Personalized Network Modeling in Psychopathology: The Importance of Contemporaneous and Temporal Connections

Supplementary materials 3: Simulation Studies

Sacha Epskamp, Claudia D. van Borkulo, Date C. van der Veen, Michelle N. Servaas, Adela-Maria Isvoranu, Harriëtte Riese and Angelique O.J. Cramer

We have investigated two common questions related to estimating graphical VAR models. First, we have investigated if linear detrending of data prior to analyzing network models should be conducted, and second we have investigated how to best handle irregular spacing of observations due to missing observations throughout the nights. In both simulation studies, graphical VAR models with 7 nodes were generated as detailed by Yin & Li (2011, page 8), using a constant of 1.1 instead of 1.5 and simulating temporal and contemporaneous networks that were 75% sparse (In both networks each possible edge had a probability of 25% to be included). Generating such models can be done using the R function `randomGVARmodel` from the *graphicalVAR* package (Epskamp, 2017). Subsequently the VAR model can be simulated using the `graphicalVARsim` function. All simulations were performed using *graphicalVAR* version 0.2.1 and R version 3.3.1 on a Linux high performance cluster. We investigate the *sensitivity* (true positive rate), *specificity* (true negative rate) and correlation between network parameters, as is common in simulation studies on network estimation (Epskamp & Fried, 2017; van Borkulo et al., 2014). If sensitivity is low, edges are not picked up well, if specificity is low, spurious edges are detected, and if the correlation is low the estimated networks poorly reflect the true network structures.

Simulation study 1: Detrending data

In the first simulation study, we simulated responses from a graphical VAR model, and subsequently added a linear trend to each variable. Linear trends were generated for each variable from a normal distribution with mean 0 and standard deviation 0, 0.01, 0.1 or 1. As such, in the condition with a standard deviation of 0 there was no linear trend, and in the condition with a standard deviation of 1 we would expect strong linear trends for most variables. Sample size was varied between 100, 250 and 500 and detrending method was varied between no detrending, detrending only significant effects and detrending all variables. Linear detrending was performed as described by Fisher, Reeves, Lawyer, Medaglia, & Rubel

(2017). Each condition was replicated 100 times, leading to a total of 3,600 simulated datasets.

Figure S1 shows the results of the first simulation study. It can be seen that detrending all variables or only variables with a significant trend performed comparably, and have no effect when there is no true trend. Notably, not detrending leads to problematic networks when there are true trends: spurious edges are included in the temporal network (low specificity) and less true edges are detected in the contemporaneous network (low sensitivity). These results led to the decision of detrending variables with a significant trend in the paper.

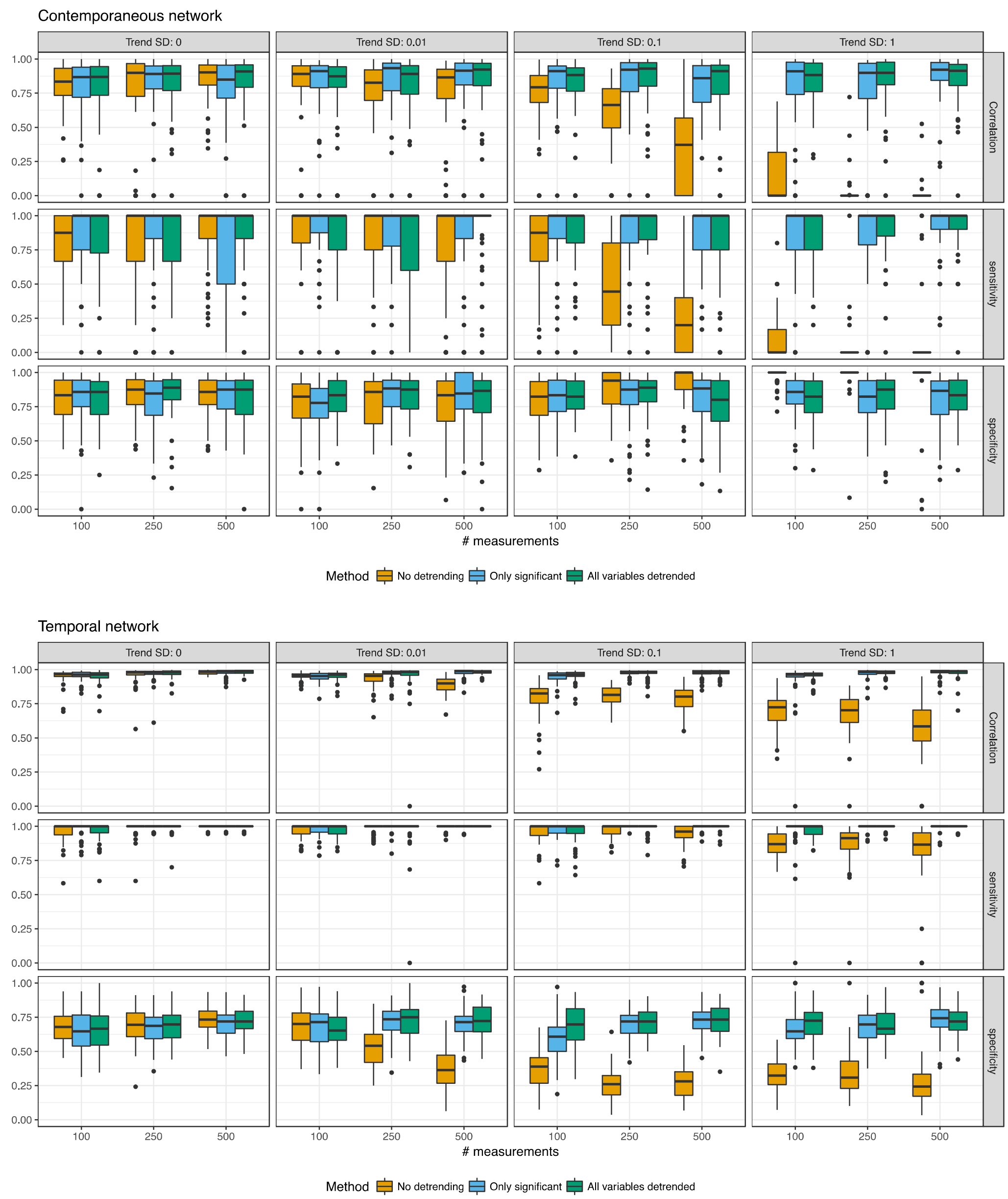

Figure S1. Results of the first simulation study.

Simulation study 2: Night effects

A second question we investigated was how to best handle the effects of missing observations throughout the night. The data we analyzed was gathered five times per day with roughly 3-hour intervals between measurements (interval queries were exactly spaced 3 hours apart, but the patient did not always respond immediately). To mimic this behavior in the simulation study, we simulated a VAR process for 5, 10, 25 or 50 days with eight realizations per day, and subsequently removed every sixth, seventh and eight measurements. This generated data with five observations per day (25, 50, 125 and 250 measurements respectively per condition). We varied the method for handling the night by either ignoring the night (regressing first observation of the day on the last observation of the previous day) or by using cubic spline interpolation as described by Fisher et al. (2017). Every condition was replicated 100 times, leading to a total of 1,200 generated datasets. Figure S2 displays the results, which show that removing nights had the best performance. This can be done in *graphicalVAR* using the `dayvar` argument.

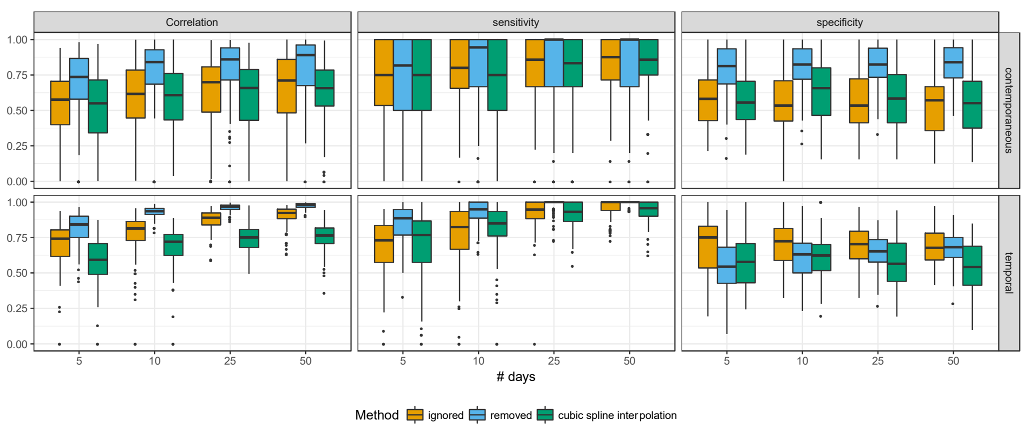

Figure S2. Results of the second simulation study.

References

Epskamp, S. (2017). graphicalVAR: Graphical VAR for Experience Sampling Data. Retrieved from http://cran.r-project.org/package=graphicalVAR

Epskamp, S., & Fried, E. I. (2017). A tutorial on Estimating Regularized Psychological Networks. Retrieved from http://arxiv.org/abs/1607.01367

Fisher, A. J., Reeves, J. W., Lawyer, G., Medaglia, J. D., & Rubel, J. A. (2017). Exploring the Idiographic Dynamics of Mood and Anxiety via Network Analysis. Retrieved from https://osf.io/5ybxt/

van Borkulo, C. D., Borsboom, D., Epskamp, S., Blanken, T. F., Boschloo, L., Schoevers, R. A., & Waldorp, L. J. (2014). A new method for constructing networks from binary data. *Nature: Scientific Reports*, *4*(5918). http://doi.org/10.1038/srep05918

Yin, J., & Li, H. (2011). A sparse conditional gaussian graphical model for analysis of genetical genomics data. *The Annals of Applied Statistics*, *5*(4), 2630–2650.