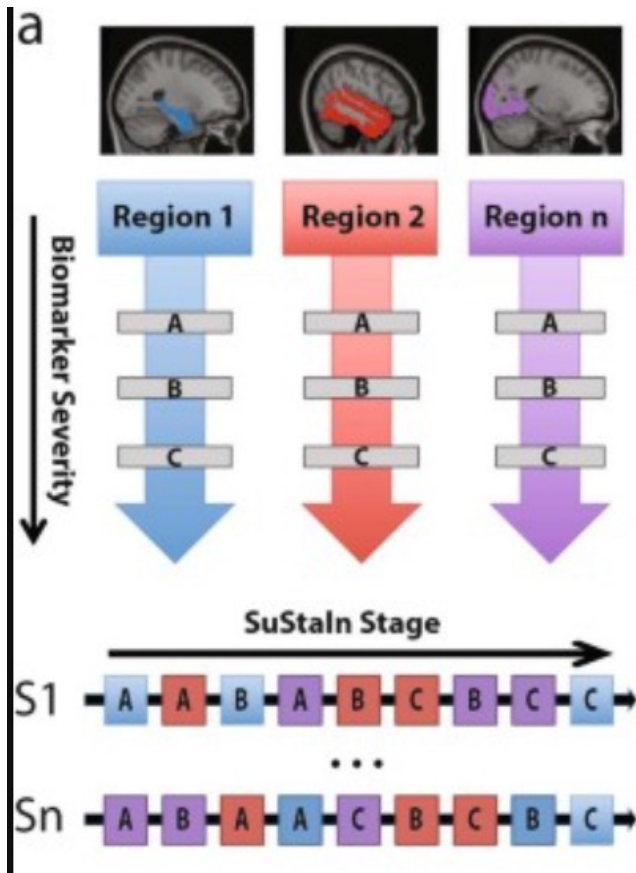


# Subtype and **Stage Inference** Model Implementation on LEADS tau PET

Marlene

2024 March

# SuStain Overview



Vogel et al., *Nat Med* 2020

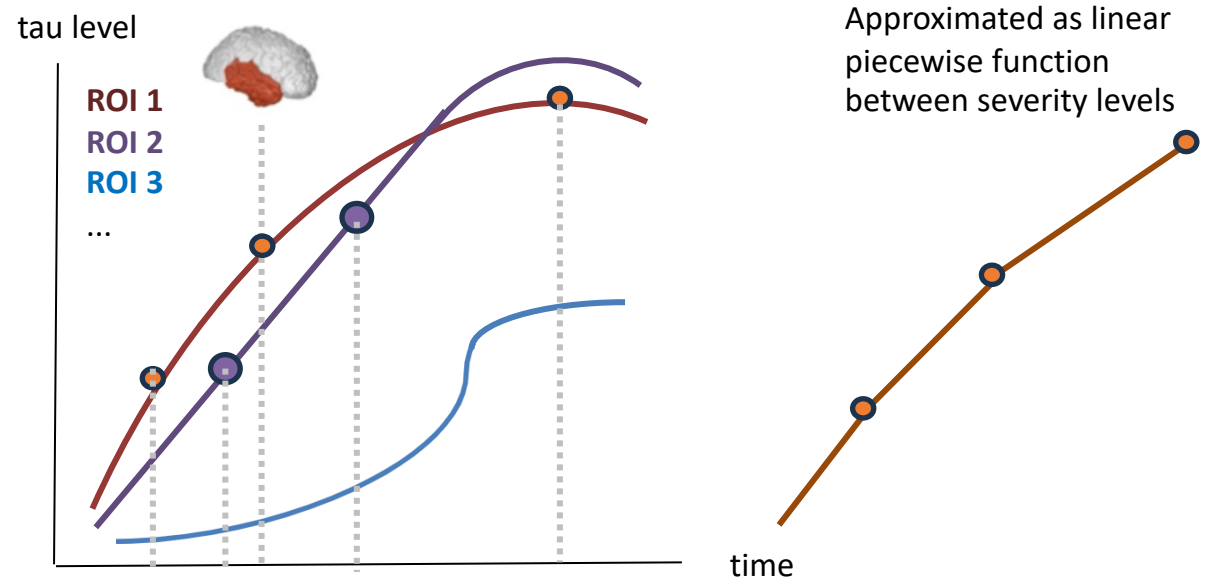
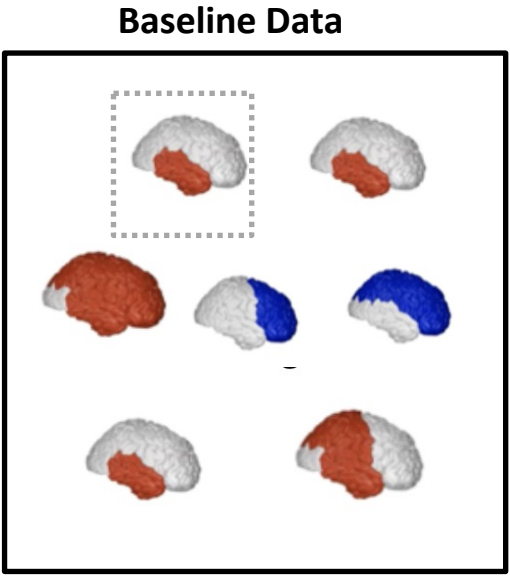
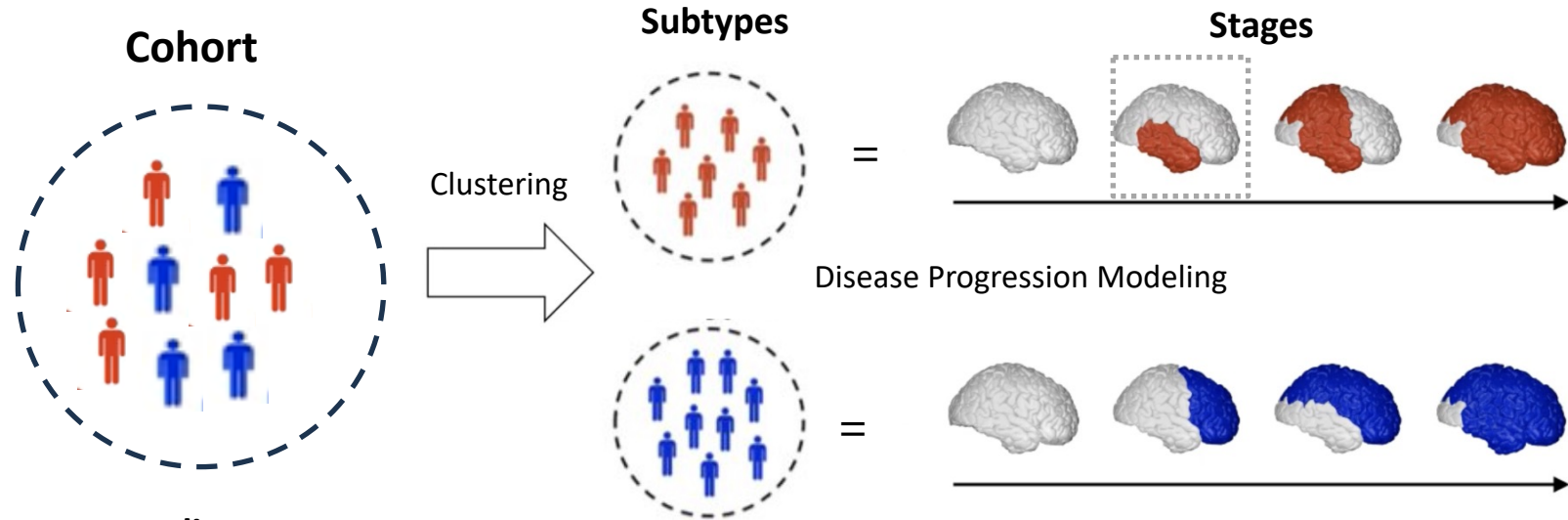
Modeling subtype and stage simultaneously with event-based modeling:

- Events = the progression of biomarker (regional tau accumulation) to certain severity levels
- Subtypes = distinct sequences of events
- Stages = occurrences of events down the sequence

Assumptions:

- 1) arbitrary timescale
- 2) reconstruction of longitudinal trajectories from cross-sectional data
- 3) Stable subtypes overtime – no crossover
- 4) monotonicity in biomarker level

# Model Intuition



# Model Intuition

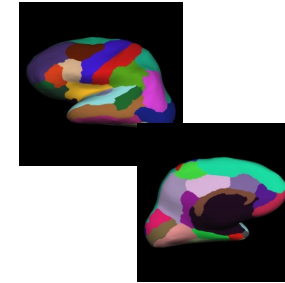
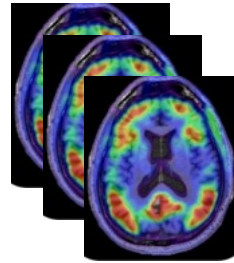
## **Given**

biomarker values (regional tau levels of patients)  
severity levels for each ROI

## **Determined through clustering**

the most likely sequence of events (order of reaching severity levels)?  
the most likely proportions of different patterns in the data?

# Data Preparation



Left & Right:  
 Frontal  
 Occipital  
 Parietal  
 Temporal  
 MTL  
 Lobar ROI

Study participants

tau PET images

SUVR images

Parcellation

Lobar ROI

Tabular data

dx	L_MTL	L_frontal	L_occipital	L_parietal	L_temporal	R_MTL	R_frontal	R_occipital	R_parietal	R_temporal
CN	1.212698	1.029354	1.103603	1.074372	1.127232	1.210549	1.025647	1.109314	1.074605	1.142007
EOAD	1.856240	2.413739	2.957153	3.012157	2.856721	1.695600	2.197533	2.754564	2.772873	2.635381
EOAD	1.329493	1.744891	2.092321	2.439468	2.290511	1.271227	1.697044	1.870114	2.304942	2.221782

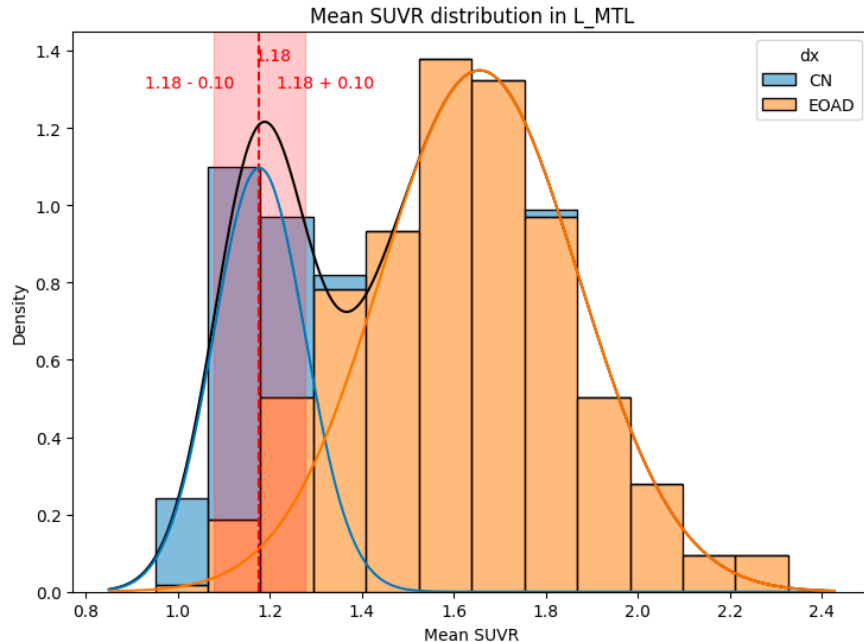
...

$X = \{x_{ij} \mid i = 1 \dots I, j = 1 \dots J\}$   
 $x_{ij}$  : mean SUVR of subject  $j$  in ROI  $i$

- LEADS (a $\beta$ - CN: 89; EOAD: 379) tau-PET images
- Co-registered to MRI, Normalized w.r.t. inferior cerebellar gray to derive SUVR images
- Parcellations using the Desikan–Killiany atlas (68 regions)
- Combined into ten lobar ROIs according to Vogel's
- For each ROI, calculate the volume-weighted mean SUVR

# Standardization using GMM

Standardization using the mean and SD of the 1st GMM component



- Two-component GMM (one for normal, one for abnormal) is fitted on all subjects for individual ROIs
- Then standardized using the mean and standard deviation of the normal component

$$\mathbf{X} = \{x_{ij} \mid i = 1 \dots I, j = 1 \dots J\}$$

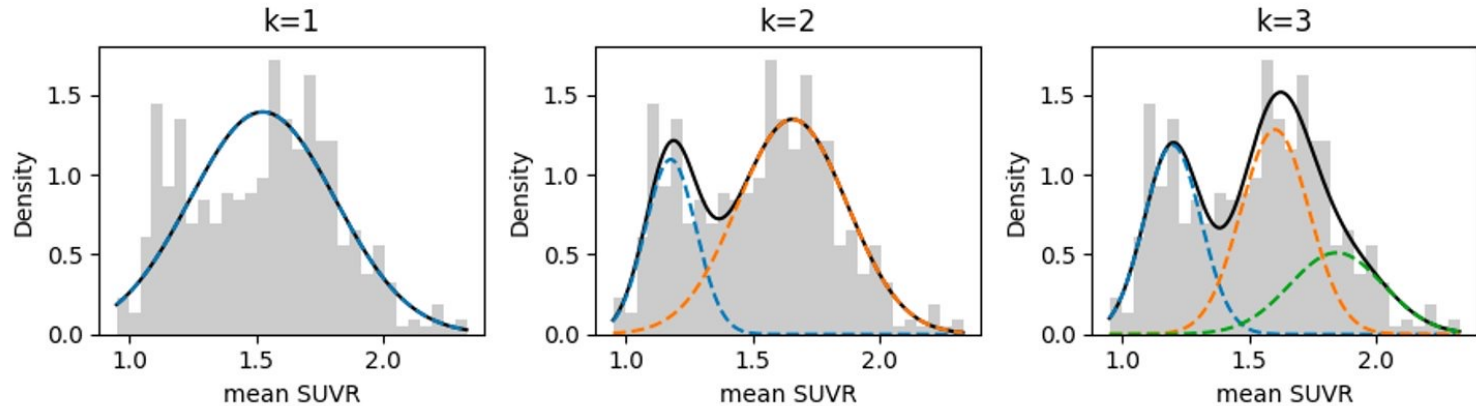
Data matrix

$$x_{ij} = (x_{ij} - \mu_{\text{normal}}) / \sigma_{\text{normal}}$$

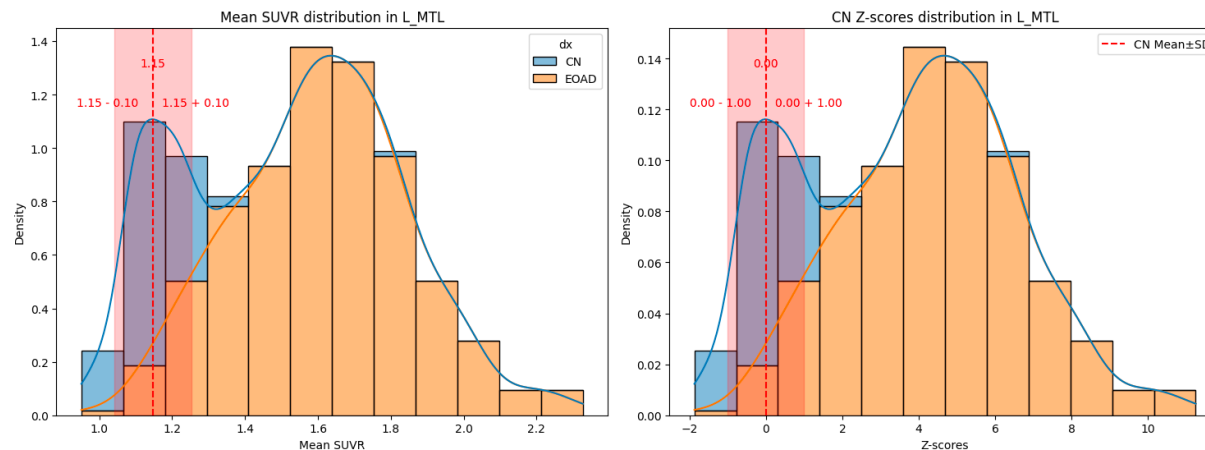
Standardization using  $\mu_{\text{normal}}$  and  $\sigma_{\text{normal}}$

# Alt. Standardization

1-3 component GMM fit on mean SUVR in L\_MTL (Best: K = 2)

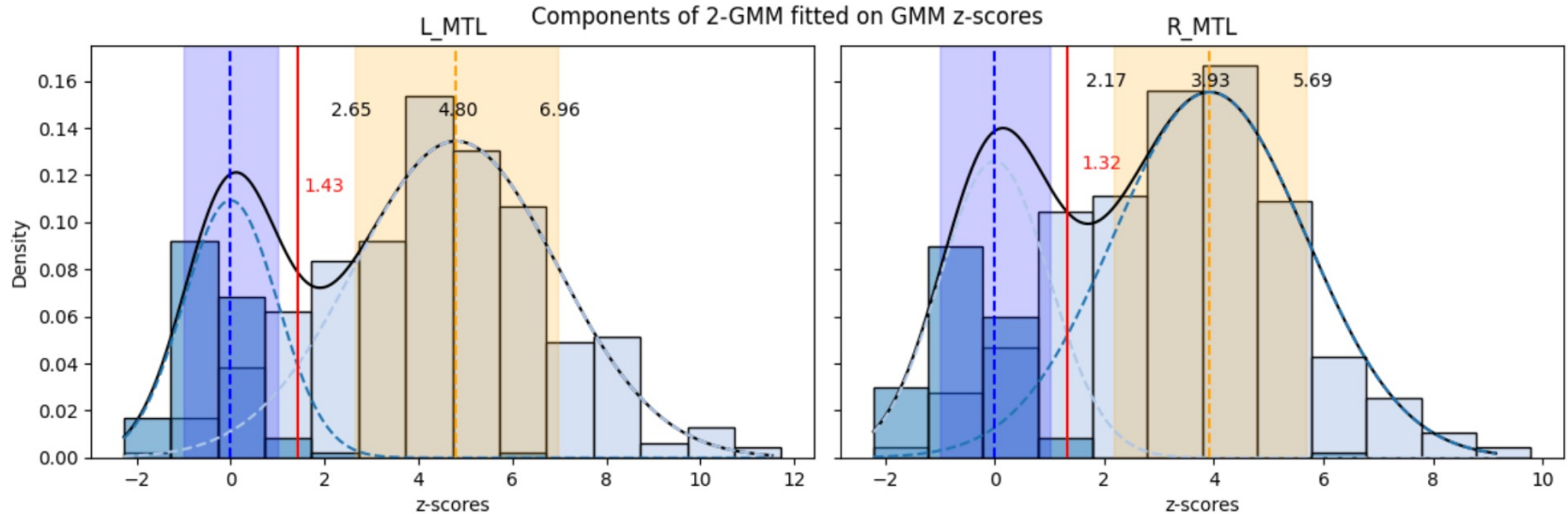


- Data-driven k-GMM confirmed most ROIs follow 2-GMM (exceptions: R & L occipitals)  
K = 1~6, optimized for 5-fold CV AIC



- Standardization using the mean and standard deviation of the CN group yields similar results (Pearson correlation coefficient  $\sim 1$  for all ROIs)

# Severity Levels



For ROI  $i$ ,  $Z_{i1}$ ,  $Z_{i2}$ , ...  $Z_{iR_i}$

(For each ROI) Severity levels are decided based on Z-scores

$$N = \sum_{i=1}^I R_i$$

The total number of z-scores events in the sequence

- The event sequence  $\mathbf{S}$  is then represented as a series of these events  $E_{iz}$  and has length  $N$
- At the end of the event sequence (stage  $N$ ), regional tau accumulates to their respective max levels
- In this implementation, all ROI has severity levels (arbitrarily chosen) of  $z = 2, 5, 10$ .  $N = 30$ .
- Alt.: the intersection of the 1st and 2nd component,  $\text{mean} \pm \text{SD}$  of the 2nd component

i.e., ROI1 = L\_MTL, ROI2 = R\_MTL,  $Z_{11}=1.43$ ,  $Z_{12}=2.65$ ,  $Z_{13}=6.96$ ;  $Z_{21}=1.32$ ,  $Z_{22}=2.17$ ,  $Z_{23}=3.69$ ;  $R_1 = R_2 = 3$ .



# Math

*Prior Probability of individual belonging to a certain stage at time t (uniform)*

$$P(\mathbf{X}|\mathbf{S}) = \prod_{j=1}^J \left[ \sum_{k=0}^N \left( \int_{t=\frac{k}{N+1}}^{t=\frac{k+1}{N+1}} \left( P(t) \prod_{i=1}^I P(x_{ij}|t) \right) \partial t \right) \right]$$

Young et al., Nat Commun 2018

*Likelihood of tau levels at time t for subject j*

*Trajectory of tau accumulations between stage k and k+1 for subject j*

*Overall trajectories from stage 0 to N for subject j*

*Likelihoods for tau distribution data for all subjects (X) given the order of events (S)*

**where,**

*Linear piecewise function that models tau accumulation in ROI i overtime*

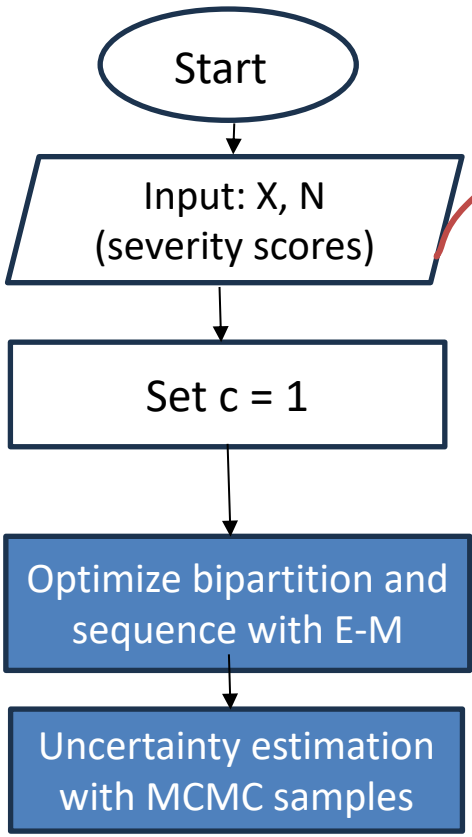
$$P(x_{ij}|t) = \text{NormPDF}(x_{ij}, \boxed{g_i(t)}, \sigma_i) . \quad \text{Probability of tau level at ROI i at time t for individual j follows normal distribution}$$

**Hence, considering all subtypes (C = number of clusters) for the data,**

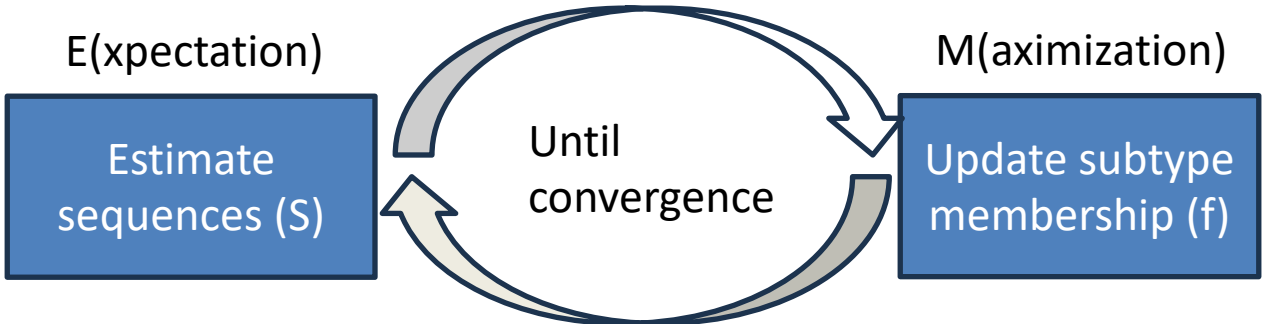
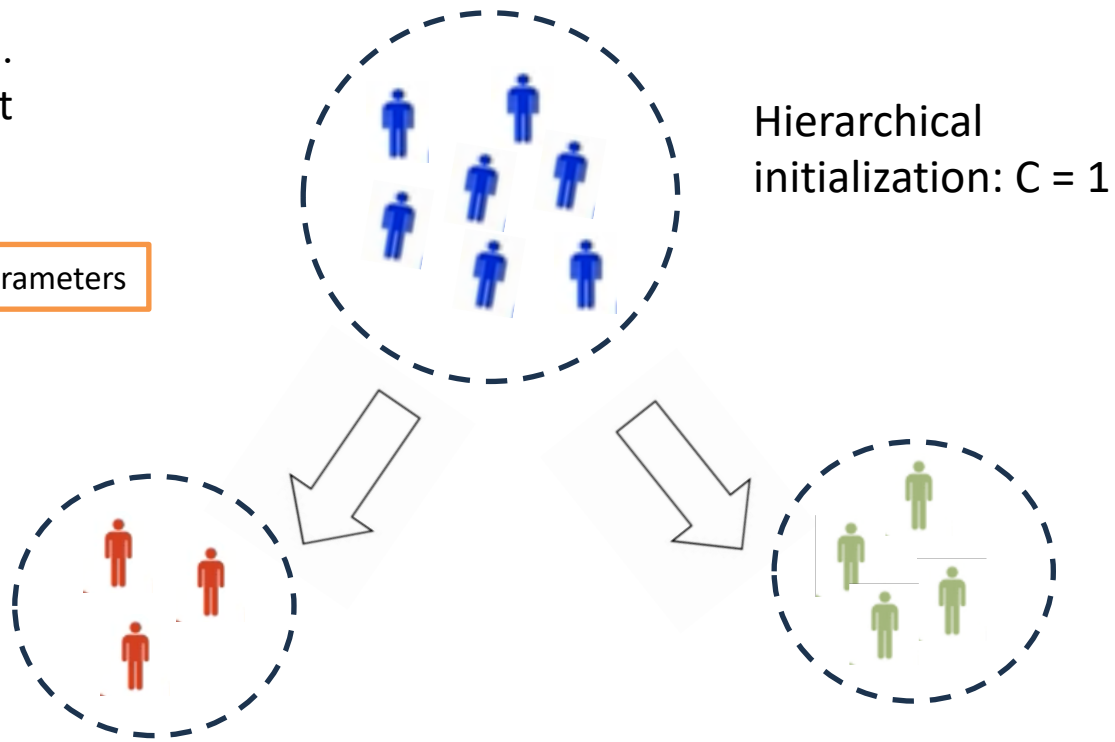
$$P(\mathbf{X}|\mathbf{M}) = \sum_{c=1}^C f_c P(\mathbf{X}|\mathbf{S}_c) \quad \begin{array}{l} M = \text{the mixture of z-scores models} \\ f_c = \text{proportion of subjects belonging to a subtype } c \end{array}$$

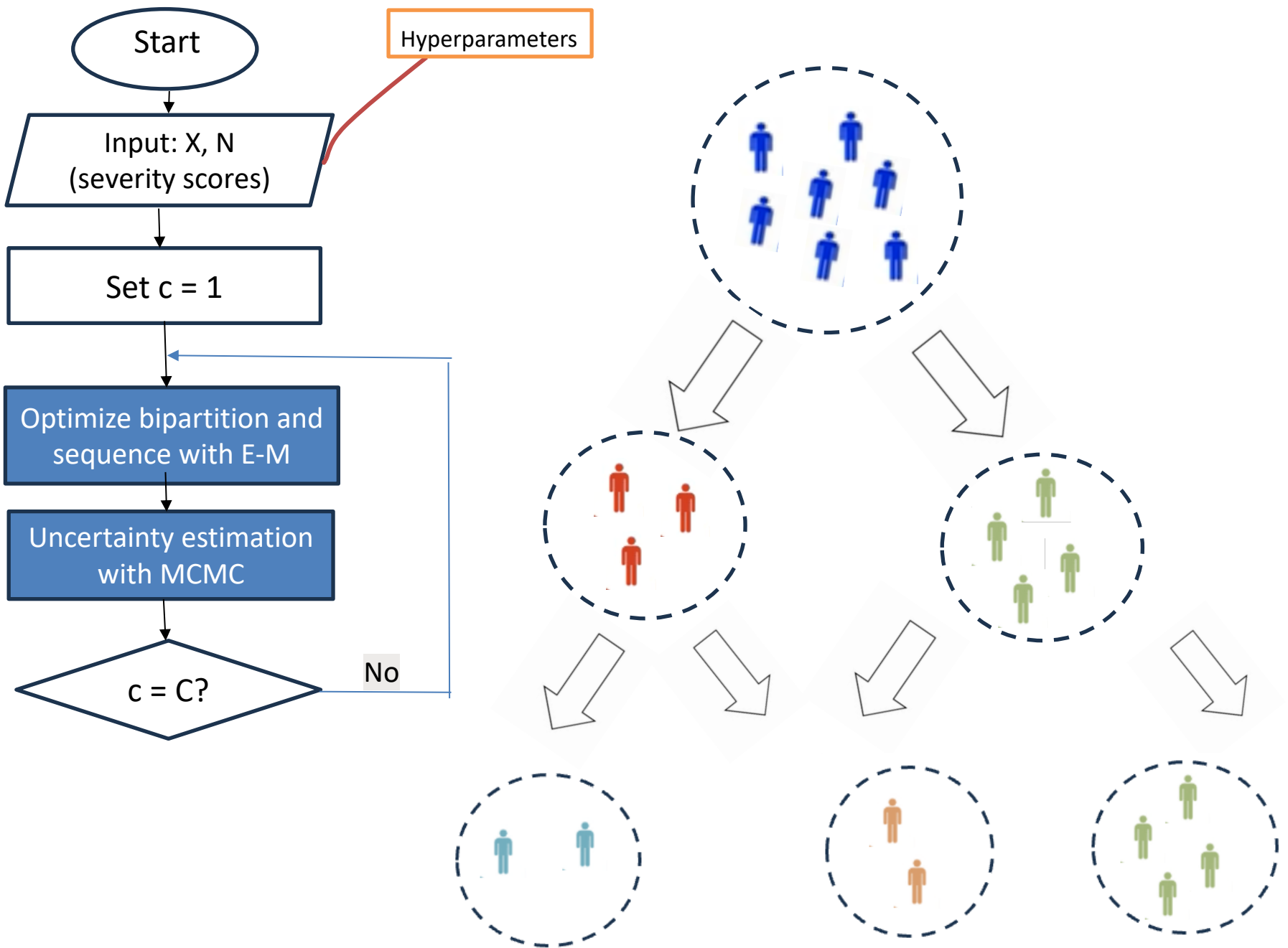
# Clustering

SuStain model  $\mathbf{M}$  with  $C$  clusters corresponding to sequences  $S_1, S_2, \dots, S_c$ , and fractions of the subjects that belong to each cluster  $f_1, f_2, \dots, f_c$



Hyperparameters





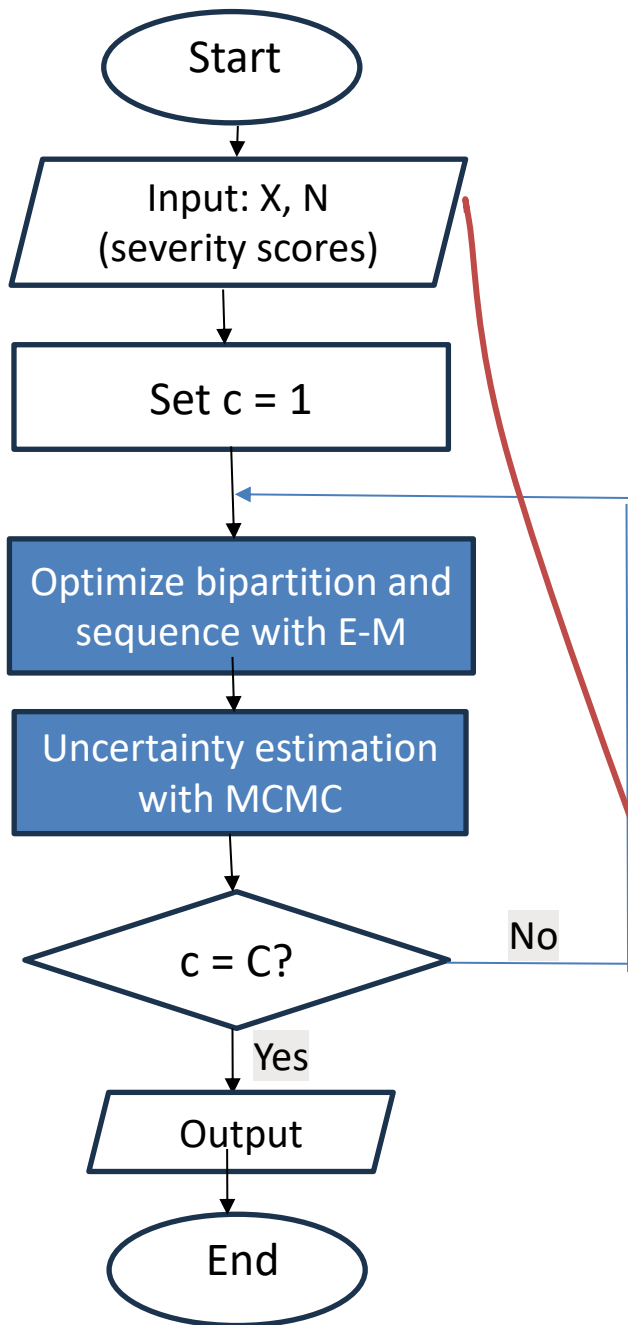
# Output

	Stage 1	Stage 2	Sum
Subtype 1			0.3
Subtype 2	0.1	0.4	0.5
Subtype 3			0.2

- Probabilities of individual subtype and stage assignment
- sequences for each subtype (with uncertainty)
- fractions of subjects belonging to each subtype

## Hyperparameters:

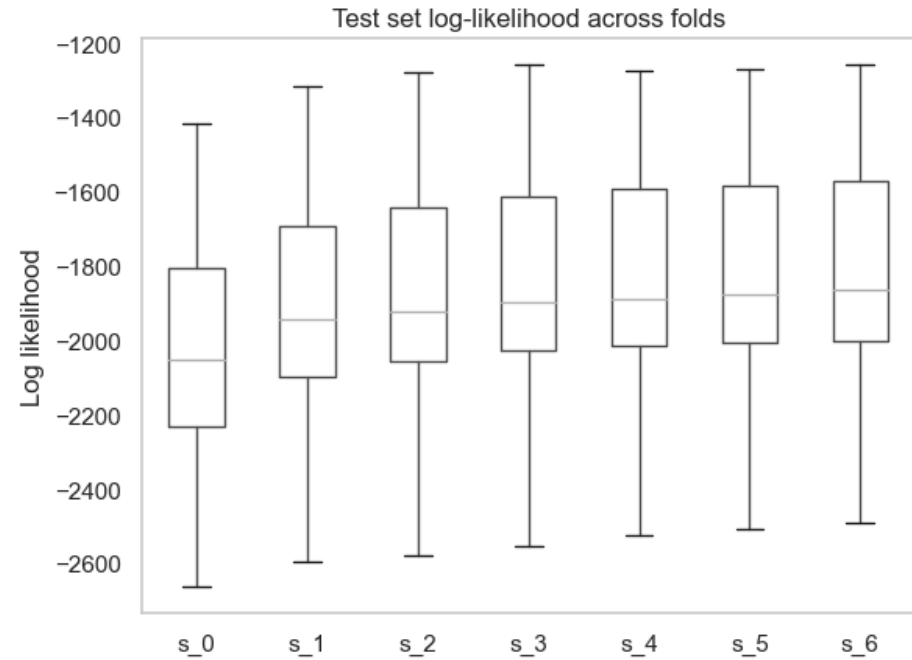
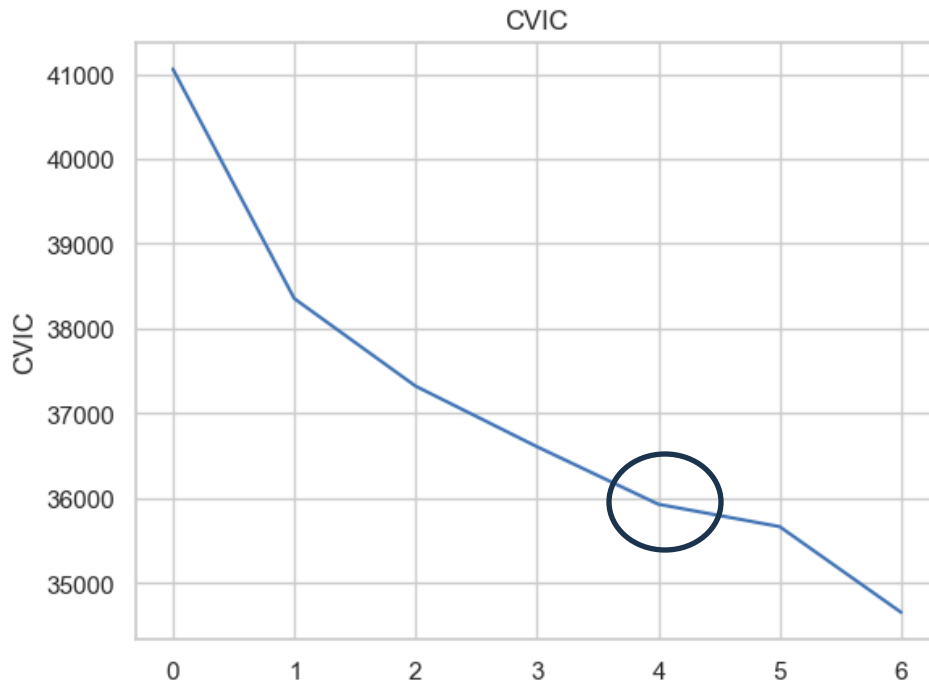
- number of start points (25) = random initializations of sequence estimation/subtype assignments
- number of MCMC iterations ( $10^5$ ) = # of different sequence orders examined to estimate uncertainty
- Number of Clusters ( $C = 7$ )



# Cross-validation

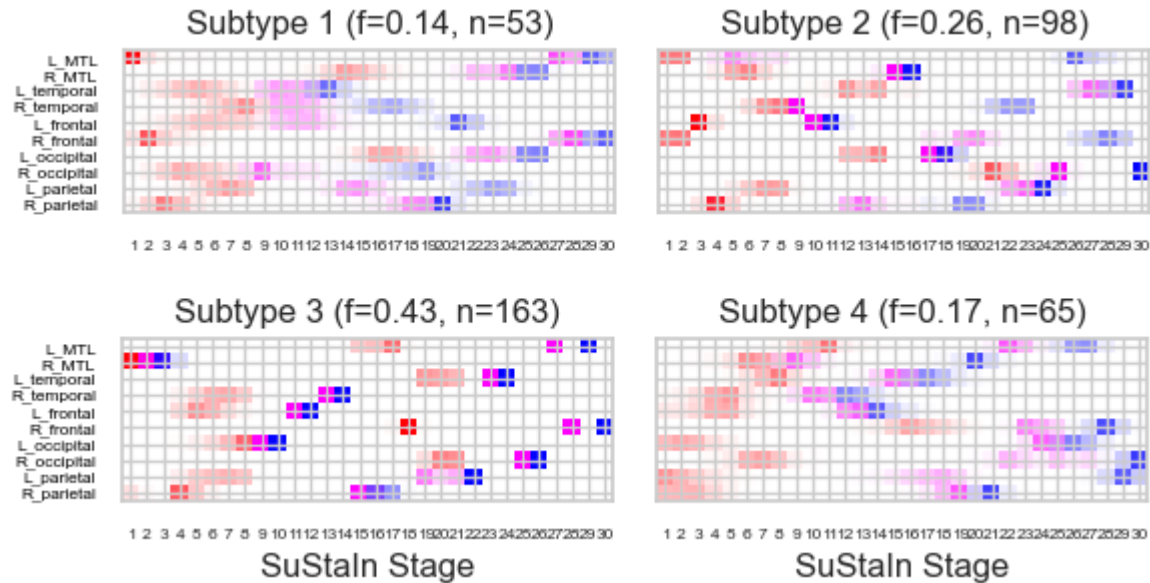
Number of subtypes ( $C = 1 \sim 7$ ) are decided using 10-fold cross-validation optimizing for the following criteria:

- Cross-validation information criteria
- Out-of-sample log likelihood
- Cross-validated maximum likelihood subtype probability



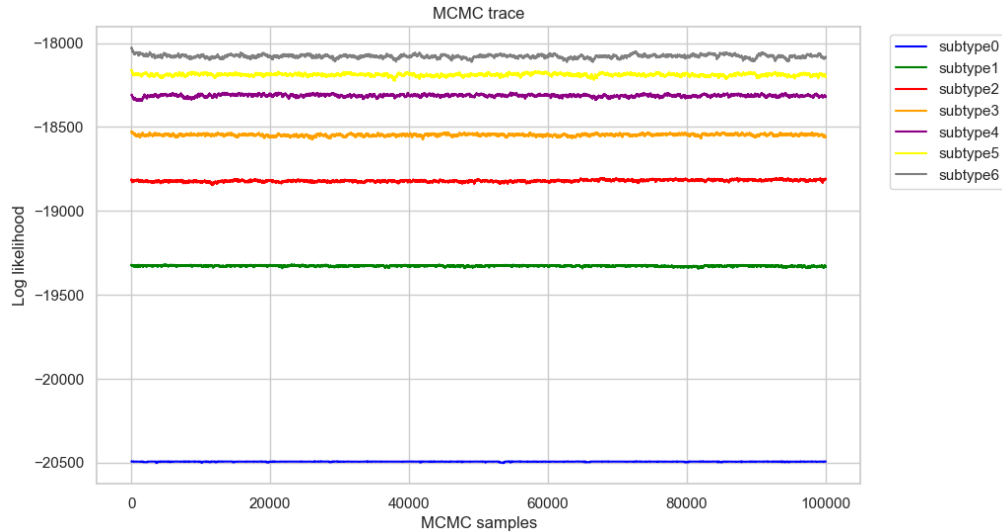
Number of splits

# Derived subtype and stage (C=5)

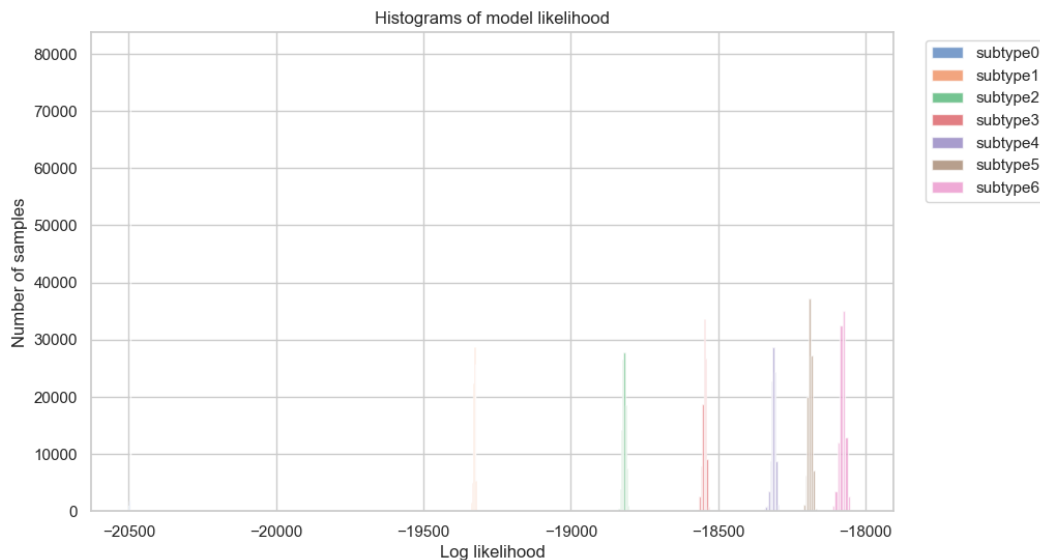


- Positional variance diagram shows most likely event order in which the regional tau distribution reaches a certain severity (red, magenta, and blue for increasing levels) and the uncertainty in the ordering.
- Subtype 3 is the subtype with the least blurry positional variance graph, which translates to confidence in a predominant disease progression trajectory;  
In contrast to subtype 1 and 4, for example.

# Model Assessment



- In each MCMC iteration a new set of parameter values (sequence orders) are evaluated and its resulting MLL is evaluated against the existing best.
- As the number of subtypes increases, the overall MLL also increase. And the MCMC traces show reasonable mixing property.



- Plotting distributions of the MLL derived from MCMC samples for each number of subtypes.
- The histograms again demonstrated as the number of subtypes increases, the MLL increases. The distributions are getting closer but no overlap is observed.

# To-do

- Severity level choice
- CV criterion, graph/table replications (sans post-hoc)
- Compare results with Vogel's SuStaln (Young's)
- Other ROI definitions, data transformation, clustering methods