprobabilistic method	Probabilistic method
Distance between clusters	Small distance	**.574** (.427)	.575 (.381)
	Medium distance	.895 (.205)	**.939** (.147)
	Large distance	.927 (.168)	**.965** (.102)
Distance within clusters	Identical data	**.988** (.059)	.935 (.192)
	Similar data	.609 (.379)	**.718** (.348)
Number of persons	30	**.790** (.335)	.789 (.338)
	60	.800 (.331)	**.828** (.299)
	120	.806 (.326)	**.861** (.257)
Number of observations	51	**.769** (.336)	.736 (.352)
	101	.808 (.327)	**.855** (.271)
	501	.820 (.327)	**.887** (.250)
Cluster size	Equal proportion	.832 (.309)	**.844** (.277)
	Majority cluster	.804 (.320)	**.842** (.284)
	Minority cluster	.760 (.357)	**.793** (.336)
Number of clusters	2	.802 (.354)	**.843** (.314)
	4	.796 (.306)	**.809** (.286)

Note. ARI = adjusted Rand index. The largest ARIs in each condition are highlighted in bold.

membership, were estimated for different parametrizations of the covariance matrices via the EM algorithm. Because no constraints for the parametrization of the covariance matrices were imposed a priori the estimation procedure, the ideal parametrization was determined through model selection (cf. Fraley & Raftery, 2002; McNicholas, 2016) employing the BIC. Estimation of the probabilistic method relied on default values for determining convergence and terminating the EM algorithm (Fraley et al., 2012). Tolerance for relative convergence of the estimated log likelihood equaled thus $1 \times 10 - 5$, the relative parameter convergence tolerance in case of iterative M-steps equaled $1.490116 \times 10 - 8$. Because the EM algorithm for multivariate normal mixtures breaks down for singular covariance matrices (e.g., Bishop, 2006, Chapter 9), estimation was terminated when the covariance of a component density appeared to be singular by a threshold value of $2.220446 \times 10 - 16$. The best fitting probabilistic clustering solution was indicated by the BIC.[7] For each estimated model, the crisp cluster membership and cluster VAR(1) slopes, Φ_i, were considered.

Simulation 2: Unequal Intercepts Between Clusters

To evaluate how much the performance would improve for between-cluster differences in intercepts, we conducted a second simulation. This simulation deviated from the simulation above in two ways. First, instead of fixing the mean of all variables to zero in every cluster, the mean of the time series varied between clusters. Means of variables were assigned values in such a way that the difference between any two clusters was equal to the difference in means

between any other two clusters, with a Euclidean distance of 4.24. Second, in this simulation, we kept the factor of between cluster differences in VAR(1) coefficients fixed to the *small differences* condition. We selected this condition because in the first simulation this condition proved most difficult for the two clustering algorithms; Performance improvements due to between cluster differences in intercepts will thus be most visible in this condition.

Results Simulation 1

Recovery Performance. In this section, we contrast the recovery performance of the clustering methods regarding (a) cluster membership and (b) cluster-wise VAR(1) slopes. In the next section, clustering methods will be compared on their computational efficacy.

Recovery of cluster membership. To assess the recovery of cluster membership, we calculated the adjusted Rand index (ARI; Hubert & Arabie, 1985) between the estimated classification of individuals into clusters and the true partitioning. The ARI is a measure of agreement between classifications, taking on a value of 1 in case of perfect agreement and 0 when the agreement between partitions could have been expected by chance. Mean ARIs for different conditions are displayed in Table 2.[8] Overall, the probabilistic clustering method outperformed the nonprobabilistic method with mean ARIs of .83 (*SD* = .30) and .80 (*SD* = .33), respectively. This occurred in more complex situations, for instance, when the proportion between clusters was unequal or when participants of a cluster differed in their underlying VAR(1) coefficients. However, due to the increased complexity of the probabilistic method its

performance decreased more strongly when the sample size or the number of observations was low.

To gain an insight into effect size, we performed a repeated-measures analysis of variance (RM-ANOVA) on ARI values including all possible two-way interactions. Results are shown in Table 3. Between-method effects give an indication of the general influence of the manipulated factors on membership recovery, while within-data effects illustrate the differential effects of these factors on the two different clustering methods. Effect sizes suggest that low distance between ($\hat{\eta}_p^2 = .63$) and large distance within clusters ($\hat{\eta}_p^2 = .57$) are most detrimental for classification recovery with a large interaction between the two factors ($\hat{\eta}_p^2 = .47$). Thus, the cluster separation was found to influence cluster recovery most severely. A high number of observations appears the third largest factor for increasing classification recovery ($\hat{\eta}_p^2 = .10$). Within-data effects suggest that the performance of especially the nonprobabilistic clustering method suffers under increased within-cluster variation ($\hat{\eta}_p^2 = .18$).[9]

Recovery of true cluster-specific VAR(1) slopes. Recovery of coefficients was assessed through the mean[10] Euclidean distance between the estimated Φ_is, $\hat{\Phi}_i$, and true Φ_is used to generate the data:[11]

$$d_{E\,mean} = \frac{\sum_{i=1}^{k} \sqrt{\sum_{s=1}^{m^2} \left(\hat{\Phi}_{is} - \Phi_{is} \right)^2}}{K}$$

Table 4 shows average mean Euclidean distances across conditions. Again, the probabilistic method shows overall superior recovery over the nonprobabilistic method with mean values of .20 ($SD = .15$) and .24 ($SD = .19$), respectively. Table 5 displays the result of a RM-ANOVA on mean Euclidean distances. Effect sizes suggest that the largest influence on parameter recovery is shown for large distance within clusters ($\hat{\eta}_p^2 = .85$), a small distance between clusters ($\hat{\eta}_p^2 = .48$), a low number of observations ($\hat{\eta}_p^2 = .41$) and small sample size ($\hat{\eta}_p^2 = .28$). There was a large interaction between between-cluster and within-cluster variance ($\hat{\eta}_p^2 = .37$).[12] Within-data effects suggest that the performance of especially the nonprobabilistic method decreases when there is more variation within clusters ($\hat{\eta}_p^2 = .47$) while the probabilistic clustering method suffers more heavily under fewer observations ($\hat{\eta}_p^2 = .22$). Type of clustering method employed accounted for 14% of the overall variance.

Sensitivity to Local Minima. We also compared clustering methods on attraction rates, the percentage of runs out of the 101 that resulted in the model that was retained as solution. The average attraction rates across simulation conditions are displayed in Table 6 and the RM-ANOVA

results in Table 7. Both tables suggest that regardless of the data analyzed, the probabilistic clustering method had much lower attraction rates with a high percentage of runs resulting in a suboptimal solution ($\hat{\eta}_p^2 = .82$). Attraction rates of both methods were lower in scenarios where variation exists between the true parameters of individuals of the same cluster ($\hat{\eta}_p^2 = .54$) or when the number of clusters was high ($\hat{\eta}_p^2 = .38$).

Results Simulation 2

The results of the second simulation are compared with the results of the first simulation's *small differences* condition in Table 8. By introducing small differences in intercepts, the performance improved with regard to all three aspects evaluated here, with exception of the attraction rates of the probabilistic clustering method. The improvement is a little smaller than the improvement of increasing the distance in VAR(1) coefficients from *small differences* to *medium differences* (cf. the medium distance rows in Tables 2, 4, and 6). Larger distances between the cluster-level intercepts have the potential to improve the performance of the clustering algorithms to an even larger extent.

Discussion of Results

Based on the results of the simulation study, we draw the following conclusions. When true parameters of individuals within a cluster are not identical, the probabilistic clustering method demonstrates better recovery of both cluster membership and VAR(1) slopes than the nonprobabilistic approach. Under the conditions of our simulation, both methods performed well, but the retrieval of cluster parameters and classifications depended crucially on clusters being well-separated. Thus, the distance between clusters is an important predictor for the success of clustering methods, particularly with increasing variation within clusters. It is apparent that clustering methods should be employed only if visible differences between groups can be expected.

Compared with the simpler nonprobabilistic method, the probabilistic clustering method estimates additional parameters, such as the variation within clusters and the cluster proportions. Consequently, a high number of observations and a high sample size are required when this method is employed. While more complex, the estimation of additional parameters by the probabilistic clustering method can offer advantages by providing more information about the data structure. Employing a finite mixture model, the within-cluster variation in dynamics can be estimated. Furthermore, taking estimated proportions into account during the determination of probabilistic cluster membership allowed for improved inferences with regard to cluster classification when proportions were unequal.

Table 3. Results of a RM-ANOVA Evaluating the ARI Values of Both Clustering Methods.

	df	F	p	$\hat{\eta}^2_p$
Between-method effects				
Distance within clusters	1	4252.000	.000	.572
Number of persons	2	31.200	.000	.019
Number of observations	2	179.000	.000	.101
Cluster size	2	65.400	.000	.039
Number of clusters	1	19.600	.000	.006
Distance between clusters	2	2729.000	.000	.631
Distance within clusters × Number of persons	2	10.500	.000	.007
Distance within clusters × Number of observations	2	0.855	.425	.001
Distance within clusters × Cluster size	2	18.400	.000	.011
Distance within clusters × Number of clusters	1	3.350	.067	.001
Distance within clusters × Distance between clusters	2	1413.000	.000	.470
Number of persons × Number of observations	4	2.070	.082	.003
Number of persons × Cluster size	4	2.540	.038	.003
Number of persons × Number of clusters	2	0.254	.776	.000
Number of persons × Distance between clusters	4	2.540	.038	.003
Number of observations × Cluster size	4	4.010	.003	.005
Number of observations × Number of clusters	2	2.420	.090	.002
Number of observations × Distance between clusters	4	42.600	.000	.051
Cluster size × Number of clusters	2	97.600	.000	.058
Cluster size × Distance between clusters	4	1.770	.132	.002
Number of clusters ×Distance between clusters	2	0.550	.577	.000
Residuals	3188			
Within-data effects				
Clustering method	1	79.400	.000	.024
Distance within clusters × Method	1	698.000	.000	.180
Number of persons × Method	2	26.900	.000	.017
Number of observations × Method	2	98.900	.000	.058
Cluster size × Method	2	6.660	.001	.004
Number of clusters × Method	1	20.800	.000	.006
Distance between clusters × Method	2	19.800	.000	.012
Distance within clusters × Number of persons × Method	2	12.400	.000	.008
Distance within clusters × Number of observations × Method	2	5.610	.004	.004
Distance within clusters × Cluster size × Method	2	31.900	.000	.020
Distance within Clusters × Number of Clusters × Method	1	5.150	.023	.002
Distance within Clusters × Distance between Clusters × Method	2	157.000	.000	.090
Number of persons × Number of observations × Method	4	0.679	.606	.001
Number of persons × Cluster size × Method	4	0.438	.782	.001
Number of persons × Number of clusters × Method	2	0.082	.922	.000
Number of persons × Distance between clusters × Method	4	15.700	.000	.019
Number of observations × Cluster size × Method	4	1.000	.406	.001
Number of observations × Number of clusters × Method	2	0.051	.950	.000
Number of observations × Distance between clusters × Method	4	108.000	.000	.120
Cluster size × Number of clusters × Method	2	9.970	.000	.006
Cluster size × Distance between clusters × Method	4	20.400	.000	.025
Number of clusters × Distance between clusters × Method	2	9.010	.000	.006

Note. RM-ANOVA = repeated-measures analysis of variance; ARI = adjusted Rand index; *df* = degrees of freedom.

An Empirical Example

Data and Modelling Procedure

We applied both clustering methods to the data of the HowNutsAreTheDutch diary study (for details, see Van der Krieke et al., 2016). The data comprises 973 participants who were assessed three times daily (morning/midday/night) over 30 or 31 days. We focus here on participants who completed at least 68 measurements (*N* = 366). We selected Positive Affect Activation, Positive Affect Deactivation,

Table 4. Mean (*SD*) Euclidean Distances Between True and Estimated VAR(1) Slopes for the Two Clustering Methods Across Conditions.

Factor	Levels	Nonprobabilistic method	Probabilistic method
Distance between clusters	Small distance	.300 (.232)	**.274** (.180)
	Medium distance	.209 (.155)	**.173** (.114)
	Large distance	.203 (.149)	**.163** (.101)
Distance within clusters	Identical data	**.083** (.050)	.127 (.091)
	Similar data	.392 (.143)	**.280** (.148)
Number of persons	30	.265 (.182)	**.254** (.161)
	60	.232 (.185)	**.199** (.139)
	120	.215 (.193)	**.157** (.115)
Number of observations	51	**.261** (.171)	.285 (.148)
	101	.237 (.181)	**.193** (.121)
	501	.214 (.207)	**.132** (.122)
Cluster size	Equal proportion	.215 (.173)	**.181** (.129)
	Majority cluster	.241 (.193)	**.202** (.151)
	Minority cluster	.256 (.194)	**.226** (.151)
Number of clusters	2	.224 (.190)	**.176** (.128)
	4	.251 (.185)	**.231** (.156)

Note. The lowest Euclidean distances in each condition are highlighted in bold.

Negative Affect Activation, and Negative Affect Deactivation (PAA, PAD, NAA, and NAD) as endogenous variables for our analysis. On every measurement, each variable was assessed as the sum score of three items, each item was scored on a continuous scale from 0 to 100. Missing data and values for night measurements were imputed to fulfil the VAR assumption of equidistant time-points. Data were imputed using Amelia II (Honaker, King, & Blackwell, 2011) taking lags of previous and leads of following measurements into account. During the imputation process participants were treated as different cross-sections to account for interindividual differences. To account for possible effects of time of day, each individual time series was centered across the observed mean score of time of day of that individual.

During the first estimation step of the probabilistic clustering method, it was determined that of all 366 participants the time series of 345 were best summarized by a VAR(1), 19 were best described by a VAR(2) and 2 by a VAR(3) model. We decided to apply the probabilistic clustering method for both VAR(1) and VAR(2) models, to keep the dimensionality of parameter vectors, x_j, within manageable size. Models were fitted for a number of clusters between 2 and 5. The BIC indicated the two cluster solution as ideal for both the VAR(1) and the VAR(2) probabilistic method. The solutions with higher cluster numbers showed clusters of few individuals with quite distinct dynamic patterns, because we are interested especially in these patterns which deviate markedly from the other clusters, we selected the five cluster solutions of the probabilistic clustering methods.

For the nonprobabilistic method the ideal number of clusters balancing goodness of fit and model complexity was determined with the CHull procedure (Bulteel, Wilderjans, Tuerlinckx, & Ceulemans, 2013; Wilderjans, Ceulemans, & Meers, 2013). The CHull procedure favored the three cluster solution, but the final VAR(1) patterns of different clusters looked similar for other number of cluster solutions.

Cluster solutions were contrasted using relevant cross-sectional individual characteristics that had not been included in the clustering process to find whether the methods were able to differentiate clusters that would correspond to different personality characteristics. Clusters are described in terms of gender proportion, age, positive and negative affect schedule (Peeters, Ponds, & Vermeeren, 1996), anxiety and panic, depression (Quick Inventory of Depressive Symptomatology sum; Rush et al., 2003), self-reported proportion of time spent with others, neuroticism, extraversion, and openness (de Fruyt & Hoekstra, 2008); for details, see Van der Krieke et al. (2016).

Empirical Results

Mean VAR coefficients of the clusters probabilistic VAR(1) and VAR(2) clusters are shown in Figure 2. A comparison of Figures 2 (a) and 2 (b) shows that lag 2 coefficients fade once they are averaged across people with similar dynamics. Furthermore, clustering methods based on VAR(1) coefficients appear better able than the VAR(2) clustering solution at distinguishing groups of individuals with different personality characteristics relating to positive and negative affect. Table 9 and Figure 3 give the cluster-wise mean scores on cross-sectional measures for the probabilistic VAR(1) solution. To illustrate the potential of the probabilistic clustering method, we focus on the probabilistic

Table 5. Results of a RM-ANOVA Evaluating the Mean Euclidean Distances of Both Clustering Methods.

	df	F	p	$\hat{\eta}^2_p$
Between-method effects				
Distance within clusters	1	17672.000	.000	.847
Number of persons	2	608.000	.000	.276
Number of observations	2	1111.000	.000	.411
Cluster size	2	207.000	.000	.115
Number of clusters	1	551.000	.000	.147
Distance between clusters	2	1472.000	.000	.480
Distance within clusters × Number of persons	2	40.100	.000	.025
Distance within clusters × Number of observations	2	114.000	.000	.066
Distance within clusters × Cluster size	2	44.300	.000	.027
Distance within clusters × Number of clusters	1	80.900	.000	.025
Distance within clusters × Distance between clusters	2	930.000	.000	.368
Number of persons × Number of observations	4	19.900	.000	.024
Number of persons × Cluster size	4	8.970	.000	.011
Number of persons × Number of clusters	2	32.600	.000	.020
Number of persons × Distance between clusters	4	10.200	.000	.013
Number of observations × Cluster size	4	5.590	.000	.007
Number of observations × Number of clusters	2	28.100	.000	.017
Number of observations × Distance between clusters	4	13.700	.000	.017
Cluster size × Number of clusters	2	232.000	.000	.127
Cluster size × Distance between clusters	4	1.140	.335	.001
Number of clusters × Distance between clusters	2	0.543	.581	.000
Residuals	3188			
Within-data effects				
Clustering method	1	527.000	.000	.142
Distance within clusters × Method	1	2786.000	.000	.466
Number of persons × Method	2	86.100	.000	.051
Number of observations × Method	2	446.000	.000	.219
Cluster size × Method	2	2.690	.068	.002
Number of clusters × Method	1	88.700	.000	.027
Distance between clusters × Method	2	8.010	.000	.005
Distance within clusters × Number of persons × Method	2	106.000	.000	.062
Distance within clusters × Number of observations × Method	2	10.300	.000	.006
Distance within clusters × Cluster Size × Method	2	2.230	.107	.001
Distance within clusters × Number of clusters × Method	1	71.200	.000	.022
Distance within clusters × Distance between clusters × Method	2	75.800	.000	.045
Number of persons × Number of observations × Method	4	4.180	.002	.005
Number of persons × Cluster size × Method	4	0.910	.457	.001
Number of persons × Number of clusters × Method	2	0.251	.778	.000
Number of persons × Distance between clusters × Method	4	13.300	.000	.016
Number of observations × Cluster size × Method	4	0.452	.771	.001
Number of observations × Number of clusters × Method	2	0.938	.392	.001
Number of observations × Distance between clusters × Method	4	94.000	.000	.105
Cluster size × Number of clusters × Method	2	0.947	.388	.001
Cluster size × Distance between clusters × Method	4	14.600	.000	.018
Number of clusters × Distance between clusters × Method	2	28.400	.000	.018

Note. RM-ANOVA = repeated-measures analysis of variance; *df* = degrees of freedom.

VAR(1) 5 cluster solution in Table 9 (and Figure 2 (a)). Detailed results for the nonprobabilistic clustering model and the probabilistic VAR(2) model are shown in the supplementary material available online, so is the estimated within-cluster variance for the probabilistic VAR(1) method.

The estimation of within-cluster variance with regard to the parameters distinguishes the proposed method from the

Table 6. Average Attraction Rates (SD) of the Two Clustering Methods Across Simulation Conditions.

Factor	Levels	Nonprobabilistic method	Probabilistic method
Distance between clusters	Small distance	**.578** (.373)	.077 (.220)
	Medium distance	**.742** (.316)	.165 (.314)
	Large distance	**.778** (.296)	.190 (.333)
Distance within cluster	Identical data	**.859** (.214)	.272 (.375)
	Similar data	**.540** (.369)	.016 (.056)
Number of persons	30	**.703** (.324)	.172 (.326)
	60	**.704** (.343)	.146 (.298)
	120	**.691** (.355)	.113 (.261)
Number of observations	51	**.679** (.354)	.020 (.070)
	101	**.701** (.341)	.128 (.280)
	501	**.718** (.326)	.283 (.383)
Cluster size	Equal proportion	**.787** (.302)	.148 (.305)
	Majority cluster	**.643** (.357)	.125 (.274)
	Minority cluster	**.667** (.345)	.159 (.311)
Number of clusters	2	**.847** (.287)	.205 (.369)
	4	**.552** (.327)	.082 (.182)

Note. Highest attraction rates in each condition are highlighted in bold.

other time series clustering methods discussed in the introduction. These estimates allow to evaluate which variables differ most markedly between clusters relative to their within-cluster variance. This provides insight into the between-cluster differences that drive our cluster solution. With regard to VAR coefficients the lagged effect of NAA on PAA had the highest within-cluster variance with a *SD* of 0.173; the lagged effect of NAD on NAA had the smallest within-cluster variance with a *SD* of 0.118 (see Table 15 in the supplementary material available online). For the intercepts, the within-cluster *SD* varied from 53.123 (PAA) to 39.506 (NAA). Thus, a deviation of one *SD* from a cluster mean coefficient would require a difference between 0.173 and 0.118 in VAR coefficients or a difference between 53.123 and 39.506 in intercepts.

The probabilistic cluster memberships give direct indication of the classification uncertainty of every person. The classification uncertainty, the probability that an individual has been misclassified (i.e., $1 - \max(\hat{\pi}_i)_j, i = 1\ldots k$), is visualized across individuals in Figure 4 for the probabilistic VAR(1) method. For most of our 366 participants, the classification is conclusive; 323 participants are associated with a classification uncertainty below .2 and 299 participants with a classification uncertainty below .1.

Emotions are intensive and transient subjective states that emerge in response to specific stimuli and are typically described in terms of an affective valance dimension (positive to negative affect or PA/NA) and an arousal dimension (low-to-high bodily activation; Russell, 1980). Both dimensions are combined in our operationalization of activated or high arousal PA (energetic/enthusiastic/cheerful), deactivated or low arousal PA (relaxed/content/calm), activated NA (anxiety/nervous/gloomy), and deactivated NA

(gloomy/dull/tired). Below we will interpret the five clusters.

The Majority Cluster 2. The majority of our participants were in Cluster 2 (63%), which can therefore be used as a point of reference for the interpretation of the other clusters. Cluster 2 shows comparatively average PA and NA levels (see Table 9) with relatively high autocorrelations (see Figure 2) for activated and deactivated PA (.19/.17) and NA (.24/.25); thus relatively persistent and stable emotions over time and situations. External criteria (in Figure 3) are consistent with these observations and show comparatively high extraversion scores (closely aligned to PA), also reflected in the proportion of time spend with other people, and average neuroticism levels (closely aligned with NA; John, Robins, & Pervin, 2008). Temporal cross-valence correlations between PA and NA were rather weak (.04 to −.07, see Table 10), which might suggest two relatively independent within-person affect dimensions rather than a bipolar (PA ↔ NA) dimension, in line with previous work (Rush & Hofer, 2014; Russell & Carroll, 1999). However, it is salient that most cross-regressive associations were close to zero. This might be indicative of individual differences in cross-regressive effects that are cancelled out once effects are aggregated across individuals with similar dynamics. Alternatively, this could be seen as support for the notion that cross-regressive effects found through ideographic methods are often the result of overfitting (Bulteel, Mestdagh, Tuerlinckx, & Ceulemans, 2018).

Positive Affect Cluster 1. When we shift focus to the Cluster 1 participants (15%) they combine high average PA and relatively low NA, which is indicative of high emotional

Table 7. Results of a RM-ANOVA Evaluating the Attraction Rates of Both Clustering Methods.

	df	F	p	$\hat{\eta}_p^2$
Between-method effects				
Distance within clusters	1	3698.000	.000	.537
Number of persons	2	18.800	.000	.012
Number of observations	2	342.000	.000	.177
Cluster size	2	108.000	.000	.064
Number of clusters	1	1962.000	.000	.381
Distance between clusters	2	411.000	.000	.205
Distance within clusters × Number of persons	2	1.870	.155	.001
Distance within clusters × Number of observations	2	241.000	.000	.131
Distance within clusters × Cluster size	2	60.000	.000	.036
Distance within clusters × Number of clusters	1	62.200	.000	.019
Distance within clusters × Distance between clusters	2	21.900	.000	.014
Number of persons × Number of observations	4	6.300	.000	.008
Number of persons × Cluster size	4	0.744	.562	.001
Number of persons × Number of clusters	2	0.150	.860	.000
Number of persons × Distance between clusters	4	1.020	.393	.001
Number of observations × Cluster size	4	1.140	.335	.001
Number of observations × Number of clusters	2	22.000	.000	.014
Number of observations × Distance between clusters	4	13.400	.000	.017
Cluster size × Number of clusters	2	146.000	.000	.084
Cluster size × Distance between clusters	4	15.400	.000	.019
Number of clusters × Distance between clusters	2	15.100	.000	.009
Residuals	3188			
Within-data effects				
Clustering method	1	14150.000	.000	.816
Distance within clusters × Method	1	46.200	.000	.014
Number of persons × Method	2	8.450	.000	.005
Number of observations × Method	2	195.000	.000	.109
Cluster size × Method	2	80.800	.000	.048
Number of clusters × Method	1	339.000	.000	.096
Distance between clusters × Method	2	34.100	.000	.021
Distance within clusters × Number of persons × Method	2	23.300	.000	.014
Distance within clusters × Number of observations × Method	2	312.000	.000	.164
Distance within clusters × Cluster size × Method	2	27.100	.000	.017
Distance within clusters × Number of clusters × Method	1	256.000	.000	.074
Distance within clusters × Distance between clusters × Method	2	398.000	.000	.200
Number of persons × Number of observations × Method	4	2.680	.030	.003
Number of persons × Cluster size × Method	4	0.919	.452	.001
Number of persons × Number of clusters × Method	2	7.920	.000	.005
Number of persons × Distance between clusters × Method	4	4.730	.001	.006
Number of observations × Cluster size × Method	4	4.100	.003	.005
Number of observations × Number of clusters × Method	2	45.200	.000	.028
Number of observations × Distance between clusters × Method	4	16.000	.000	.020
Cluster size × Number of clusters × Method	2	128.000	.000	.074
Cluster size × Distance between clusters × Method	4	24.100	.000	.029
Number of clusters × Distance between clusters × Method	2	37.100	.000	.023

Note. RM-ANOVA = repeated-measures analysis of variance; *df* = degrees of freedom.

well-being (Diener, Scollon, & Lucas, 2009). These high PA scores align with comparatively high extraversion scores (see Figure 3), whereas low NA concedes with low neuroticism (also called emotional stability). In terms of emotion dynamics (see Table 10), Cluster 1 people experienced shorter lived activated NA than people in Cluster 2 (such as anxiety, .10 vs. .24), while their deactivated NA (such as weariness) and PA stabilities were comparable. Also, the temporal link between PA and NA was comparable to Cluster 2 (.10 to −.06, see Table 10).

Table 8. Differences Between the Simulations With Equal and Unequal Intercepts Between Clusters.

Performance aspect	Intercepts equal		Intercepts unequal	
	Nonprobabilistic method	Probabilistic method	Nonprobabilistic method	Probabilistic method
ARI	.574 (.427)	**.575** (.381)	.849 (.242)	**.870** (.234)
Euclidean distance	.300 (.232)	**.274** (.180)	.282 (.217)	**.212** (.126)
Attraction rate	**.578** (.373)	.077 (.220)	.677 (.342)	**.046** (.152)

Note. ARI = adjusted Rand index. Comparisons between the simulations are made with the distance between clusters fixed to small distance. The simulations are compared on mean (*SD*) ARI values, Euclidean distances between true and estimated VAR(1) slopes and attraction rates. The best performance aspects are highlighted in bold.
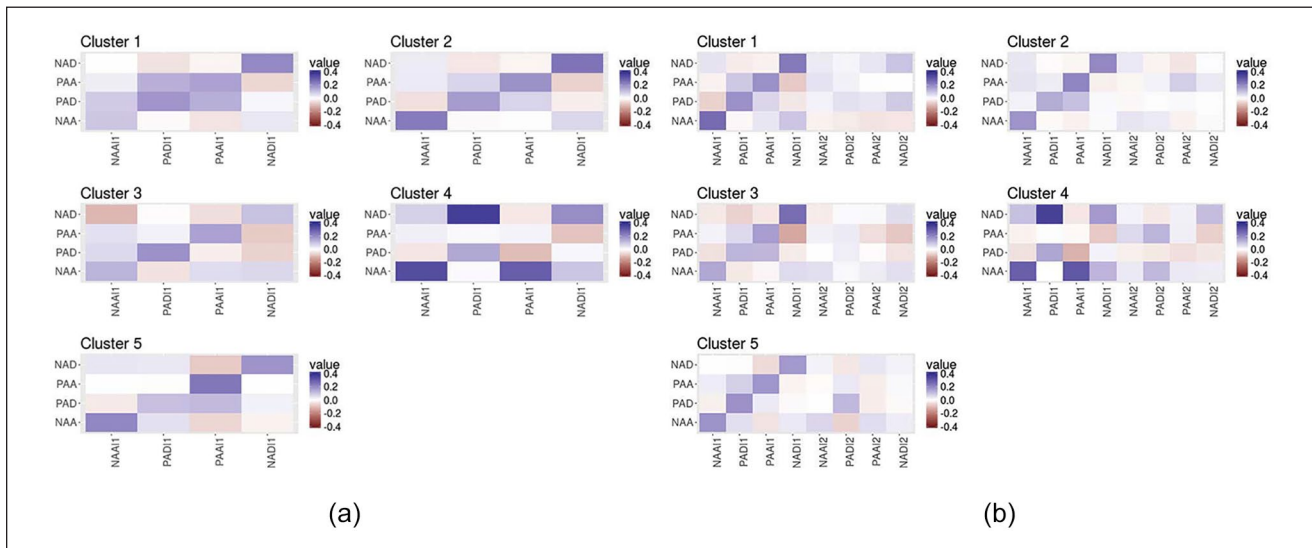

(a)   (b)

Figure 2. Clusters are based on the multivariate associations between the following variables: Negative Affect Deactivation (NAD), Positive Affect Activation (PAA), Positive Affect Deactivation (PAD), and Negative Affect Activation (NAA). (a) Mean VAR(1) coefficients per cluster, reached through the probabilistic VAR(1) clustering method. (b) Mean VAR(2) coefficients per cluster, reached through the probabilistic VAR(2) clustering method.
Note. The postfix l1 indicates columns hold the lag-1 coefficients that give the influence the variable will have on an emotion (listed in the rows) at the next measurement 6 hours later. The postfix l2 indicates columns which hold lag-2 coefficients that give the influence the variable will have on an emotion (listed in the rows) at the measurement 12 hours later.

Table 9. Mean (*SD*) Scores on Static Variables for the Clusters Reached by the Probabilistic Clustering VAR(1) Method.

	% Male	Age	PANAS PA	PANAS NA	Anxiety/panic	QIDS sum	Proportion
Cluster 1	0.21 (0.41)	47.00 (16.21)	35.65 (5.26)	15.82 (5.14)	0.40 (0.53)	4.00 (3.07)	.15
Cluster 2	0.17 (0.37)	45.05 (13.39)	32.71 (6.83)	20.11 (6.15)	0.73 (0.74)	6.16 (4.01)	.63
Cluster 3	0.33 (0.49)	50.78 (12.74)	34.67 (5.67)	15.11 (2.97)	0.28 (0.46)	3.67 (3.45)	.05
Cluster 4	0.00 (0.00)	51.50 (7.78)	28.50 (14.85)	26.00 (1.41)	1.50 (0.71)	15.00 (2.83)	.01
Cluster 5	0.19 (0.40)	42.25 (11.86)	27.77 (6.67)	28.07 (7.78)	1.44 (0.98)	10.12 (5.02)	.16

Note. PANAS = positive and negative affect schedule; QIDS = Quick Inventory of Depressive Symptomatology.

Well-Adjusted Cluster 3. Cluster 3 captures the 5% of the participants who were emotionally most stable and well-adjusted, as reflected in their comparatively low symptom levels of anxiety and depression (Table 9). The relatively high PA and low NA scores indicate high emotional well-being, and are in keeping with low neuroticism (i.e., emotional stability) and intermediate extraversion (see Figure 3). In terms of emotion dynamics Cluster 3 participants showed comparatively short-lived NA stability (activated .13, deactivated .11), compared with Cluster 2, but comparable PA stabilities (.16 and .19) and temporal cross-valence correlations (.06 to −.08).
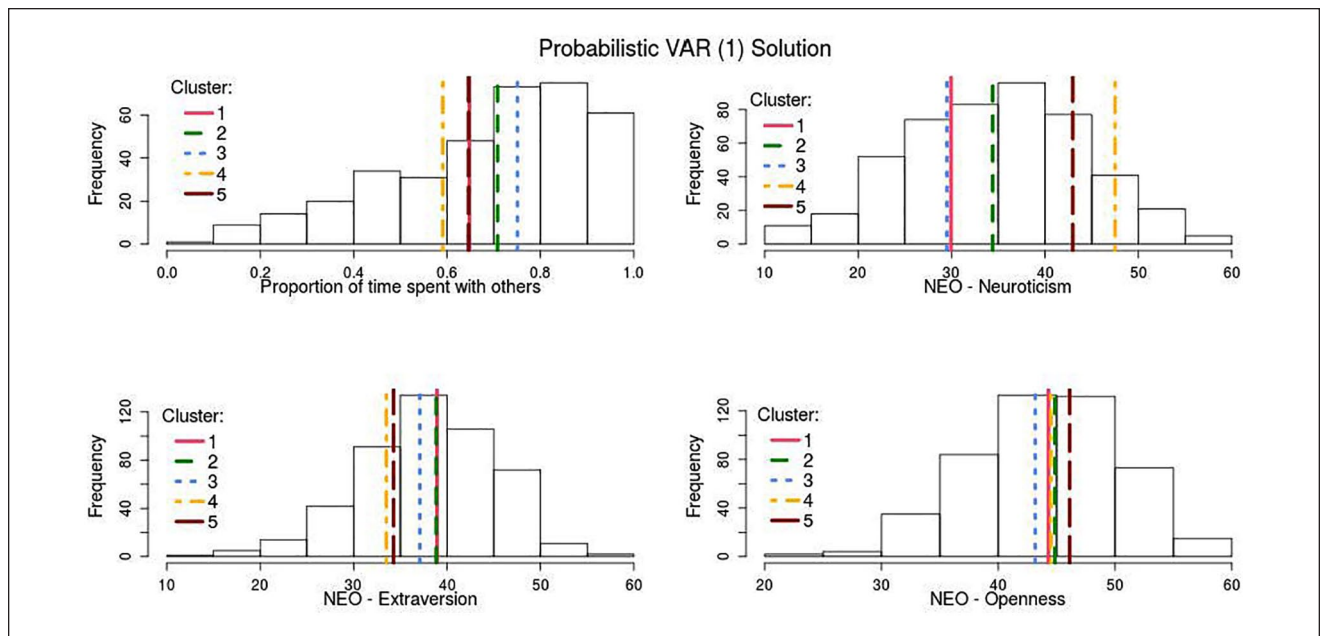
Figure 3. Distribution of scores on personality measurements of the whole sample.
Note. Vertical lines show mean values of clusters reached through the probabilistic VAR(1) clustering method.
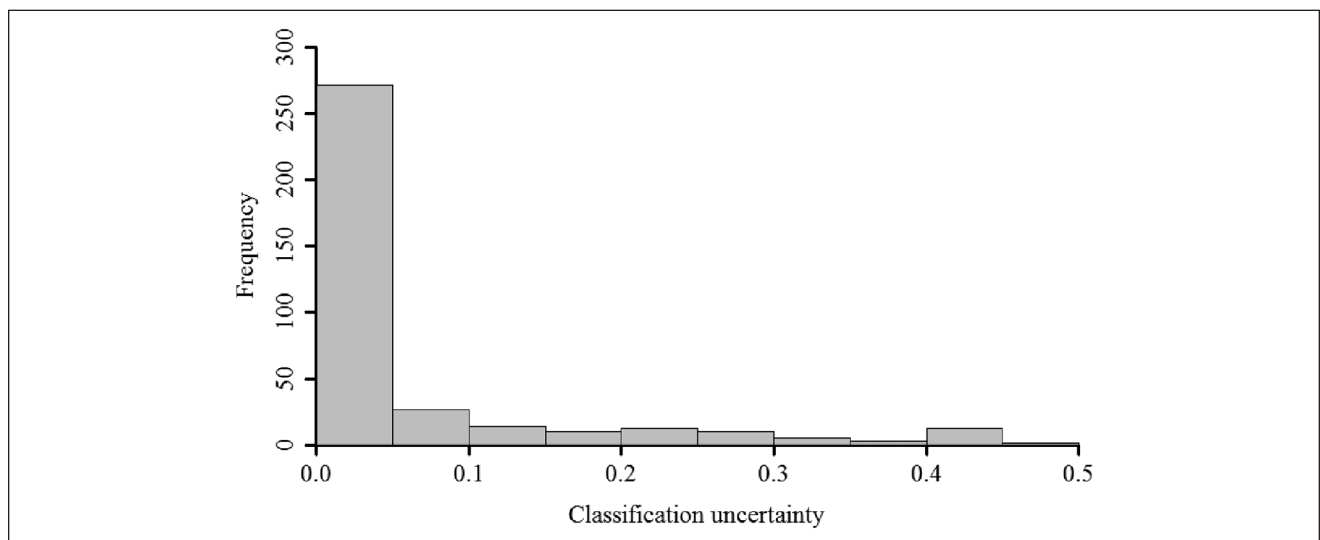

Figure 4. Classification uncertainty for all 366 participants in the solution reached by the probabilistic VAR(1) clustering method.

Depressed Cluster 4. Cluster 4 was rather small (1%) and captured two participants with comparatively high symptom levels of depression and anxiety, high NA and low PA (i.e., low emotional well-being). Unsurprisingly, Cluster 4 participants reported highest neuroticism and lowest extraversion levels (Figure 3), and spend most time alone. In terms of emotion dynamics their activated NA (.33) showed higher autocorrelation compared with the other clusters, whereas the stability of their activated PA was remarkably low (.02), which is in line with having a depressed mood (Watson, Clark, & Tellegen, 1988). Another unusual pattern was the strong temporal link between activated NA and PA (.30) and deactivated NA and PA (.37); When these participants felt relaxed/content (deactivated PA), they were more likely to be tired or gloomy (deactivated NA) 6 hours later. Overall, the results suggest that Cluster 4 people showed affective experiences with a stronger temporal relationship and a more bipolar nature.

Table 10. Mean VAR Coefficients of the Probabilistic Clustering VAR(1) Method.

	Cluster 1				Cluster 2				Cluster 3				Cluster 4				Cluster 5			
	NAA	PAD	PAA	NAD	NAA	PAD	PAA	NAD	NAA	PAD	PAA	NAD	NAA	PAD	PAA	NAD	NAA	PAD	PAA	NAD
Int	44.876	135.775	118.154	65.729	42.636	143.796	123.703	76.372	28.067	148.287	133.914	67.397	1.256	42.546	50.869	184.080	102.627	118.864	106.294	111.890
NAAl1	.096	.087	.030	-.001	.235	-.050	.035	.034	.127	.062	.050	-.113	.334	-.040	.026	.078	.212	-.033	-.002	.042
PADl1	-.009	.183	.136	-.046	-.008	.171	.074	-.038	-.045	.191	.025	-.005	.009	.150	.012	.371	.053	.106	.004	.039
PAAl1	-.044	.137	.162	-.017	.004	.075	.190	-.018	.058	-.025	.162	-.053	.296	-.107	.020	-.037	-.065	.117	.243	-.085
NADl1	.039	.016	-.064	.208	.067	-.027	-.071	.251	.068	-.070	-.083	.106	.097	.014	-.092	.201	-.019	.024	-.002	.187

Note. Rows hold the lag-1 coefficients that give the influence the variable will have on an emotion (listed in the columns) at the next measurement 6 hours later.

High Arousal Positive Affect Cluster 5. Cluster 5 captured 16% of the participants who reported relatively high depressive and anxious symptom levels and high NA, but also high PA, in contrast to Cluster 4 (Table 9). Cluster 5 participants were relatively high on neuroticism, although somewhat less than Cluster 4 people, and reported relatively low extraversion, but high openness (Figure 3). In terms of emotion dynamics the autocorrelations for activated and deactivated NA (.21/.19) were somewhat weaker but comparable to Cluster 2. However, the stabilities for activated PA was somewhat stronger (.24 vs. .19) but the stability of deactivated PA somewhat weaker (.11 vs. .17) than for Cluster 2. Cluster 5 participants reported relatively persistent high arousal positive experiences, in line with their younger age, but not with their comparatively low extraversion. The temporal cross-valence correlations (.06 to −.09) were not unusual, and supported relatively independent (or bivariate) affect dimensions.

Overall, relatively distinct affect dimensions were observed, except in Cluster 4. This is in line with previous research (Rush & Hofer, 2014) reporting on inversely correlated PA and NA factors at the within-person level and two independent PA and NA factors at the between-person level. As can be seen in Table 9 and Figure 3, provided that the *SD* of correlations is typically about four times larger within individuals than for groups (Fisher, Medaglia, & Jeronimus, 2018), these results stress the importance of accounting for both measurement levels and using within-person dynamics when clustering on between-person differences.

Conclusion

Dynamic clustering procedures might fill the methodological gap in time series analysis by allowing drawing accurate generalizations from intensive longitudinal data while accounting for qualitatively different processes between individuals. Employing probability distributions to model within-cluster variation appears a promising avenue for psychological research where most investigated constructs are assumed to be represented on a continuum with individuals displaying similar, but hardly ever equal dynamics. The probabilistic clustering method addresses two issues of concern inherent to the dynamic clustering of human data: (a) human data is likely to be characterized by variation within clusters and (b) most individual properties investigated in the social sciences cannot be assumed to occur in balanced proportion across the population.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

ORCID iD

Anja F. Ernst [ID] https://orcid.org/0000-0003-4398-7231

Supplemental Material

Supplemental material for this article is available online.

Notes

1. Prominent examples include *k*-means and *k*-median clustering.
2. Alternatively, a zero padding approach could be used where each time series is summarized by the most parsimonious model and nonestimated parameters of higher lag order are set to zero. We do not employ this approach to keep parameters at identical lags comparable and to not violate the normality assumption of VAR coefficients.
3. A distribution of stationary VAR coefficients is bounded above and below by 1 and −1, respectively. Even in extreme cases, however, their distribution within a large enough sample will approximate a truncated normal distribution. Gaussian mixture models have been shown robust to these deviations from normality (see, e.g., the triangular distribution condition in Steinley & Brusco, 2011).
4. While the regularity conditions of the BIC are not met in the context of a finite mixture model, an abundance of analytical and applied results demonstrates the appropriateness of the BIC for mixture model selection (see Fraley & Raftery, 2002, for a detailed review).
5. Consequently, in one condition (30 participants evenly distributed over four clusters) the sample size per cluster would be determined by a noninteger value. In this condition, two clusters were randomly selected that included eight individuals with the remaining two clusters containing seven individuals each.
6. Within-cluster distance is relative to the between-cluster distance. The simulation contains six different cluster distance conditions with within-cluster distance being manipulated at two levels, between-cluster distance at three levels.
7. See footnote 4.
8. Steinley (2004) provides guidelines for interpreting ARIs: values exceeding .9 are regarded as excellent, values over .8 indicate good recovery, values above .65 are moderate and everything below this threshold is taken as a sign of poor cluster recovery. However, a caveat in interpreting coefficients of membership recovery is that the concept of true cluster membership is only sensible as long as clusters are well-separated.
9. High interaction effects between cluster size and the number of clusters can be explained by the proportional differences between clusters being most extreme in the minority condition when the number of clusters equals two and in the majority condition when the number of clusters equals four.

10. To account for differences in number of clusters across conditions, we use mean Euclidean distances of every condition.

11. Before true and estimated cluster coefficients could be compared the two partitions had to be permuted so that corresponding clusters would be compared with one another. Clusters were thus arranged in such a way that overall overlap in classification between corresponding clusters of the two partitions was highest.

12. See footnote 9.

References

Adolf, J. K., & Fried, E. I. (2019). Ergodicity is sufficient but not necessary for group-to-individual generalizability. *Proceedings of the National Academy of Sciences of the United States of America*, *116*, 6540-6541. doi:10.1073/pnas.1818675116

Adolf, J. K., Schuurman, N. K., Borkenau, P., Borsboom, D., & Dolan, C. V. (2014). Measurement invariance within and between individuals: A distinct problem in testing the equivalence of intra- and inter-individual model structures. *Frontiers in Psychology*, *5*, 883. doi:10.3389/fpsyg.2014.00883

Anderlucci, L., & Viroli, C. (2015). Covariance pattern mixture models for the analysis of multivariate heterogeneous longitudinal data. *Annals of Applied Statistics*, *9*, 777-800. doi:10.1214/15-AOAS816

Banfield, J. D., & Raftery, A. E. (1993). Model-based Gaussian and non-Gaussian clustering. *Biometrics: Journal of the Biometric Society*, *49*, 803-821. doi:10.2307/2532201

Bellman, R. E. (1957). *Dynamic programming*. Princeton, NJ: Princeton University Press.

Beltz, A. M., Wright, A. G. C., Sprague, B. N., & Molenaar, P. C. M. (2016). Bridging the nomothetic and idiographic approaches to the analysis of clinical data. *Assessment*, *23*, 447-458. doi:10.1177/1073191116648209

Bishop, C. M. (2006). *Pattern recognition and machine learning* (Information Science and Statistics). Secaucus, NJ: Springer.

Bos, F. M., Schoevers, R. A., & aan het Rot, M. (2015). Experience sampling and ecological momentary assessment studies in psychopharmacology: A systematic review. *European Neuropsychopharmacology*, *25*, 1853-1864. doi:10.1016/j.euroneuro.2015.08.008

Bouveyron, C., & Brunet-Saumard, C. (2014). Model-based clustering of high-dimensional data: A review. *Computational Statistics & Data Analysis*, *71*, 52-78. doi:10.1016/j.csda.2012.12.008

Bringmann, L. F., Hamaker, E. L., Vigo, D. E., Aubert, A., Borsboom, D., & Tuerlinckx, F. (2017). Changing dynamics: Time-varying autoregressive models using generalized additive modeling. *Psychological Methods*, *22*, 409-425. doi:10.1037/met0000085

Brose, A., Schmiedek, F., Koval, P., & Kuppens, P. (2015). Emotional inertia contributes to depressive symptoms beyond perseverative thinking. *Cognition & Emotion*, *29*, 527-538. doi:10.1080/02699931.2014.916252

Bulteel, K., Mestdagh, M., Tuerlinckx, F., & Ceulemans, E. (2018). VAR(1) based models do not always outpredict AR(1) models in typical psychological applications. *Psychological Methods*, *23*, 740-756. doi:10.1037/met0000178

Bulteel, K., Tuerlinckx, F., Brose, A., & Ceulemans, E. (2016). Clustering vector autoregressive models: Capturing qualitative differences in within-person dynamics. *Frontiers in Psychology*, *7*, 1540. doi:10.3389/fpsyg.2016.01540

Bulteel, K., Wilderjans, T. F., Tuerlinckx, F., & Ceulemans, E. (2013). CHull as an alternative to AIC and BIC in the context of mixtures of factor analyzers. *Behavior Research Methods*, *45*, 782-791. doi:10.3758/s13428-012-0238-5

Corduas, M., & Piccolo, D. (2008). Time series clustering and classification by the autoregressive metric. *Computational Statistics & Data Analysis*, *52*, 1860-1872. doi:10.1016/j.csda.2007.06.001

de Fruyt, F., & Hoekstra, H. (2008). *NEO-PI-3 persoonlijkheidsvragenlijst* [NEO-PI-3 Personality Questionnaires]. Amsterdam, Netherlands: Hogrefe.

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood for incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B: Methodological*, *39*, 1-38.

Diener, E., Scollon, C. N., & Lucas, R. E. (2009). The evolving concept of subjective well-being: The multifaceted nature of happiness. In *Assessing well-being: The collected works of Ed Diener* (pp. 67-100). Berlin, Germany: Springer Science + Business Media.

D'Urso, P., De Giovanni, L., & Massari, R. (2015). Time series clustering by a robust autoregressive metric with application to air pollution. *Chemometrics and Intelligent Laboratory Systems*, *141*, 107-124. doi:10.1016/j.chemolab.2014.11.003

D'Urso, P., De Giovanni, L., Massari, R., & Di Lallo, D. (2013). Noise fuzzy clustering of time series by autoregressive metric. *Metron*, *71*, 217-243. doi:10.1007/s40300-013-0024-x

Ernst, A. F., Albers, C. J., Jeronimus, B. F., & Timmerman, M. E. (2019). *Inter-individual differences in multivariate time series: Dynamic adaptive cluster modelling based on finite mixtures of vector-autoregressive processes*. Manuscript submitted for publication.

Fisher, A. J., Medaglia, J. D., & Jeronimus, B. F. (2018). Lack of group-to-individual generalizability is a threat to human subjects research. *Proceedings of the National Academy of Sciences of the United States of America*, *115*, E6106-E6115. doi:10.1073/pnas.1711978115

Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis and density estimation. *Journal of the American Statistical Association*, *97*, 611-631.

Fraley, C., Raftery, A. E., Murphy, T. B., & Scrucca, L. (2012). *mclust Version 4 for R: Normal mixture modeling for model-based clustering, classification, and density estimation* (Technical Report No. 597). Retrieved from https://www.stat.washington.edu/sites/default/files/files/reports/2012/tr597.pdf

Gates, K. M., Lane, S. T., Varangis, E., Giovanello, K., & Guskiewicz, K. (2017). Unsupervised classification during time-series model building. *Multivariate Behavioral Research*, *52*, 129-148. doi:10.1080/00273171.2016.1256187

Hamaker, E. L., Grasman, R. P. P. P., & Kamphuis, J. H. (2016). Modeling BAS dysregulation in bipolar disorder: Illustrating the potential of time series analysis. *Assessment*, *23*, 436-446. doi:10.1177/1073191116632339

Honaker, J., King, G., & Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software*, *45*(7), 1-47.

Hubert, L. J., & Arabie, P. (1985). Comparing partitions. *Journal of Classification*, *2*, 193-218. doi:10.1007/BF01908075

Jacques, J., & Preda, C. (2014). Functional data clustering: A survey. *Advances in Data Analysis and Classification*, *8*, 231-255. doi:10.1007/978-3-7908-2062-1

Jebb, A. T., Tay, L., Wang, W., & Huang, Q. (2015). Time series analysis for psychological research: Examining and forecasting change. *Frontiers in Psychology*, *6*, 727. doi:10.3389/fpsyg.2015.00727

John, O. P., Robins, R. W., & Pervin, L. A. (2008). *Handbook of personality: Theory and research* (3rd ed.). New York, NY: Guilford Press.

Koval, P., & Kuppens, P. (2012). Changing emotion dynamics: individual differences in the effect of anticipatory social stress on emotional inertia. *Emotion*, *12*, 256-267. doi:10.1037/a0024756

Koval, P., Kuppens, P., Allen, N. B., & Sheeber, L. (2012). Getting stuck in depression: The roles of rumination and emotional inertia. *Cognition & Emotion*, *26*, 1412-1427. doi:10.1080/02699931.2012.667392

Krone, T., Albers, C. J., Kuppens, P., & Timmerman, M. E. (2018). A multivariate statistical model for emotion dynamics. *Emotion*, *18*, 739-754. doi:10.1037/emo0000384

Lane, S. T., Gates, K. M., Pike, H. K., Beltz, A. M., & Wright, A. G. C. (2019). Uncovering general, shared, and unique temporal patterns in ambulatory assessment data. *Psychological Methods*, *24*, 54-69. doi:10.1037/met0000192

Liao, W. T. (2005). Clustering of time series data: A survey. *Pattern Recognition*, *38*, 1857-1874. doi:10.1016/j.patcog.2005.01.025

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Berlin, Germany: Springer.

Maitra, R., & Melnykov, V. (2010). Simulating data to study performance of finite mixture modeling and clustering algorithms. *Journal of Computational and Graphical Statistics*, *19*, 354-376. doi:10.1198/jcgs.2009.08054

McLachlan, G. J., & Basford, K. E. (1988). *Mixture models*. New York, NY: Marcel Dekker.

McLachlan, G. J., & Peel, D. (2004). Finite mixture models. New York, NY: Wiley-Interscience.

McNicholas, P. D. (2016). Model-based clustering. *Journal of Classification*, *33*, 331-373. doi:10.1007/s00357-016-9211-9

McNicholas, P. D., & Murphy, T. B. (2010). Model-based clustering of longitudinal data. *Canadian Journal of Statistics*, *38*, 153-168. doi:10.1002/cjs.10047

Melnykov, V., Chen, W.-C., & Maitra, R. (2012). MixSim: An R package for simulating data to study performance of clustering algorithms. *Journal of Statistical Software*, *51*(12), 1-25.

Michael, S., & Melnykov, V. (2016). Finite mixture modeling of Gaussian regression time series with application to dendrochronology. *Journal of Classification*, *33*, 412-441. doi:10.1007/s00357-016-9216-4

Molenaar, P. C. M. (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research and Perspectives*, *2*, 201-218. doi:10.1207/s15366359mea0204_1

Peeters, F., Ponds, R., & Vermeeren, M. (1996). Affectiviteit en zelfbeoordeling van depressie en angst [Affectivity and self-assessment of depression and anxiety]. *Tijdschrift Voor De Psychiatrie*, *3*, 240-250.

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rush, A. J., Trivedi, M. H., Ibrahim, H. M., Carmody, T. J., Arnow, B., Klein, D. N., . . . Keller, M. B. (2003). The 16-item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): A psychometric evaluation in patients with chronic major depression. *Biological Psychiatry*, *54*, 573-583. doi:10.1016/S0006-3223(02)01866-8

Rush, J., & Hofer, S. M. (2014). Differences in within- and between-person factor structure of positive and negative affect: Analysis of two intensive measurement studies using multilevel structural equation modeling. *Psychological Assessment*, *26*, 462-473. doi:10.1037/a0035666

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*, 1161-1178. doi:10.1037/h0077714

Russell, J. A., & Carroll, J. M. (1999). On the bipolarity of positive and negative affect. *Psychological Bulletin*, *125*, 3-30. doi:10.1037/0033-2909.125.1.3

Schenk, H. M., Bos, E. H., Slaets, J. P., de Jonge, P., & Rosmalen, J. G. (2017). Differential association between affect and somatic symptoms at the between- and within-individual level. *British Journal of Health Psychology*, *22*, 270-280. doi:10.1111/bjhp.12229

Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics*, *6*, 461-464. doi:10.1214/aos/1176344136

Scott, S. B., Sliwinski, M. J., & Fields, F. B. (2013). Age differences in emotional responses to daily stress: The role of timing, severity, and global perceived stress. *Psychology and Aging*, *28*, 1076-1087. doi:10.1037/a0034000

Shifren, K., Hooker, K., Wood, P., & Nesselroade, J. R. (1997). Structure and variation of mood in individuals with Parkinson's disease: A dynamic factor analysis. *Psychology and Aging*, *12*, 328-339. doi:10.1037/0882-7974.12.2.328

Steinley, D. (2004). Properties of the Hubert-Arabie adjusted Rand index. *Psychological Methods*, *9*, 386-396. doi:10.1037/1082-989X.9.3.386

Steinley, D., & Brusco, M. J. (2011). Evaluating mixture modeling for clustering: Recommendations and cautions. *Psychological Methods*, *16*, 63-79. doi:10.1037/a0022673

van der Krieke, L., Blaauw, F. J., Emerencia, A. C., Schenk, H. M., Slaets, J. P. J., Bos, E. H., . . . Jeronimus, B. F. (2017). Temporal dynamics of health and well-being: A crowdsourcing approach to momentary assessments and automated generation of personalized feedback. *Psychosomatic Medicine*, *79*, 213-223. doi:10.1037/t03592-000

van der Krieke, L., Jeronimus, B. F., Blaauw, F. J., Wanders, R. B. K., Emerencia, A. C., Schenk, H. M., . . . de Jonge, P. (2016). HowNutsAreTheDutch (HoeGekIsNL): A crowdsourcing

study of mental symptoms and strengths. *International Journal of Methods in Psychiatric Research*, *25*, 123-144. doi:10.1002/mpr.1495

Viroli, C. (2011). Finite mixtures of matrix normal distributions for classifying three-way data. *Statistics and Computing*, *21*, 511-522. doi:10.1007/s11222-010-9188-x

Voelkle, M. C., Brose, A., Schmiedek, F., & Lindenberger, U. (2014). Toward a unified framework for the study of between-person and within-person structures: Building a bridge between two research paradigms. *Multivariate Behavioral Research*, *49*, 193-213. doi:10.1080/00273171.2014.889593

Ward, J. H., Jr. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, *58*, 236-244. doi:10.2307/2282967

Watson, D., Clark, L., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Emotion*, *54*, 1063-1070.

Wilderjans, T. F., Ceulemans, E., & Meers, K. (2013). CHull: A generic convex-hull-based model selection method. *Behavior Research Methods*, *45*, 1-15. doi:10.3758/s13428-012-0238-5

Wright, A. G. C., & Zimmermann, J. (2019). Uncovering general, shared, and unique temporal patterns in ambulatory assessment data. *Psychological Methods*, *24*, 54-69.