



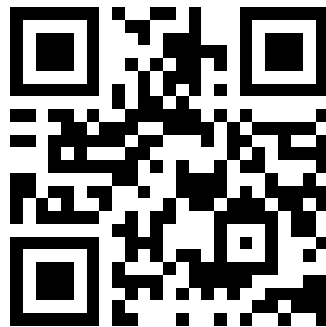
# mIVAR ou GIMME : les jeux sont faits, rien ne va plus !

## PREMIERE PARTIE

Jean-Luc Kop

Université de Lorraine, 2LPN-CEMA

<http://2lpn.univ-lorraine.fr/>



[https://frama.link/LDFf\\_gAW](https://frama.link/LDFf_gAW)

7<sup>e</sup> ATELIER MODEVAIIA, Sommières (Gard), 27-29 mai 2019



# Préambule

# Préambule

2015

Un nouveau continent ...

OPEN ACCESS Freely available online

PLOS ONE

## A Network Approach to Psychopathology: New Insights into Clinical Longitudinal Data

Laura F. Bringmann<sup>1\*</sup>, Nathalie Vissers<sup>1</sup>, Marieke Wichers<sup>2</sup>, Nicole Geschwind<sup>3</sup>, Peter Kuppens<sup>1</sup>, Frenk Peeters<sup>2</sup>, Denny Borsboom<sup>4</sup>, Francis Tuerlinckx<sup>1</sup>

<sup>1</sup> Department of Psychology, University of Leuven, Leuven, Belgium, <sup>2</sup> Department of Psychiatry and Neuropsychology, Maastricht University, Maastricht, The Netherlands, <sup>3</sup> Department of Clinical Psychological Science, Maastricht University, Maastricht, The Netherlands, <sup>4</sup> Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands



### Un modèle auto-régressif multivarié et multi-niveau de données ESM émotionnelles

Jean-Luc Kop<sup>1</sup>, Bruno Davier<sup>2</sup>, Sarah Le Vigouroux<sup>2</sup>, Jean-Baptiste Pavani<sup>2</sup> et Anne Congard<sup>2</sup>

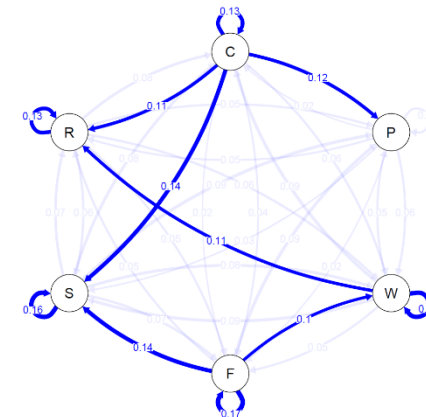
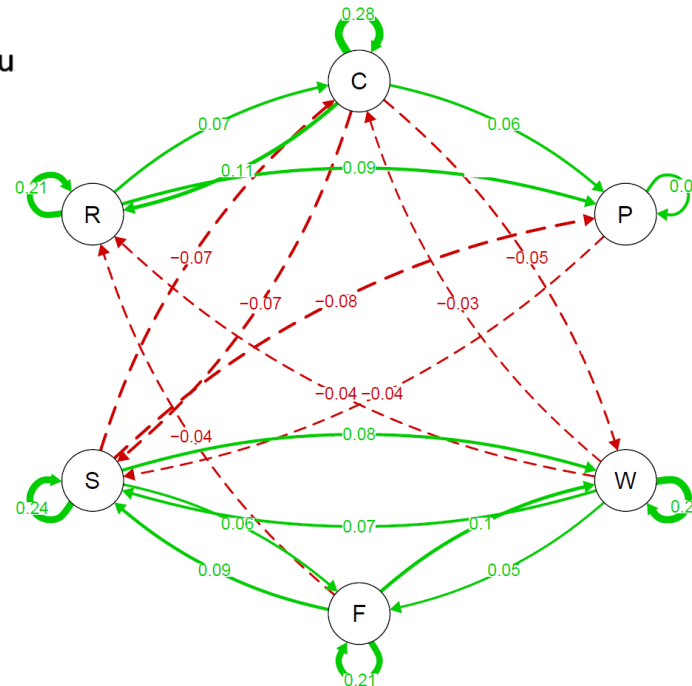
<sup>1</sup> Université de Lorraine

<sup>2</sup> Aix-Marseille Université

5<sup>e</sup> ATELIER MODEVAIIA, Camaret-sur-Mer - Presqu'île de Crozon (Finistère), 22-24 juin 2015

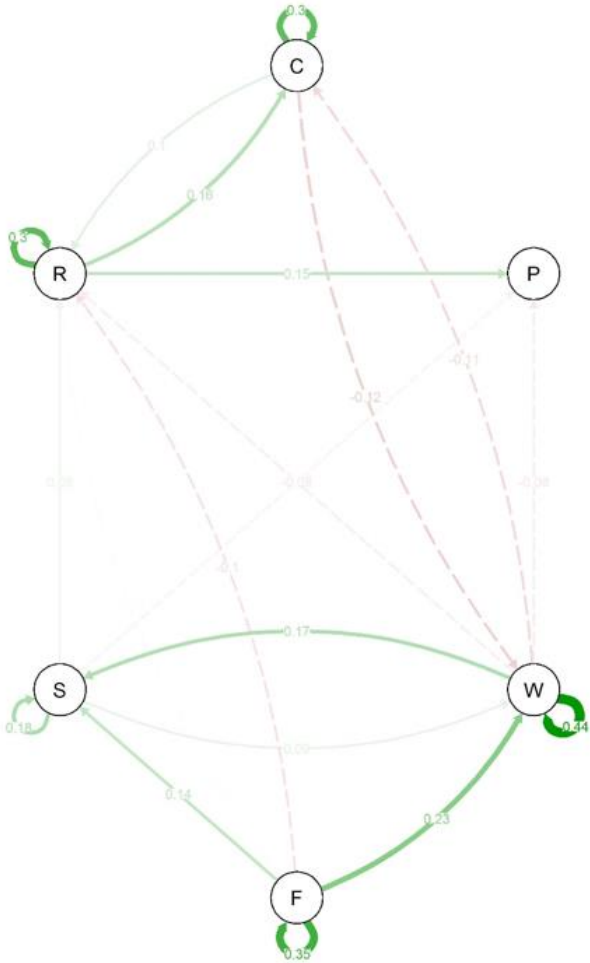
Cheerful  
Pleasant  
Worry

Fearful  
Sad  
Relaxed

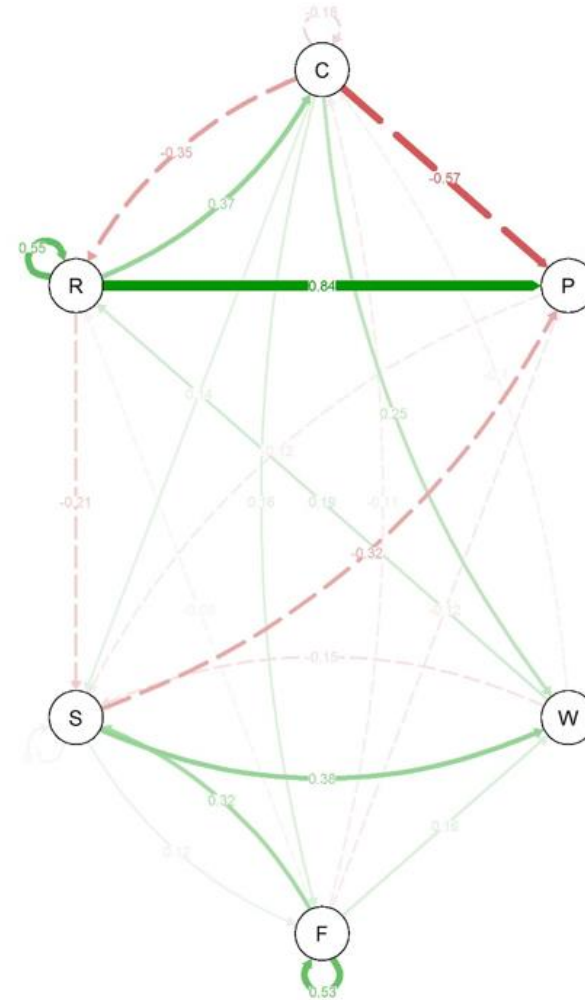


# Préambule

Mais déjà des doutes... Et les personnes dans tout cela ?  
Elles sont où ?



Réseau modélisé de l'individu 1 (en fonction de son écart au modèle moyen)



Réseau descriptif de l'individu 1 (en ne tenant pas compte du modèle moyen)

# Préambule

2017

Le temps des errements... et d'une idée simple sensée tout résoudre

- Pour « retrouver les personnes », c'est absurde de passer par un modèle moyen
  - Partir des réseaux dynamiques individuels (package « Graphical VAR »)
  - Construire des groupes typologiques sur la base de la similarité des réseaux individuels
- ➔ Modèle individuel et modèle(s) général(aux) au lieu de modèle moyen et écart au modèle moyen

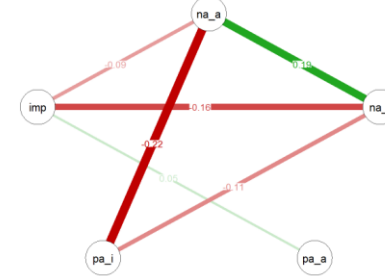
# Préambule

2017

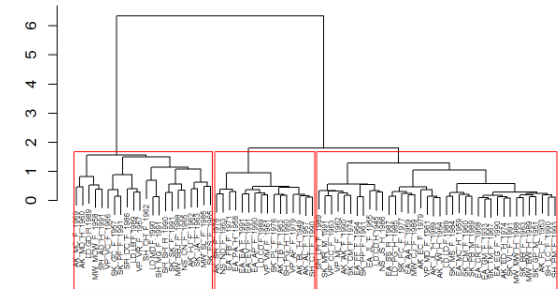
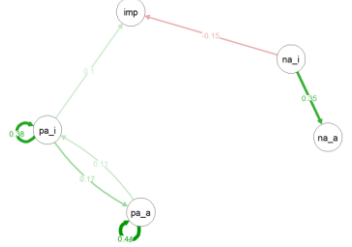
Où l'idée simple s'avère une usine à gaz dans sa mise en oeuvre

- Comment gérer les deux réseaux individuels (réseau contemporain et réseau décalé) ?
- Quel niveau de régularisation appliquer sur les réseaux individuels ?
- Quelle analyse en clusters utiliser ?
- Quelle méthode choisir pour décider du bon nombre de clusters ?
- ....

Réseau contemporain, période 1, participant SH\_LC\_F\_1990



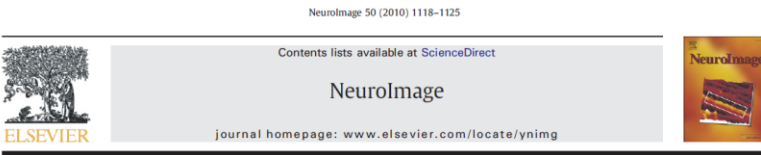
Réseau dirigé, période 1, participant NS\_NO\_F\_1963



# Préambule

2018

Où l'on s'aperçoit que cela ne sert à rien d'inventer la roue quand elle existe déjà...



Automatic search for fMRI connectivity mapping: An alternative to Granger causality testing using formal equivalences among SEM path modeling, VAR, and unified SEM

Kathleen M. Gates <sup>a,\*</sup>, Peter C.M. Molenaar <sup>a</sup>, Frank G. Hillary <sup>b</sup>, Nilam Ram <sup>a</sup>, Michael J. Rovine <sup>a</sup>

<sup>a</sup> Department of Human Development and Family Studies, The Pennsylvania State University, 5110 Henderson, University Park, PA, 16802, USA  
<sup>b</sup> Department of Psychology, The Pennsylvania State University, USA



Extended unified SEM approach for modeling event-related fMRI data

Kathleen M. Gates <sup>a,\*</sup>, Peter C.M. Molenaar <sup>a</sup>, Frank G. Hillary <sup>b</sup>, Semyon Slobounov <sup>c</sup>

<sup>a</sup> Department of Human Development and Family Studies, The Pennsylvania State University, University Park, PA, USA  
<sup>b</sup> Department of Psychology, The Pennsylvania State University, University Park, PA, USA  
<sup>c</sup> Department of Kinesiology, The Pennsylvania State University, University Park, PA, USA



Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples

Kathleen M. Gates <sup>a</sup>, Peter C.M. Molenaar

<sup>a</sup> Department of Human Development and Family Studies, Pennsylvania State University, USA



## Unsupervised Classification During Time-Series Model Building

Kathleen M. Gates, Stephanie T. Lane, E. Varangis, K. Giovanello & K. Guiskewicz

To cite this article: Kathleen M. Gates, Stephanie T. Lane, E. Varangis, K. Giovanello & K. Guiskewicz (2016): Unsupervised Classification During Time-Series Model Building, Multivariate Behavioral Research



## Automated Selection of Robust Individual-Level Structural Equation Models for Time Series Data

Stephanie T. Lane & Kathleen M. Gates

To cite this article: Stephanie T. Lane & Kathleen M. Gates (2017): Automated Selection of Robust Individual-Level Structural Equation Models for Time Series Data, Structural Equation Modeling: A Multidisciplinary Journal, DOI: [10.1080/10705511.2017.1309978](https://doi.org/10.1080/10705511.2017.1309978)



## Evaluating the Use of the Automated Unified Structural Equation Model for Daily Diary Data

Stephanie T. Lane & Kathleen M. Gates

To cite this article: Stephanie T. Lane & Kathleen M. Gates (2017) Evaluating the Use of the Automated Unified Structural Equation Model for Daily Diary Data, Multivariate Behavioral Research, 52:1, 126-127, DOI: [10.1080/00273171.2016.1265439](https://doi.org/10.1080/00273171.2016.1265439)



## Uncovering General, Shared, and Unique Temporal Patterns in Ambulatory Assessment Data

Stephanie T. Lane, Kathleen M. Gates, and Hallie K. Pike  
University of North Carolina at Chapel Hill

Adriene M. Beltz  
University of Michigan

Aidan G. C. Wright  
University of Pittsburgh



# Les modèles d'équations structurales unifiés (*uSEM*)

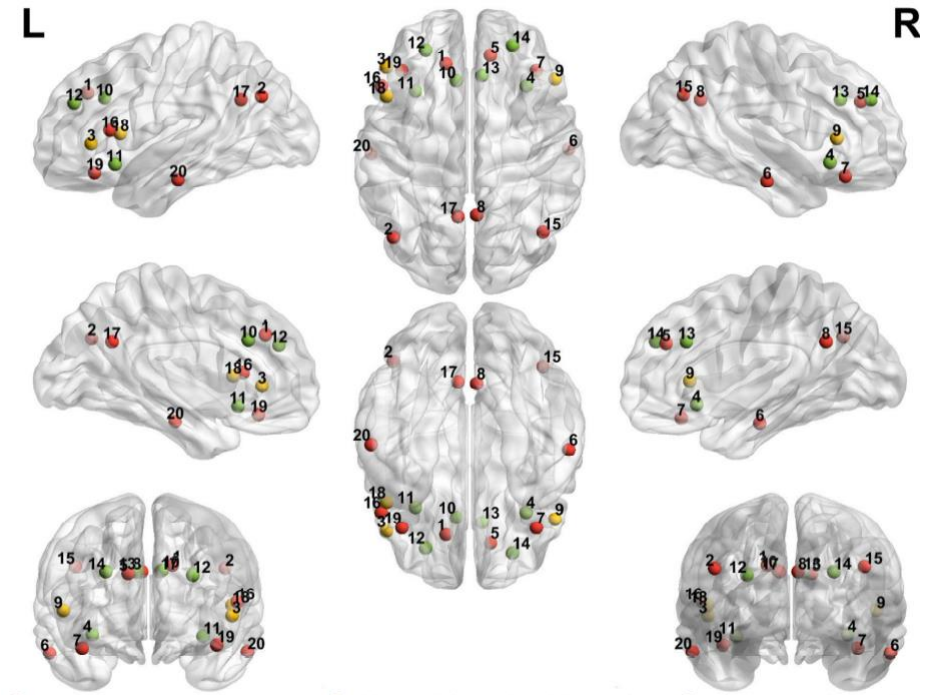


- Les données issues de l'enregistrement de l'activation de régions cérébrales par IRMf génèrent des séries temporelles

Plus d'une trentaine de méthodes existent (SEM [Structural Equations Model], DCM [Dynamic Causal Modeling], VAR [Vector Auto Regression], uSEM [unified Structural Equations Model] ...)

pour traiter ces données (Smith, S. M., Miller, K. L., Salimi-Khorshidi, G., Webster, M., Beckmann, C. F., Nichols, T. E., ... Woolrich, M. W. (2011). Network modelling methods for FMRI. *NeuroImage*, 54(2), 875-891. <https://doi.org/10.1016/j.neuroimage.2010.08.063>)

- Les modèles *uSEM* combinent modèles SEM et modèles VAR : les relations contemporaines (SEM) et décalées (VAR) y sont estimées *de manière simultanée*

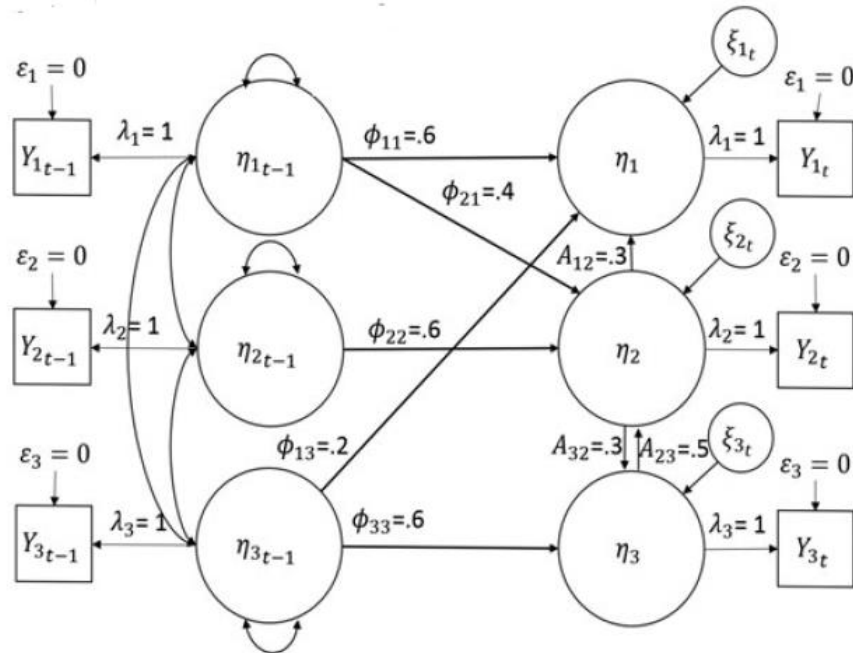


Henry, T. R., Feczko, E., Cordova, M., Earl, E., Williams, S., Nigg, J. T., ... Gates, K. M. (2019). Comparing directed functional connectivity between groups with confirmatory subgrouping GIMME. *NeuroImage*, 188, 642-653. <https://doi.org/10.1016/j.neuroimage.2018.12.040>

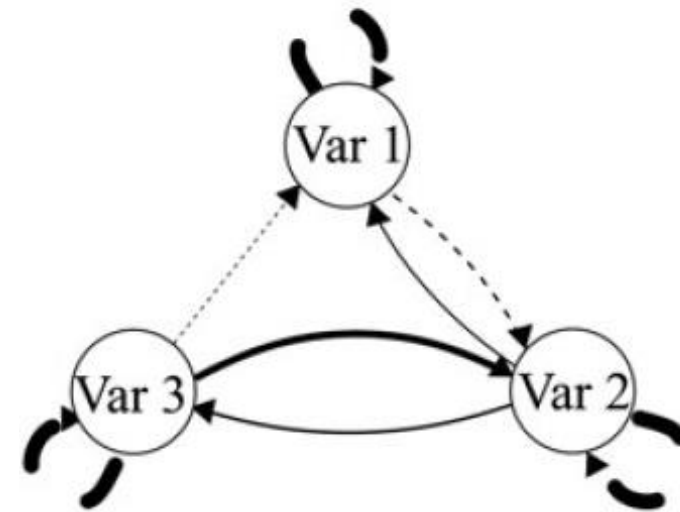
# Les modèles d'équations structurales unifiés (uSEM) :

(Kim, J., Zhu, W., Chang, L., Bentler, P. M., & Ernst, T. (2007). Unified structural equation modeling approach for the analysis of multisubject, multivariate functional MRI data. *Human Brain Mapping*, 28(2), 85-93. <https://doi.org/10.1002/hbm.20259>)

- Le modèle s'écrit :  $\eta_t = A\eta_t + \Phi\eta_{t-1} + \zeta_t$



Représentation détaillée d'un modèle uSEM (Gates et al., 2017)



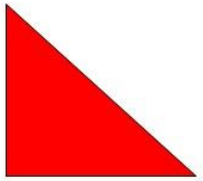
Lagged Effects ( $\Phi$ )

Contemp. Effects ( $A$ )

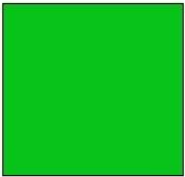
Représentation simplifiée d'un modèle uSEM (Gates et al., 2017)

# Les données analysées dans un *uSEM*

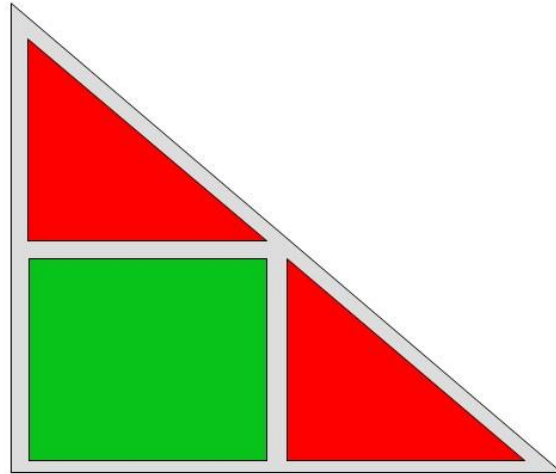
- Matrice *block-Toeplitz*



Covariances contemporaines (lag 0)



Covariances décalées (lag 1)



Matrice block-Toeplitz

## Les paramètres d'un *uSEM*

- Exemple avec 4 variables et un décalage temporel

$$B = \begin{pmatrix} & V_{1_{t-1}} & V_{2_{t-1}} & V_{3_{t-1}} & V_{4_{t-1}} & V_{1_t} & V_{1_t} & V_{3_t} & V_{4_t} \\ V_{1_{t-1}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ V_{2_{t-1}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ V_{3_{t-1}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ V_{4_{t-1}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ V_{1_t} & \beta_{1,5} & \beta_{2,5} & \beta_{3,5} & \beta_{4,5} & 0 & \beta_{6,5} & \beta_{7,5} & \beta_{8,5} \\ V_{2_t} & \beta_{1,6} & \beta_{2,6} & \beta_{3,6} & \beta_{4,6} & \beta_{5,6} & 0 & \beta_{7,6} & \beta_{8,6} \\ V_{3_t} & \beta_{1,7} & \beta_{2,7} & \beta_{3,7} & \beta_{4,7} & \beta_{5,7} & \beta_{6,7} & 0 & \beta_{8,7} \\ V_{4_t} & \beta_{1,8} & \beta_{2,8} & \beta_{3,8} & \beta_{4,8} & \beta_{5,8} & \beta_{6,8} & \beta_{7,8} & 0 \end{pmatrix}$$

$$B = \begin{pmatrix} & V_{1_{t-1}} & V_{2_{t-1}} & V_{3_{t-1}} & V_{4_{t-1}} & V_{1_t} & V_{1_t} & V_{3_t} & V_{4_t} \\ V_{1_{t-1}} & \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} \\ V_{2_{t-1}} & \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} \\ V_{3_{t-1}} & \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} \\ V_{4_{t-1}} & \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} & \boxed{0} \\ V_{1_t} & \beta_{1,5} & \beta_{2,5} & \beta_{3,5} & \beta_{4,5} & 0 & \beta_{6,5} & \beta_{7,5} & \beta_{8,5} \\ V_{2_t} & \beta_{1,6} & \beta_{2,6} & \beta_{3,6} & \beta_{4,6} & \beta_{5,6} & 0 & \beta_{7,6} & \beta_{8,6} \\ V_{3_t} & \beta_{1,7} & \beta_{2,7} & \beta_{3,7} & \beta_{4,7} & \beta_{5,7} & \beta_{6,7} & 0 & \beta_{8,7} \\ V_{4_t} & \beta_{1,8} & \beta_{2,8} & \beta_{3,8} & \beta_{4,8} & \beta_{5,8} & \beta_{6,8} & \beta_{7,8} & 0 \end{pmatrix}$$

Les 64 paramètres

$$\eta_t = A\eta_t + \Phi\eta_{t-1} + \zeta_t$$

A noter que les relations contemporaines sont dirigées (e.g.  $\beta_{5,6} \neq \beta_{6,5}$ ) et que ce n'est pas absurde :

- l'effet d'une variable sur une autre peut se produire dans un délai plus court que celui programmé pour le recueil des données (quasi systématique avec des données d'IRMf) ;
- il existe un effet décalé non mesuré sur deux variables ; cet effet non mesuré ne peut apparaître que sous forme d'une piste contemporaine.

## Estimation des paramètres

Gates, K. M., Molenaar, P. C. M., Hillary, F. G., Ram, N., & Rovine, M. J. (2010). Automatic search for fMRI connectivity mapping: An alternative to Granger causality testing using formal equivalences among SEM path modeling, VAR, and unified SEM. *NeuroImage*, 50(3), 1118-1125. <https://doi.org/10.1016/j.neuroimage.2009.12.117>

$$B = \begin{pmatrix} & V_{1,t-1} & V_{2,t-1} & V_{3,t-1} & V_{4,t-1} & V_{1,t} & V_{1,t} & V_{3,t} & V_{4,t} \\ V_{1,t-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ V_{2,t-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ V_{3,t-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ V_{4,t-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ V_{1,t} & \beta_{1,5} & \beta_{2,5} & \beta_{3,5} & \beta_{4,5} & 0 & \beta_{6,5} & \beta_{7,5} & \beta_{8,5} \\ V_{2,t} & \beta_{1,6} & \beta_{2,6} & \beta_{3,6} & \beta_{4,6} & \beta_{5,6} & 0 & \beta_{7,6} & \beta_{8,6} \\ V_{3,t} & \beta_{1,7} & \beta_{2,7} & \beta_{3,7} & \beta_{4,7} & \beta_{5,7} & \beta_{6,7} & 0 & \beta_{8,7} \\ V_{4,t} & \beta_{1,8} & \beta_{2,8} & \beta_{3,8} & \beta_{4,8} & \beta_{5,8} & \beta_{6,8} & \beta_{7,8} & 0 \end{pmatrix}$$

- Modèle initial = modèle nul (tous les paramètres sont fixés à 0)
- Calcul des multiplicateurs de Lagrange (i.e. indices de modification)
- Libération du paramètre ayant le multiplicateur de Lagrange le plus élevé
- Recalcul des multiplicateurs de Lagrange sur les autres paramètres
- Arrêt de la procédure lorsqu'aucun multiplicateur de Lagrange n'est statistiquement significatif



La recherche d'un modèle général tenant compte  
de l'hétérogénéité des modèles individuels

La procédure GIMME  
*(Group Iterative Multiple Model Estimation)*

## Comment agréger les données de modèles *uSEM* ?

Application d'*uSEM* à des données agrégées (moyennées)



Le modèle final est un modèle « moyen » qui peut ne correspondre à aucun individu

Agrégation « bottom up » à partir des modèles individuels



Les modèles individuels comportent une composante d'erreur importante qui peut amener à considérer certains réseaux différents alors qu'ils sont similaires

# La procédure GIMME est censée répondre à ces difficultés : modéliser de manière simultanée l'homogénéité et l'hétérogénéité

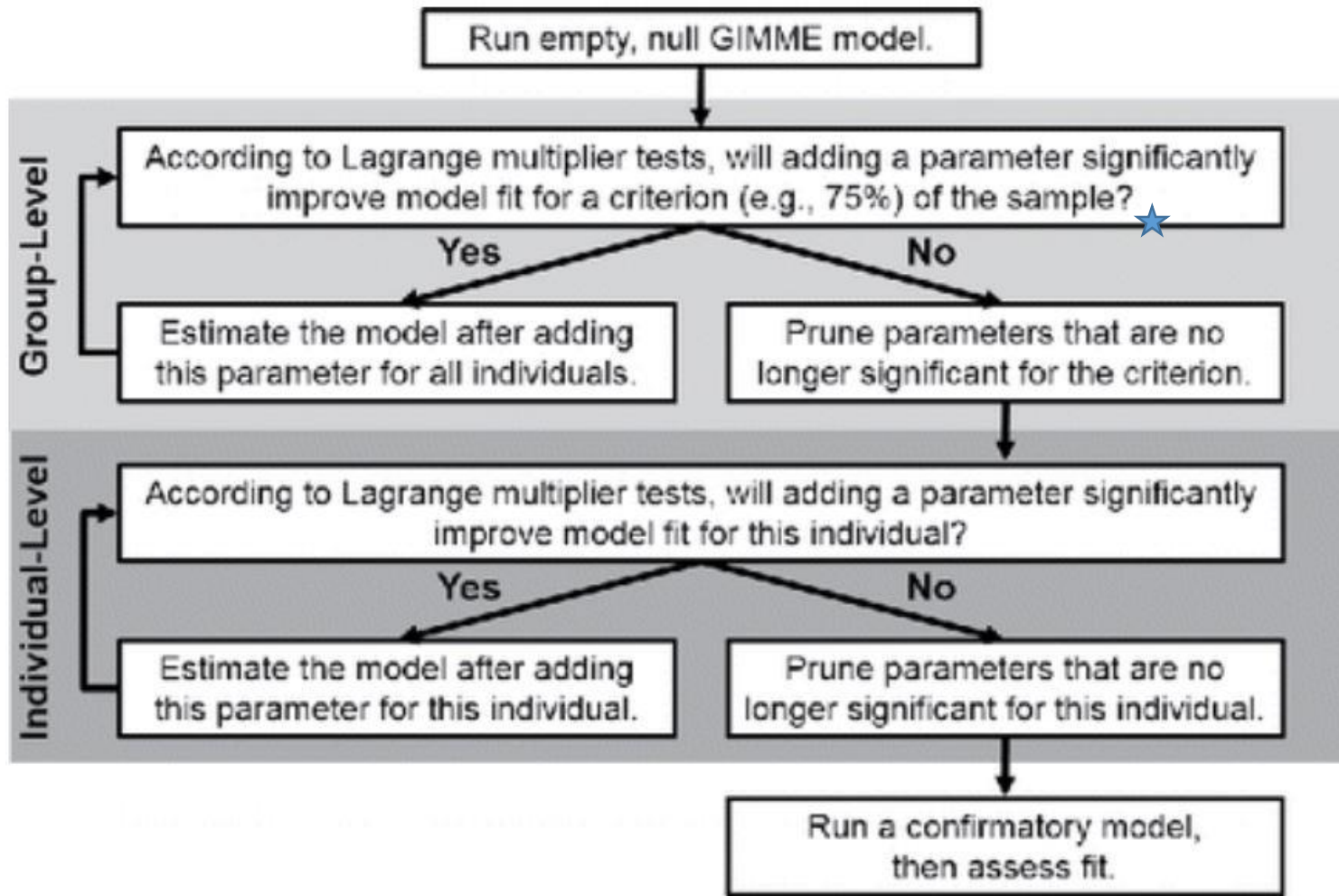
Gates, K. M., & Molenaar, P. C. M. (2012). Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage*, 63(1), 310-319. <https://doi.org/10.1016/j.neuroimage.2012.06.026>

- Objectif : identifier un modèle de groupe
- Modèle de groupe = modèle commun
- Modèle de groupe  $\leftrightarrow$  modèle moyen
- Le modèle de groupe est obtenu en retenant les paramètres du modèle uSEM qui améliorent l'adéquation de la majorité des modèles individuels

$$B = \begin{pmatrix} & V_{1_{t-1}} & V_{2_{t-1}} & V_{3_{t-1}} & V_{4_{t-1}} & V_{1_t} & V_{1_t} & V_{3_t} & V_{4_t} \\ V_{1_{t-1}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ V_{2_{t-1}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ V_{3_{t-1}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ V_{4_{t-1}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ V_{1_t} & \beta_{1,5} & \beta_{2,5} & \beta_{3,5} & \beta_{4,5} & 0 & \beta_{6,5} & \beta_{7,5} & \beta_{8,5} \\ V_{2_t} & \beta_{1,6} & \beta_{2,6} & \beta_{3,6} & \beta_{4,6} & \beta_{5,6} & 0 & \beta_{7,6} & \beta_{8,6} \\ V_{3_t} & \beta_{1,7} & \beta_{2,7} & \beta_{3,7} & \beta_{4,7} & \beta_{5,7} & \beta_{6,7} & 0 & \beta_{8,7} \\ V_{4_t} & \beta_{1,8} & \beta_{2,8} & \beta_{3,8} & \beta_{4,8} & \beta_{5,8} & \beta_{6,8} & \beta_{7,8} & 0 \end{pmatrix}$$



# Principe schématique de la procédure GIMME



$$B = \begin{pmatrix} & V_{1,t-1} & V_{2,t-1} & V_{3,t-1} & V_{4,t-1} & V_{1,t} & V_{1,t} & V_{3,t} & V_{4,t} \\ V_{1,t-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ V_{2,t-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ V_{3,t-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ V_{4,t-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ V_{1,t} & \beta_{1,5} & \beta_{2,5} & \beta_{3,5} & \beta_{4,5} & 0 & \beta_{6,5} & \beta_{7,5} & \beta_{8,5} \\ V_{2,t} & \beta_{1,6} & \beta_{2,6} & \beta_{3,6} & \beta_{4,6} & \beta_{5,6} & 0 & \beta_{7,6} & \beta_{8,6} \\ V_{3,t} & \beta_{1,7} & \beta_{2,7} & \beta_{3,7} & \beta_{4,7} & \beta_{5,7} & \beta_{6,7} & 0 & \beta_{8,7} \\ V_{4,t} & \beta_{1,8} & \beta_{2,8} & \beta_{3,8} & \beta_{4,8} & \beta_{5,8} & \beta_{6,8} & \beta_{7,8} & 0 \end{pmatrix}$$

- L'estimation des modèles individuels à partir du modèle général améliore la récupération correcte des paramètres individuels (séparation entre signal et bruit)
- La recherche de paramètres au niveau individuel s'arrête lorsque deux indicateurs globaux d'adéquation (RMSEA, NFI...) sont bons (→ contribue à la parcimonie des modèles)



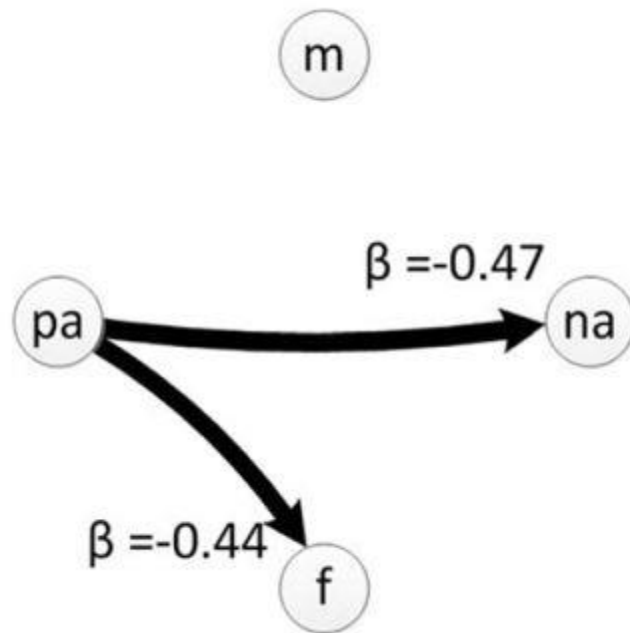
## Un exemple

Bouwmans, M. E. J., Beltz, A. M., Bos, E. H., Oldehinkel, A. J., de Jonge, P., & Molenaar, P. C. M. (2018). The person-specific interplay of melatonin, affect, and fatigue in the context of sleep and depression. *Personality and Individual Differences*, 123, 163-170. <https://doi.org/10.1016/j.paid.2017.11.022>

30 participants

3 mesures par jour pendant 30 jours

4 mesures : taux de mélatonine (m), affectivité positive (pa), affectivité négative (na) et fatigue (f)



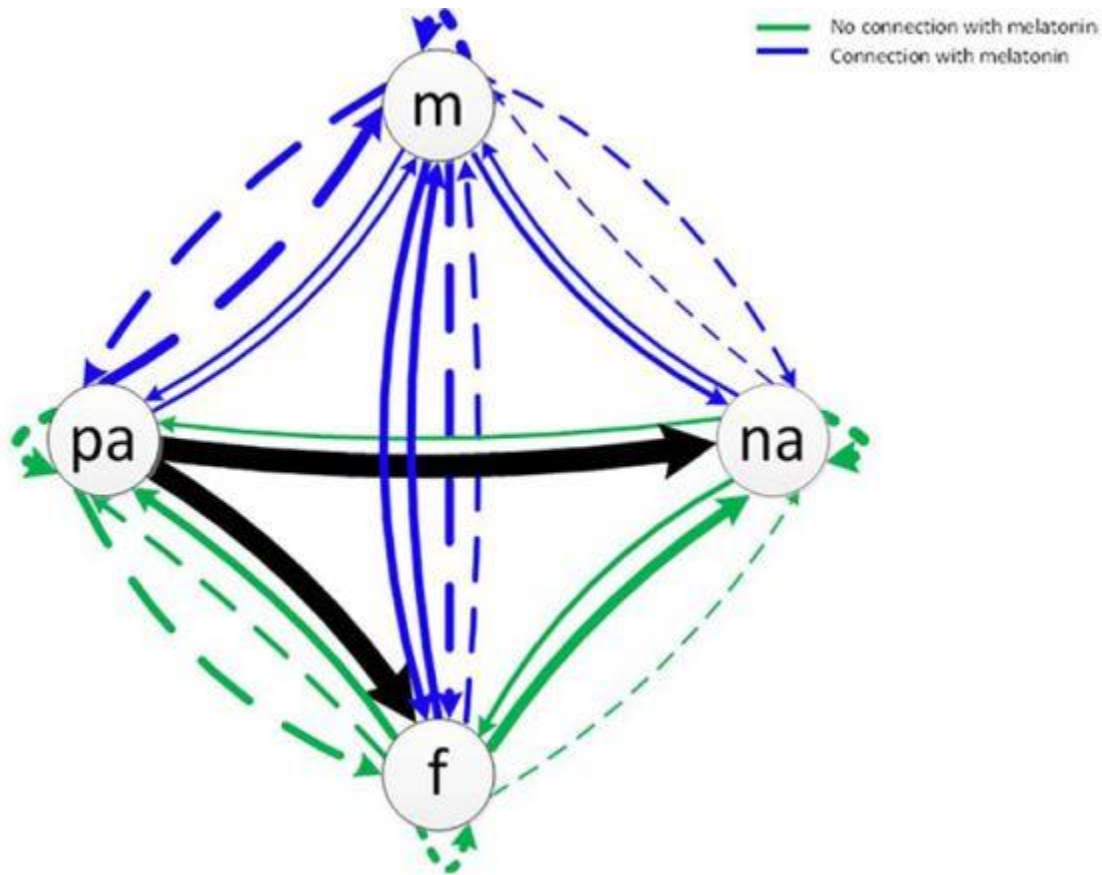
Modèle général

Traits pleins : relations contemporaines

$\beta$  : moyenne des coefficients individuels

# Un exemple

Bouwmans, M. E. J., Beltz, A. M., Bos, E. H., Oldehinkel, A. J., de Jonge, P., & Molenaar, P. C. M. (2018). The person-specific interplay of melatonin, affect, and fatigue in the context of sleep and depression. *Personality and Individual Differences*, 123, 163-170. <https://doi.org/10.1016/j.paid.2017.11.022>



Modèle général et modèles individuels  
Traits pleins : relations contemporaines  
Traits tiretés : relations décalées  
Traits noirs : relations présentes dans le modèle général  
Traits colorés : relations présentes dans les modèles individuels (la largeur des traits est proportionnelle au nombre de modèles individuels dans lesquels la piste est présente)



# Les extensions de la procédure GIMME

# Plusieurs extensions de la procédure GIMME

- **euSEM (extended uSEM) : modélisation conjointe d'effets exogènes variant dans le temps**  
Gates, K. M., Molenaar, P. C. M., Hillary, F. G., & Slobounov, S. (2011). Extended unified SEM approach for modeling event-related fMRI data. *NeuroImage*, 54(2), 1151-1158. <https://doi.org/10.1016/j.neuroimage.2010.08.051>
- **MS-GIMME (Multiple Solutions GIMME)**  
Beltz, A. M., & Molenaar, P. C. M. (2016). Dealing with multiple solutions in structural vector autoregressive models. *Multivariate Behavioral Research*, 51(2-3), 357-373. <https://doi.org/10.1080/00273171.2016.1151333>
- **Intégration de variables latentes (depuis version 0.6 du package [23 mai 2019])**  
<http://gimme.web.unc.edu/63-2/latent-variable-gimme-lv-gimme/>  
Gates, K. M., Fisher, Z. F., & Bollen, K. A. (2019). *Latent variable GIMME using model implied instrumental variables (MIIVs) (submitted)*. [<http://gimme.web.unc.edu/files/2019/05/LV-gimme-FINAL.pdf>]
- **S-GIMME (Subgrouping GIMME) : recherche de sous-groupes homogènes**  
Gates, K. M., Lane, S. T., Varangis, E., Giovanello, K., & Guiskewicz, K. (2017). Unsupervised classification during time-series model building. *Multivariate Behavioral Research*, 52(2), 129-148. <https://doi.org/10.1080/00273171.2016.1256187>
- **CS-GIMME (Confirmatory Subgrouping GIMME)**  
Henry, T. R., Feczko, E., Cordova, M., Earl, E., Williams, S., Nigg, J. T., ... Gates, K. M. (2019). Comparing directed functional connectivity between groups with confirmatory subgrouping GIMME. *NeuroImage*, 188, 642-653. <https://doi.org/10.1016/j.neuroimage.2018.12.040>

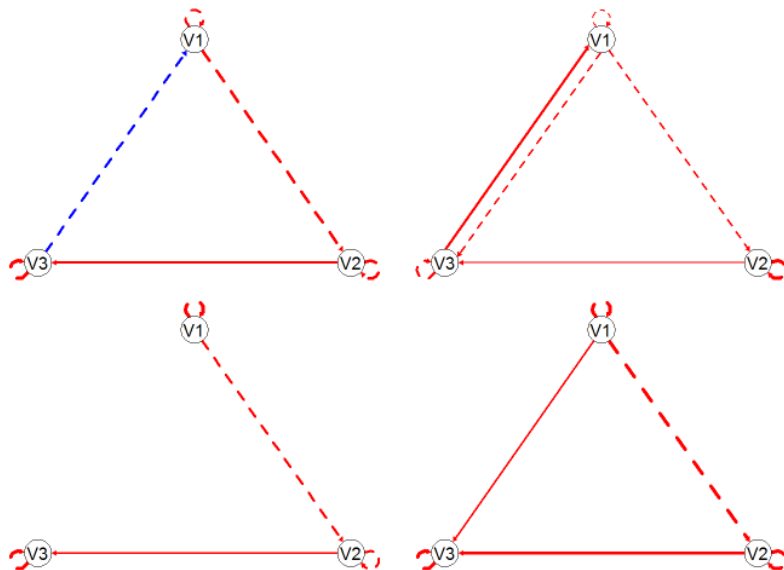
## **S-GIMME : recherche de sous-groupes homogènes**

- Rechercher des sous-groupes homogènes est un objectif en soi
- Mais identifier des sous-groupes homogènes permet aussi d'améliorer la fiabilité de l'estimation des modèles individuels (tout comme c'était le cas pour l'identification du modèle général)

## Construire une matrice de similarités entre les individus

- Il existe déjà une littérature conséquente quant aux caractéristiques à prendre en compte dans les séries temporelles pour évaluer leurs similarités (e.g. proximité entre les matrices block-Toeplitz) Liao, W. T. (2005). Clustering of time series data: A survey. *Pattern Recognition*, 38(11), 1857-1874. <https://doi.org/10.1016/j.patcog.2005.01.025>

- S-GIMME définit la proximité entre deux individus en termes **de nombre de pistes de même signe\*** entre les mêmes variables\*\*



\* Si une piste est présente dans le modèle général, mais non significative dans le modèle individuel, elle est considérée comme absente du modèle individuel

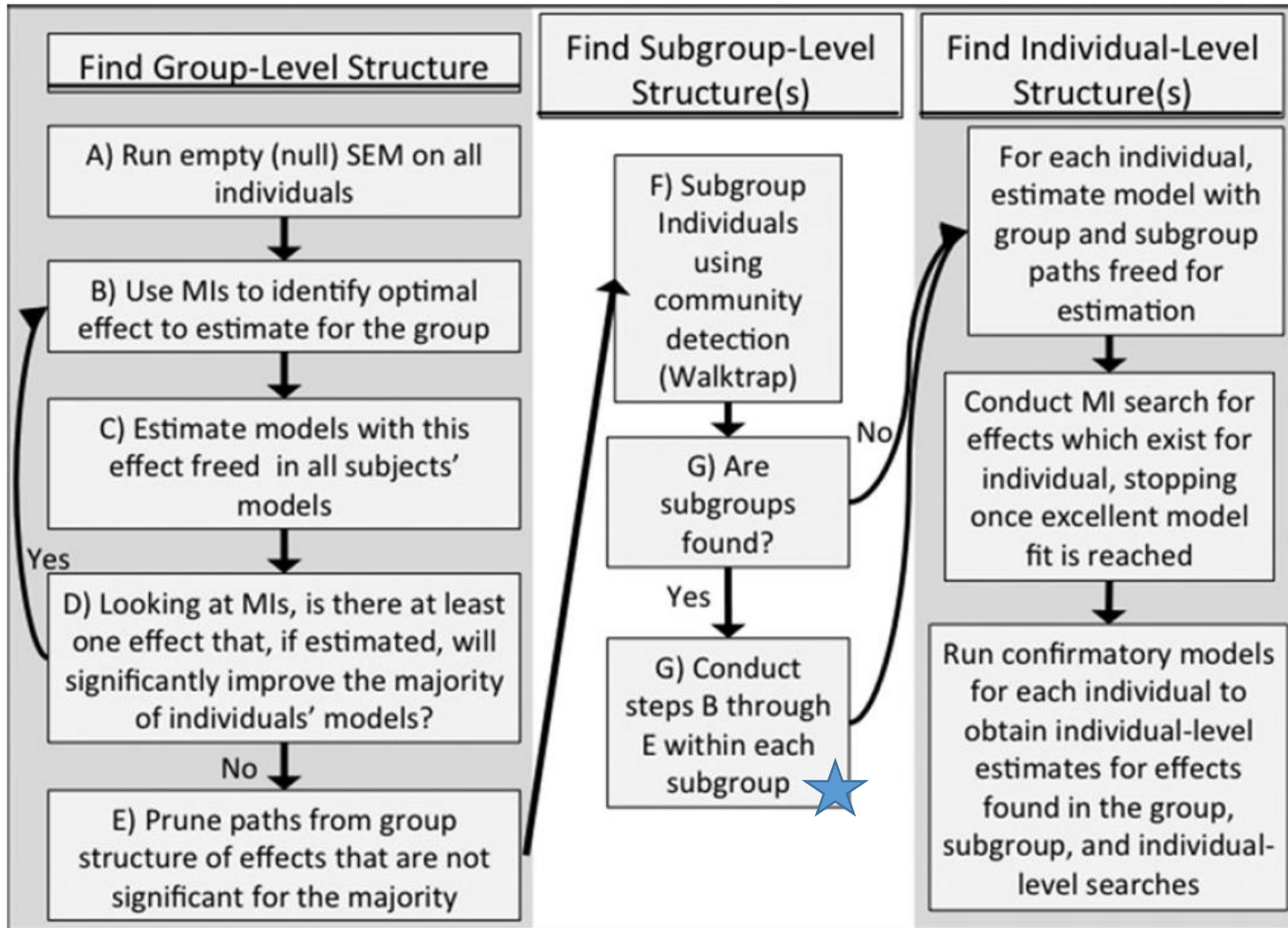
\*\* Pour éviter l'éparpillement, la valeur la plus faible de la matrice de proximités entre les individus est soustraite de toutes les valeurs

## Méthode de classification

- C'est la méthode *Walktrap* de détection de communautés qui est retenue comme méthode de classification Pons, P., & Latapy, M. (2006). Computing communities in large networks using random walks. *Journal of Graph Algorithms and Applications*, 10(2), 191-218. <https://doi.org/10.7155/jgaa.00124>
- Elle permet d'établir automatiquement le nombre optimal de clusters à partir d'un score dit de « modularité » (rapport entre la similarité intra-cluster par rapport à la similarité inter-clusters) Newman, M. E. J. (2006). Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23), 8577-8582. <https://doi.org/10.1073/pnas.0601602103>



# Principe schématique de la procédure S-GIMME



$$B = \begin{pmatrix} & V_{1,t-1} & V_{2,t-1} & V_{3,t-1} & V_{4,t-1} & V_{1,t} & V_{1,t} & V_{3,t} & V_{4,t} \\ V_{1,t-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ V_{2,t-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ V_{3,t-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ V_{4,t-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ V_{1,t} & \beta_{1,5} & \beta_{2,5} & \beta_{3,5} & \beta_{4,5} & 0 & \beta_{6,5} & \beta_{7,5} & \beta_{8,5} \\ V_{2,t} & \beta_{1,6} & \beta_{2,6} & \beta_{3,6} & \beta_{4,6} & \beta_{5,6} & 0 & \beta_{7,6} & \beta_{8,6} \\ V_{3,t} & \beta_{1,7} & \beta_{2,7} & \beta_{3,7} & \beta_{4,7} & \beta_{5,7} & \beta_{6,7} & 0 & \beta_{8,7} \\ V_{4,t} & \beta_{1,8} & \beta_{2,8} & \beta_{3,8} & \beta_{4,8} & \beta_{5,8} & \beta_{6,8} & \beta_{7,8} & 0 \end{pmatrix}$$



On fixe en général à 75% le seuil pour retenir une piste dans le modèle général et à 50% pour retenir une piste dans un modèle de sous-groupe

# Le modèle S-GIMME formalisé

$$B = \begin{pmatrix} V1_{t-1} & V2_{t-1} & V3_{t-1} & V4_{t-1} & V1_t & V1_t & V3_t & V4_t \\ V1_{t-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ V2_{t-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ V3_{t-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ V4_{t-1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ V1_t & \beta_{1,5} & \beta_{2,5} & \beta_{3,5} & \beta_{4,5} & 0 & \beta_{6,5} & \beta_{7,5} & \beta_{8,5} \\ V2_t & \beta_{1,6} & \beta_{2,6} & \beta_{3,6} & \beta_{4,6} & \beta_{5,6} & 0 & \beta_{7,6} & \beta_{8,6} \\ V3_t & \beta_{1,7} & \beta_{2,7} & \beta_{3,7} & \beta_{4,7} & \beta_{5,7} & \beta_{6,7} & 0 & \beta_{8,7} \\ V4_t & \beta_{1,8} & \beta_{2,8} & \beta_{3,8} & \beta_{4,8} & \beta_{5,8} & \beta_{6,8} & \beta_{7,8} & 0 \end{pmatrix}$$

$$\eta_{i,t} = (A_i + A_{i,k}^s + A_i^g)\eta_{i,t} + (\Phi_i + \Phi_{i,k}^s + \Phi_i^g)\eta_{i,t-1} + \zeta_{i,t}$$

i : individu

k : sous-groupe

g : modèle général

s : modèle du sous-groupe k

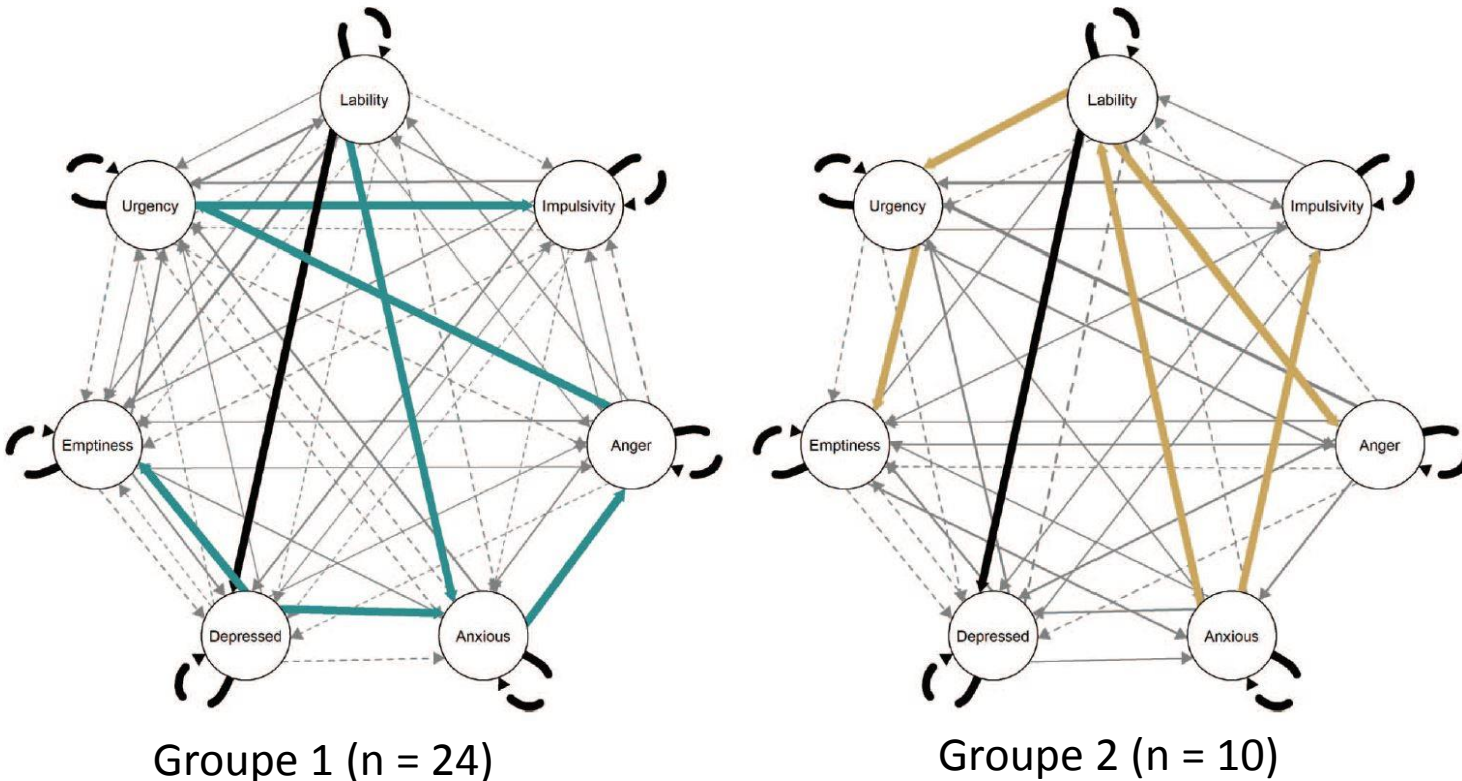
# Un exemple

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(1), 54-69. <https://doi.org/10.1037/met0000192>

36 participants

1 mesure par jour pendant 100 jours

7 mesures : mood lability (“My mood was up and down”) ; anxiousness (“I felt anxious”) ; depression (“I felt depressed”) ; anger (“I lost my temper”) ; impulsivity (“I did something on impulse”) ; emptiness (“My relationships felt empty”) ; urgency (“I acted on my emotions”)



Traits pleins : relations contemporaines  
Traits tiretés : relations décalées  
Traits noirs : modèle de groupe  
Traits en couleur : modèles de sous-groupe  
Traits gris : modèles individuels

\* Un individu constitue à lui seul un 3<sup>e</sup> groupe ; le 36<sup>e</sup> individu a disparu de l'article...



# Considérations méthodologiques

# Analyser, c'est bien. Faire attention aux pré-traitements, c'est mieux !

Bastiaansen, J. A., Kunkels, Y. K., Blaauw, F., Boker, S. M., Ceulemans, E., Chen, M., ... Bringmann, L. F. (2019). *Time to get personal? The impact of researchers' choices on the selection of treatment targets using the experience sampling methodology*. <https://doi.org/10.31234/osf.io/c8vp7>

## Time to get personal?

The impact of researchers' choices on the selection of treatment targets using the experience sampling methodology

### Authors

Jojanneke A. Bastiaansen<sup>1,2</sup>, Yoram K. Kunkels<sup>1</sup>, Frank J. Blaauw<sup>3,4</sup>, Steven M. Boker<sup>5</sup>, Eva Ceulemans<sup>6</sup>, Meng Chen<sup>7</sup>, Sy-Miin Chow<sup>7</sup>, Peter de Jonge<sup>3</sup>, Ando C. Emerencia<sup>3</sup>, Sacha Epskamp<sup>8</sup>, Aaron J. Fisher<sup>9</sup>, Ellen L. Hamaker<sup>10</sup>, Peter Kuppens<sup>6</sup>, Wolfgang Lutz<sup>11</sup>, M. Joseph Meyer<sup>5</sup>, Robert Moulden<sup>5</sup>, Zita Oravecz<sup>7</sup>, Harriette Riese<sup>1</sup>, Julian Rubel<sup>11</sup>, Oisín Ryan<sup>10</sup>, Michelle N. Servaas<sup>1</sup>, Gustav Sjöbeck<sup>5</sup>, Evelien Snippe<sup>1</sup>, Timothy J. Trull<sup>12</sup>, Wolfgang Tschacher<sup>13</sup>, Date C. van der Veen<sup>1</sup>, Marieke Wichers<sup>1</sup>, Phillip K. Wood<sup>12</sup>, William C. Woods<sup>14</sup>, Aidan G.C. Wright<sup>14</sup>, Casper J. Albers<sup>3</sup>, Laura F. Bringmann<sup>1,3</sup>

One of the promises of the experience sampling methodology (ESM) is that it could be used to identify relevant targets for treatment, based on a statistical analysis of an individual's emotions, cognitions and behaviors in everyday-life. A requisite for clinical implementation is that outcomes of person-centered analyses are not wholly contingent on the researcher performing them. To evaluate how much researchers vary in their analytical approach and to what degree outcomes vary based on analytical choices, we crowdsourced the analysis of one individual patient's ESM data to 12 prominent research teams, asking them what symptom(s) they would advise the treating clinician to target in subsequent treatment. The dataset was from a 25-year-old male with a primary diagnosis of major depressive disorder and comorbid generalized anxiety disorder, who completed momentary assessments related to depression and anxiety psychopathology prior to psychotherapy. Variation was evident at different stages of the analysis, from preprocessing steps (e.g., variable selection, clustering, handling of missing data) to the type of statistics. Most teams did include a type of vector autoregressive model, which examines relations between variables (e.g., symptoms) over time. Although most teams were confident their selected targets would provide useful information to the clinician, not one advice was similar: both the number (0-16) and nature of selected targets varied widely. This study makes transparent that the selection of treatment targets based on personalized models using ESM data is currently highly conditional on subjective analytical choices and highlights key methodological issues that need to be addressed in moving toward clinical implementation.

# Définition des seuils pour le modèle général et les modèles des sous-groupes

## Modèle général

- Les auteurs recommandent 75%
- A moduler en fonction du niveau d'hétérogénéité souhaité dans le modèle général

## Modèles des sous-groupes

- Les auteurs recommandent 50% (mais l'hétérogénéité est alors assez importante)

# Les coefficients d'auto-régression

- La procédure d'estimation débute généralement par un modèle nul
- Des études Monte Carlo semblent toutefois montrer qu'il est préférable de libérer toutes les auto-corrélations dans le modèle initial

## Le nombre de variables

- La plupart des études ont été réalisées avec un nombre de variables comprises entre 4 et 10
- Aucune raison toutefois de ne pas étendre (entre 3 et 20 ?)
- Mais les temps de calcul augmentent de manière exponentielle avec le nombre de variables
- Les études Monte Carlo donnent de meilleurs résultats avec un nombre de variables restreint



## Le nombre de participants

- Pas de contraintes fortes et pas de différences sensibles dans les études Monte Carlo
- Certains recommandent un minimum de 10 participants pour chercher un modèle général

# La longueur des séries temporelles

- Les auteurs recommandent au moins 60 observations
- Les performances de l'algorithme s'améliorent avec l'augmentation du nombre d'observations
- Avec des séries courtes, l'algorithme pêche à détecter des liens qui existent (c'est d'autant plus marqué qu'il y a peu de variables)
- La direction des pistes est très incertaine pour des séries courtes (certains considèrent qu'il faudrait au moins 200 observations)

# L'intervalle entre les moments temporels

- L'algorithme nécessite des intervalles égaux entre les moments temporels
- Si ce n'est pas le cas, possibilité de faire des interpolations ou de rééchantillonner les données
- Avec des données ESM, quid de l'intervalle entre la première mesure du matin et la dernière mesure de la veille ?

# La stationnarité des séries temporelles

- Stationnarité = propriétés statistiques (moyennes, variances, covariances) constantes au cours du temps
- La suppression des tendances existantes est un préalable indispensable
- C'est facile pour des tendances simples (ex : utiliser les résidus d'une régression [linéaire, polynomiale, GAM] en fonction du temps)
- C'est plus difficile avec des tendances plus complexes (touchant les covariances par exemple)

# L'ordre des décalages

- L'algorithme GIMME ne permet pour l'instant de ne prendre en compte que des décalages de 1 moment temporel (*lag 1*)
- La vérification de l'absence de décalages supérieurs à 1 peut se faire sur les résidus du modèle final (?? Comment accéder à ces résidus ??)
- Les décalages supérieurs à 1 peuvent être supprimés lors du pré-traitement (utilisation des résidus des fonctions d'auto-corrélations et de corrélations croisées)

## Données manquantes

- Les données manquantes sont tolérées par l'algorithme ; l'estimation se fait avec FIML (pas de problème si les données manquantes sont aléatoires...)
- On peut imputer les données manquantes (e.g. packages iVAR avec la technique des fenêtres glissantes ou AMELIA II qui tient compte de la structure dynamique des données)

# Normalité des distributions

- Les données initiales sont transformées en covariances ; celles-ci sont biaisées en cas de non-normalité
- Transformer les distributions est une option possible (e.g. package `bestNormalize`)

## Variances nulles

- Les variables avec variance nulle ne peuvent pas être conservées dans les analyses...
- ... et donc le participant qui a produit une variable sans variance est supprimé...



# Spécificités des données psychologiques par rapport à des données de neuro-imagerie (1)

- Emergence de variables avec faibles variances pour certains participants
- Séries temporelles moins longues
- Intervalles de temps non strictement équivalents
- Intervalles de temps plus espacés (2 s pour l'IRMf !) → effets auto-régressifs moins marqués
- Beaucoup plus d'hétérogénéité inter-individus que dans des processus biologiques

# Spécificités des données psychologiques par rapport à des données de neuro-imagerie (2)

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(1), 54-69. <https://doi.org/10.1037/met0000192>

- La performance de GIMME avec des séries temporelles présentant des spécificités de données psychologiques est un peu moins bonne
- Il est indispensable de libérer systématiquement les auto-corrélations au début de l'algorithme (en lieu et place d'un modèle nul)
- De critères statistiques moins stricts pourraient être utilisés pour décider d'inclure une piste dans le réseau