

Short thesis for the degree of doctor of philosophy (PhD)

**Robust Graphical Modeling of High-Dimensional
Gaussian and Non-Gaussian Processes: Theory and
Applications**

by **Abdaljbbar Babiker Abdaljbbar Dawod**

Supervisor: **Prof.em Dr. Terdik György**



UNIVERSITY OF DEBRECEN

Doctoral School of Informatics

Debrecen, 2025

Contents

1	Introduction	2
2	Problem Statement and Research Objectives	3
3	Methodology	4
4	Results	6
4.1	Thesis Group I: Theoretical and simulated results	6
4.1.1	Thesis 1: MSGH distribution	6
4.1.2	Thesis 2: MPE distribution	9
4.2	Thesis Group II: Applications of GLASSO . . .	11
4.2.1	Thesis 3: Cognitive-Behavioral Networks	11
4.2.2	Thesis 4: Process Quality Inspection . .	14
4.2.3	Thesis 5: Nutrients Interaction Network	18
	References	22

1. Introduction

Modeling dependencies in high-dimensional data is a central problem in statistical inference. Undirected graphical models represent conditional independence by the absence of edges. In Gaussian graphical models (GGMs), zero entries in the precision matrix correspond to conditional independence [24]. However, real-world data often deviates from Gaussian assumptions, exhibiting heavy tails, skewness, and nonlinear dependence structures. In many applications, the number of variables also exceeds the number of observations.

Traditional maximum likelihood estimators yield singular covariance matrices in high-dimensional graphical models. Regularization methods like the ℓ_1 -penalized GLASSO solve this by shrinking entries to zero, guaranteeing consistent sparse estimation. While originally constrained by Gaussian assumptions, robust modern extensions—including elliptical distributions, copulas, and semiparametric frameworks—now successfully accommodate a wide class of non-Gaussian processes [25].

Several extensions have been proposed to tackle this issue by incorporating other distributions. For instance, preserving the covariance-based dependence structure when having heavier-tailed processes by using an elliptical distribution [10]. Also, semiparametric and copula-based graphical models soften the normality assumptions by modelling the marginal distributions separately from the dependence structure [15].

The main publications related to this thesis are the following:

[J1, J2, J3, J4, C1, C2].

2. Problem Statement and Research Objectives

Current graphical models exhibit critical limitations: they rely on fragile Gaussian assumptions, lack clear theoretical links to conditional independence for non-Gaussian data, and fail to provide a unified computational framework. Furthermore, practical applications demand methods that remain robust despite data deficiency. This dissertation addresses these gaps. Our objective is to achieve robust estimation and application of high-dimensional graphical models across both Gaussian and non-Gaussian environments.

Research Objectives

- **Evaluate** the theoretical link between the GLASSO framework and conditional independence.
- **Extend** GLASSO to model non-Gaussian data, specifically Multivariate Symmetric Generalized Hyperbolic (MSGH) and Multivariate Power Exponential (MPE) distributions.
- **Apply** these approaches to improve the detection sensitivity of statistical process control (SPC) charts.
- **Characterize** complex network structures in behavioral science applications.

- **Map** nutrient interaction networks in baobab leaves to advance nutritional profiling in food systems.

3. Methodology

For multivariate normal distributions, partial correlation strictly dictates conditional independence. Kelker [13] and Baba et al. [17] explored this relationship across broader elliptically symmetric families. They proved a critical limitation: zero partial correlation guarantees conditional independence if and only if the data are multivariate normal. For all non-Gaussian elliptical distributions, it merely implies zero conditional correlation.

In the context of elliptical distributions, a zero partial correlation implies zero conditional correlation, since the best predictor is linear. This result follows from the linearity of the best predictor and is slightly more general than the corresponding Gaussian case.

Lemma 1 *Assume $\mathbf{Y} = [Y_i, Y_j, \mathbf{Y}_{\setminus ij}]$ is a multivariate process, where $\mathbf{Y} = \mathbf{X} - \mu_{\mathbf{X}}$ for notational convenience. Considering that the best prediction of Y_i with respect to $\mathbf{Y}_{\setminus i}$ is linear that is,*

$$\mathbb{E}(Y_i | \mathbf{Y}_{\setminus i}) = \hat{Y}_{i|\mathbf{Y}_{\setminus i}}.$$

If $Y_i \perp Y_j \nmid \mathbf{Y}_{\setminus ij}$, then $Y_i \perp Y_j | \mathbf{Y}_{\setminus ij}$. In other words if Y_i and Y_j are partially uncorrelated, then they are conditionally uncorrelated as well.

EM-GLASSO_H estimator: For a Gaussian vector $\mathbf{X} \sim \mathcal{N}_d(\boldsymbol{\mu}, \Sigma)$, the standard GLASSO estimator $\hat{\Theta}$ relies on the sample covariance matrix $\hat{\mathbf{S}}_{\mathbf{X}}$:

$$\hat{\Theta} = \arg \max_{\Theta} \left(\frac{n}{2} \log |\Theta| - \frac{n}{2} \text{tr}[\Theta \hat{\mathbf{S}}_{\mathbf{X}}] - \rho \|\Theta\|_1 \right).$$

Under normal variance mixture models with an unobserved latent weight variable τ , we estimate parameters via the EM algorithm. We substitute τ with its exact conditional expectation $\hat{\tau} = \mathbb{E}(\tau \mid \mathbf{X})$:

$$\hat{\tau}(\mathbf{x}) = \sqrt{\frac{\psi}{\chi + \delta(\mathbf{x})}} \frac{K_{\frac{d}{2}-\lambda+1} \left(\sqrt{\psi(\chi + \delta(\mathbf{x}))} \right)}{K_{\frac{d}{2}-\lambda} \left(\sqrt{\psi(\chi + \delta(\mathbf{x}))} \right)}.$$

We initialize this EM algorithm by computing $\hat{\Theta}_0$ using the standard GLASSO estimator.

E-step. Use the current estimate $\hat{\Theta}_k$ to calculate $\hat{\tau}_i, i = 1 : n$, put

$$\hat{\boldsymbol{\mu}}_k = \frac{1}{\sum_{i=1}^n \hat{\tau}_i} \sum_{i=1}^n \hat{\tau}_i \mathbf{x}_i, \quad \hat{\mathbf{S}}_{\mathbf{X}, \tau} = \frac{1}{n} \sum_{i=1}^n \hat{\tau}_i (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_k) (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_k)^\top,$$

M-step. Obtain the next estimate $\hat{\Theta}_{i+1}$ by

$$\hat{\Theta}_{i+1} = \arg \max_{\Theta} \left(\frac{n}{2} \log |\Theta| - \frac{n}{2} \text{tr} \left[\Theta \hat{\mathbf{S}}_{\mathbf{X}, \tau} \right] - \rho \|\Theta\|_1 \right).$$

The iteration of these E- and M-steps will be referred to as the EM-GLASSO algorithm.

EM-GLASSO_p estimator

$$\mathbb{E} \tau | \mathbf{X} = \beta \delta(\mathbf{x})^{\beta-1}. \quad (1)$$

This result yields the best predictor of the weight variable τ given the observation $\mathbf{x}$. Accordingly, the resulting estimator is termed EM-GLASSO_p and is implemented via the EM-algorithm, using the weight specified in (1).

4. Results

4.1. Thesis Group I: Theoretical and simulated results

We evaluated EM-GLASSO, GLASSO, γ -LASSO, and SGLASSO via average ROC curves generated across 200 replications. We applied a 10^{-3} cutoff to define the true graph and parameterized the baselines using standard literature: $\lambda = 10^{-5}$ for SGLASSO [16] and $\gamma = 0.05$ for γ -LASSO [21]. Because EM-GLASSO performs GLASSO optimization at each M-step, it incurs a higher computational cost. To guarantee reproducibility and bypass likelihood-based limitations, we tuned our model using StARS, which robustly isolates the graph structure by measuring edge stability across subsamples.

4.1.1. Thesis 1: MSGH distribution

Case 1: $\mathcal{MSGH}_d, d = 50$; **sparsity** = 30%; $n = 100$: We simulated datasets of size $n = 50$ and $n = 100$ using the MSGH distribution: $\mathcal{MSGH}_d(\lambda, \chi = 0.5, \psi = 0.5, \mu = 0, \Sigma = \Theta^{-1})$,

testing the shape parameters $\lambda \in \{2, 5, 10, 15\}$.

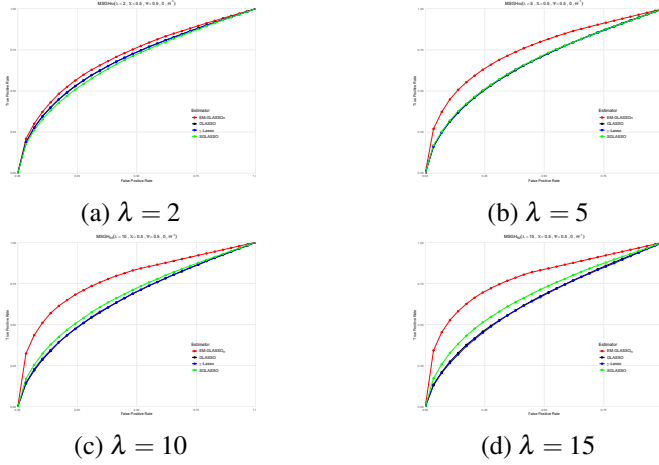


Figure 1: ROC curves for the GLASSO, EM-GLASSO_H, γ -LASSO, under $\mathcal{MSGH}_{50}(\lambda, \chi = 0.5, \psi = 0.5, \mu = 0, \Sigma = \Theta^{-1})$ distribution.

Figure 1 compares the ROC curves for the four estimators. At $\lambda = 2$, all methods exhibit tightly clustered, comparable recovery of the true sparsity pattern. However, as λ increases (5, 10, 15), EM-GLASSO_H strictly dominates, achieving the highest True Positive Rate (TPR) across all False Positive Rates (FPR). SGLASSO ranks second. Conversely, GLASSO and γ -LASSO degrade to nearly identical, lower-tier performance, clustering at the bottom of the ROC space.

Case 2: $\mathcal{MSGH}, d = 100$; **sparsity**= 10%; $n = 50$: The result in Figure 2 illustrates estimator performance in high-dimensional ($d > n$) MSGH settings. At $\lambda = 2$, all methods recover the true

sparsity pattern comparably well. However, as λ increases (5, 10, 15), heavier tails emerge, and EM-GLASSO_H strictly dominates the ROC space. It maximizes the TPR for any given False Positive Rate. SGLASSO consistently ranks second. GLASSO and γ -LASSO fail under these heavy-tailed conditions, clustering together with inferior sensitivity.

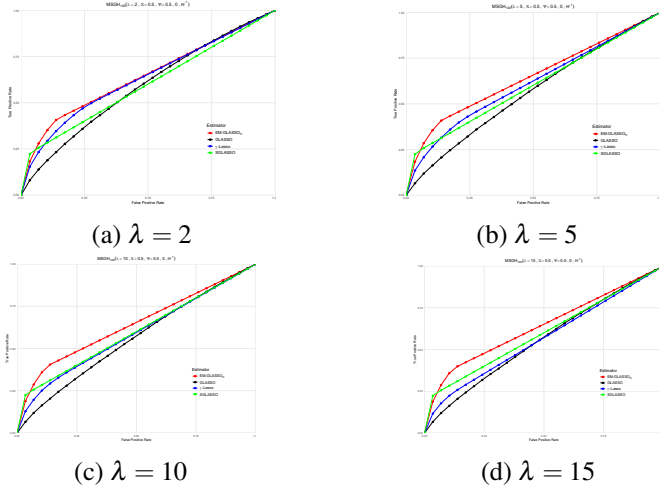


Figure 2: ROC curves of GLASSO, emLAASSO_H, γ -LASSO, and SGLASSO under $\mathcal{MSGH}_{100}(\lambda, \chi = 0.5, \psi = 0.5, \mu = 0, \Sigma = \Theta^{-1})$ with $n = 100$.

Objective: Extend the GLASSO framework under the MSGH distribution while explicitly preserving interpretable conditional relationships.

Key Findings: EM-GLASSO_H accurately models heavy-tailed MSGH distributions to extract interpretable networks. Get the open-source R code here: <https://github.com/AbdoDawod/emLASSO>.

Related publication: 3.

4.1.2. Thesis 2: MPE distribution

Case 1: $\mathcal{MPE}, d = 50$; sparsity= 30%; $n = 100$: We simulated datasets ($n \in \{50, 100\}$) from the $\mathcal{MPE}_{50}(0, \Theta^{-1}, \beta)$ distribution to evaluate variations in the shape parameter ($\beta \in \{1/3, 1/2, 2/3, 3/4, 1\}$). Figure 3 illustrates the resulting estimator performance. EM-GLASSO_P strictly dominates the ROC space across all non-Gaussian settings ($\beta < 1$), maximizing the TPR for any given False Positive Rate. SGLASSO consistently ranks second. Meanwhile, GLASSO and γ -LASSO fail to accommodate the heavier tails, clustering at the bottom with inferior sensitivity. As expected, when $\beta \rightarrow 1$ (the Gaussian limit), this performance gap naturally closes, and EM-GLASSO_P performs comparably to standard GLASSO.

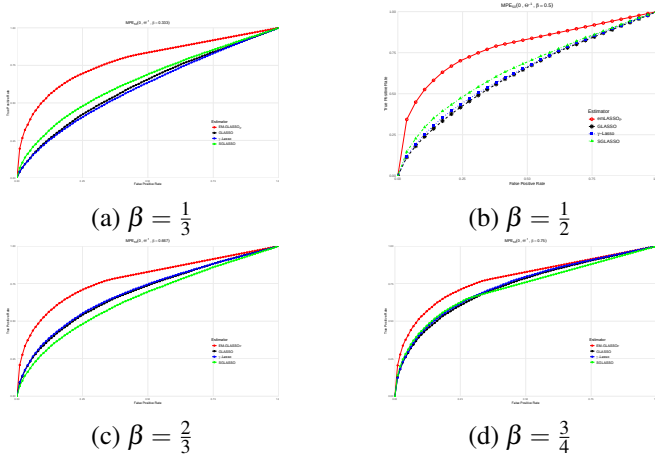


Figure 3: ROC curves of the GLASSO, EM-GLASSO_P, γ -LASSO, and SGLASSO under $\mathcal{MPE}_{50}(0, \Theta^{-1}, \beta)$ distribution.

Case 2: $\mathcal{MPE}, d = 100$; **sparsity**= 10%; $n = 50$: In the case of a high-dimensional context as $d > n$, Figure 4 shows the ROC curves with a high dimension $d = 100$ and a limited sample size ($n = 50$). The EM-GLASSO_p estimator demonstrates a commanding performance advantage over competing methods across all considered shape parameters ($\beta \in \{1/3, 1/2, 2/3, 3/4\}$). As illustrated in the ROC curves, EM-GLASSO_p consistently achieves the highest TPR for low FPR, exhibiting significant robustness to the heavy-tailed nature of the MPE distribution. The SGLASSO estimator consistently ranks second, clearly separating itself from the remaining estimators. Conversely, the GLASSO and γ -LASSO estimators show substantially degraded performance in this "large d , small n " regime, clustering together with much lower sensitivity compared to the top-performing EM-GLASSO_p.

Objective: Extend the GLASSO framework under the MPE distribution while explicitly preserving interpretable conditional relationships.

Key Findings: EM-GLASSO_p accurately models heavy-tailed MPE distributions to extract interpretable, sparse network structures.

Related publication: 3.

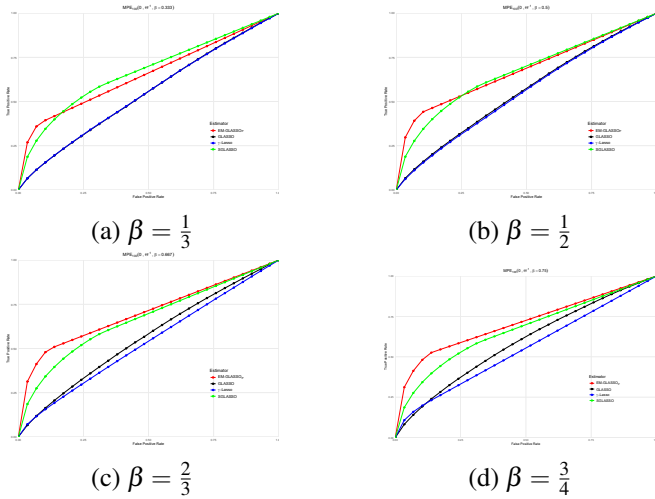


Figure 4: ROC curves of the GLASSO, EM-GLASSO_p, γ -LASSO, and SGLASSO under $\mathcal{MPE}_{100}(0, \Theta^{-1}, \beta)$ with $n = 50$.

4.2. Thesis Group II: Applications of GLASSO

4.2.1. Thesis 3: Cognitive-Behavioral Networks

This case study does not propose a new psychological model; rather, it serves as an empirical benchmark for the EM-GLASSO algorithm under standard GLASSO settings. We assess the interpretability, sparsity, and stability of the EM-GLASSO framework.

Post-Traumatic Stress Disorder (PTSD) The dataset was provided by Armour et al. [3]¹ and contains $n = 221$ observations across $d = 20$ PTSD symptoms. The symptoms are

¹<https://osf.io/mj5wa/files/osfstorage>

grouped into four domains: Intrusion (B1–B5), Avoidance (C1–C2), Negative alterations in cognition/mood (D1–D7), and Arousal/reactivity (E1–E6).

Both GLASSO and EM-GLASSO_p results align with the established clinical literature. Trauma-related amnesia (D1) and avoidance symptoms (C1, C2) remain isolated in the network as they lack conditional dependence on other symptoms, confirming prior findings [5]. Furthermore, the models identify active conditional dependencies among negative cognitive and mood alterations (D5–D7). These alterations link directly to intrusion symptoms [11]. Finally, intrusion symptoms exhibit strong conditional dependencies with negative beliefs, self-blame, and trauma-related emotions (D2–D4). These specific connections persist even when controlling for all other variables [18]. The results presented in Figures 5a and 5b.

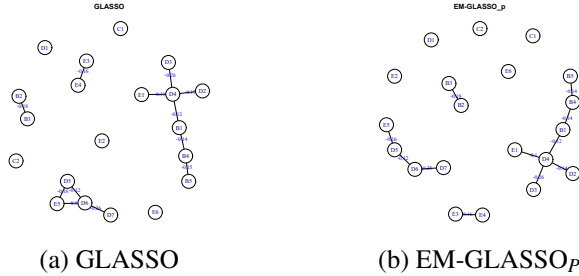


Figure 5: Sparse conditional dependence network of PTSD symptoms

Emotional monitoring: The dataset contains biosensor-based physiological and behavioral data for student responses in ide-

ological and political education settings to monitor engagement and emotional states². The dataset includes physiological metrics where higher values generally indicate stress or engagement: **heart rate** (bpm), **body temperature** (°C), **cortisol**, **skin conductance** (arousal), and **pupil diameter** (cognitive load), alongside **EEG** frequencies. Behavioral features include **smile/frown intensity** (0.0–1.0) and physical **activity level**. Environmental variables (**ambient noise**, **lighting**) track potential distractions. These features are used to analyze the student’s **emotional**, **cognitive**, and **engagement states**.

Figures 6a and 6b present the estimated networks. Both GLASSO and EM-GLASSO_P yield sparse structures that align tightly with established psychophysiological literature. We draw two primary structural conclusions. First, cortisol secretion and smile intensity function as parallel manifestations of the underlying emotional state, as they do not interact directly. This state explains their strict conditional independence when controlling for emotional state [6, 27, 7]. Second, ambient noise, lighting, engagement, and cognitive state remain isolated, exerting only indirect effects. They lack unique conditional dependencies when the model accounts for the entire network [14]. Ultimately, these sparse psychological graphs are the expected outcome of ℓ_1 -regularized estimation [9]. In addition, the results reveal the conditional dependency structure of student engagement, showing that it is not an isolated variable. It exhibits active condi-

²<https://www.kaggle.com/datasets/ziya07/emotional-monitoring-dataset/data>

tional dependencies with exactly three features: skin conductance, EEG, and lighting. It remains strictly conditionally independent of all other network variables.

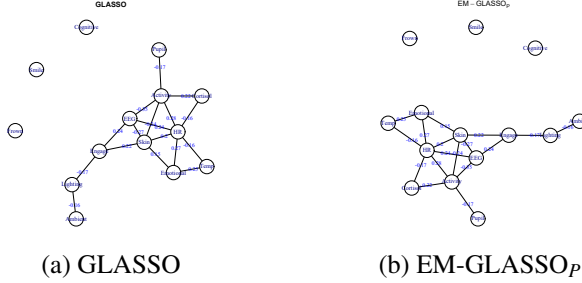


Figure 6: Estimated sparse conditional dependence network of student emotion monitoring variables

Objective: Evaluate the interpretability, sparsity, and stability of the EM-GLASSO framework relative to the GLASSO benchmark using cognitive-behavioral datasets.

Key Findings: EM-GLASSO_p performs comparably to GLASSO when the Gaussian assumption is correctly specified.

Related publication: 6.

4.2.2. Thesis 4: Process Quality Inspection

Adegoke et al. [1] used GLASSO to design directional MCV control charts that outperform classical methods under small subgroup sizes. Extending this approach, we apply GLASSO and EM-GLASSO to enhance process quality monitoring. Ultimately, these estimators drastically improve the detection of

small and moderate process shifts.

GLASSO-based MCV control charts: For a Gaussian sample $\mathbf{X}_{1:n} \sim \mathcal{N}_d(\mu, \Sigma)$, the theoretical Multivariate Coefficient of Variation (MCV) is $\varphi = (\mu^\top \Sigma^{-1} \mu)^{-1/2}$ [20]. In practice, we substitute the sample mean $\bar{\mathbf{X}}$ and covariance $\mathbf{S}$ to obtain the empirical estimator:

$$\hat{\varphi} = \left(\bar{\mathbf{X}}^\top \mathbf{S}^{-1} \bar{\mathbf{X}} \right)^{-1/2}$$

Assuming $n > d$, $\hat{\varphi}$ follows a non-central F -distribution with non-centrality parameter ω [26]. Its inverse Cumulative Distribution Function is given by:

$$F_{\hat{\varphi}}^{-1}(\alpha|n, d, \omega) = \sqrt{\frac{n(n-d)}{(n-1)d} \left[\frac{1}{F_F^{-1}(1-\alpha|d, n-d, \omega)} \right]}$$

Because this distribution is skewed, standard two-sided monitoring yields a biased average run length (ARL) [19]. We resolve this by employing separate one-sided upward and downward control charts.

We define the lower (LCL) and upper (UCL) control limits for the directional MCV charts as:

$$LCL = F_{\hat{\varphi}}^{-1}(\alpha|n, d, \omega), \quad UCL = F_{\hat{\varphi}}^{-1}(1-\alpha|n, d, \omega),$$

where $\omega = n/\varphi_0$ and φ_0 represents the baseline Phase I MCV.

Using the GLASSO framework, we get $\hat{\Theta}^{GL}$, then using the

estimated precision matrix, we get

$$\boldsymbol{\varphi}^{GL} = \left(\overline{\mathbf{X}}^\top \hat{\boldsymbol{\Theta}}^{GL} \overline{\mathbf{X}} \right)^{-1/2} \quad (2)$$

Therefore, the GLASSO-based MCV control chart limits are given as:

$$LCL = F_{\hat{\varphi}^{GL}}(\alpha|d, n), \quad UCL = F_{\hat{\varphi}^{GL}}(1 - \alpha|d, n)$$

Similar process we apply for the EM-GLASSO framework. The control chart triggers an out-of-control signal if either $\hat{\varphi}^{GL} < LCL$ or $\hat{\varphi}^{GL} > UCL$.

Results and Discussion This example evaluates production stability using the Wine Quality dataset with 11 physicochemical variables [8]. In **Phase I**, the process is assumed to be stable and in control, as only samples with quality scores greater than 5 has been considered. The control chart parameters (UCL, CL, LCL, ω , and ρ) were estimated using 57 batches, each of size 15. The tuning parameter ρ was selected using the eBIC criterion, resulting in $\rho = 0.2846$. The remaining parameters were estimated using the precision matrix estimator and are given as

- GLASSO: LCL = 0.00686, CL = 0.06625, UCL = 0.07141.
- EM-GLASSO: LCL = 0.00665, CL = 0.06421, UCL = 0.06921.

Phase II corresponds to the inspection phase, in which the Phase I control charts are applied to monitor the process under real operating conditions that may include out-of-control sam-

ples. In this phase, 40 batches, each of size 15, were monitored across 11 variables.

The EM-GLASSO framework demonstrates greater sensitivity than the standard GLASSO approach, as illustrated in Figures 7 and 8. The GLASSO chart detected 4 batches out-of-control batches, whereas the EM-GLASSO chart identified 10 batches. This higher detection rate indicates that EM-GLASSO is more effective at detecting subtle structural shifts in the process. In industrial quality monitoring, these control charts assess the multivariate coefficient of variation by tracking the relative dispersion of complex production data, thereby providing early warnings of process instability before critical failures occur. These findings are consistent with the simulation results reported by Adegoke et al. [1] and with the simulation results presented for EM-GLASSO_P.

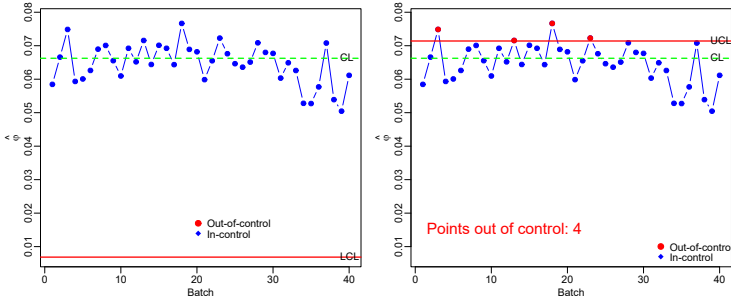


Figure 7: GLASSO-MCV chart.

This study employs one-sided control charts to monitor directional shifts in the process MCV. We estimate the underlying

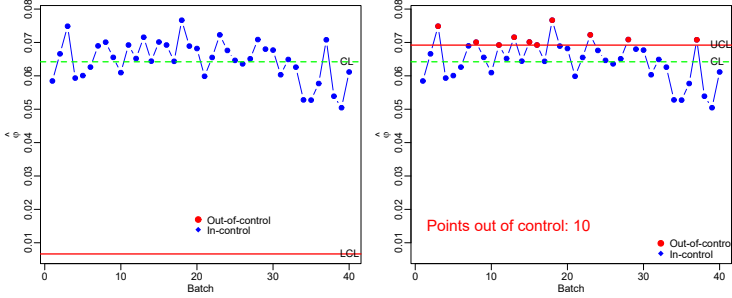


Figure 8: EM-GLASSO_p-MCV chart.

precision matrices using both the GLASSO and EM-GLASSO frameworks. The empirical results demonstrate that EM-GLASSO charts consistently outperform their standard GLASSO.

Objective: Develop and evaluate one-sided upward and downward control charts for detecting transient process changes under the GLASSO and EM-GLASSO_p frameworks.

Key Findings: The empirical results show that the EM-GLASSO_p charts consistently outperform the standard GLASSO charts when the normality assumption is violated.

Related publications: 1, 2, 5.

4.2.3. Thesis 5: Nutrients Interaction Network

GGMs provide a systems-level map of nutrient covariation in baobab leaves, identifying critical biochemical interactions to advance data-driven food-system modeling. Yet, because GGMs strictly assume normality, the skewed distributions common in

nutritional data can severely distort results [2]. To handle this non-normality, we apply Semiparametric Gaussian Copula Graphical Models (SGCGMs). It extract true conditional dependencies by applying a monotonic transformation $f(\mathbf{X}) = (f_1(X_1), \dots, f_d(X_d))^T$ to the original vector $\mathbf{X} = (X_1, \dots, X_d)^T$, such that the transformed space $f(\mathbf{X})$ achieves multivariate normality.

Results and Discussion: Figure 9 displays the Pearson correlation heatmap for the analyzed nutrients. Positive associations among ash, Ca, and K indicate linked symplastic transport [4]. Negative correlations between phenolics and minerals reflect active metal–polyphenol chelation [12]. Additionally, the high ash–pH correlation ($r \approx 0.9$) demonstrates alkaline buffering [23], whereas fat and minerals correlate negatively due to low tissue affinity [22]. Ultimately, this Pearson heatmap reveals only marginal associations; it cannot disentangle direct relationships from indirect confounding.

For the dataset of 60 observations and 13 nutrients, Mardia’s tests confirmed severe non-normality. Although tail behavior was normal (kurtosis = 197.11, $p = 0.679$), the data exhibited profound asymmetry (skewness = 71.32, $p < 0.001$). Because this violates Gaussian assumptions, we deploy SGCGMs to estimate the underlying graphical model.

Figure 10a reveals that SGCGMs provides a clear and flexible framework for uncovering nutrient interaction networks in baobab leaves. It showed four clusters: Mineral cluster, Composition cluster, pH-centered interaction layer, and Polyphenol

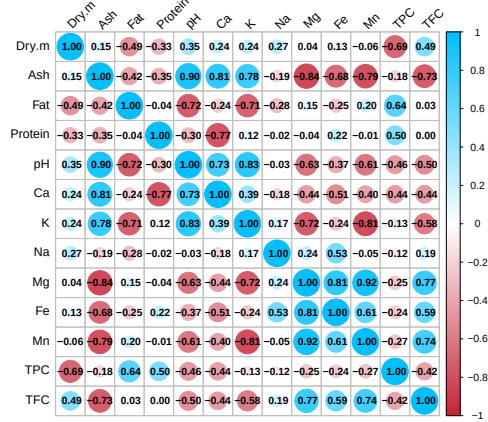


Figure 9: Heatmap of pairwise Pearson correlations among the analyzed elements.

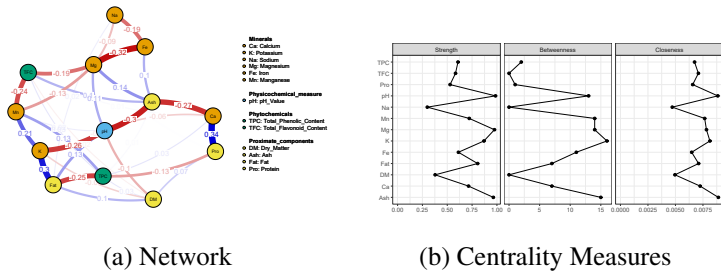


Figure 10: The network analysis of Baobab leaves using SGCGMs

cluster. The role of Ash and Ca is transitional. Ash shows strong associations with pH (-0.297) and Ca (-0.274), as well as with Mg and Fe. Ca links strongly to Pro (0.335) but weakly elsewhere. These variables act as bridges. They connect composition to minerals, but neither dominates. Across clusters, the strongest effects run through a few edges. The Fat-K link (0.305), Pro-Ca link (0.335), Mg-Fe link (-0.317), and Mn-TFC link (-0.236) stand out. These edges shape the network by defining how information or influence can flow.

The centrality measures shown in Figure 10b confirm a sparse yet well-organized structure, consistent with the network. The measures do not contradict the network; instead, they sharpen its interpretation by showing which nodes matter most once weak edges have been removed.

Figures 11a and 11b illustrate the EM-GLASSO_P network and centrality measures. EM-GLASSO_H recovers a sparse biochemical network driven by K as the central hub. The model isolates distinct dependencies: TPC and TFC show strong positive partial correlation, while Fe and Mg correlate negatively. Peripheral variables (pH, DM, fat) remain completely isolated with zero betweenness. By aggressively shrinking weak correlations, the algorithm successfully disentangles core pathways from conditionally independent variables.

Unknown data distributions threaten parametric models. Specifically, the rigid assumptions of EM-GLASSO_H can induce spurious partial correlations, inflating dependencies like DM-TFC.

SGCGM avoids this by using a semiparametric Gaussian copula to model dependence without specifying marginals. This approach natively accommodates skewness and unknown generating processes.

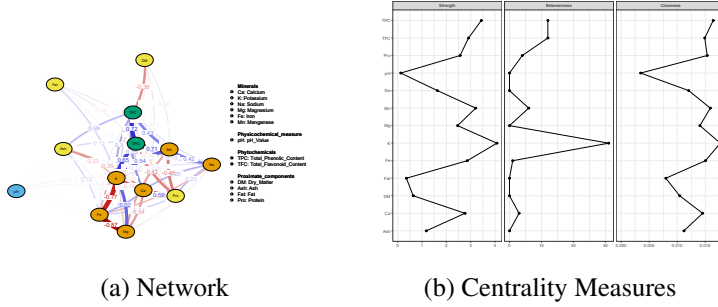


Figure 11: The network analysis of Baobab leaves using EM-GLASSO_H

Objective: Infer and interpret the conditional dependence structure among nutrient components in baobab leaf samples using robust GLASSO and SGCGMs under non-Gaussian distributional assumptions.

Key Findings: When parametric assumptions fail, SGCGMs outperform EM-GLASSO_H. They strictly separate signal from noise. The resulting networks are highly interpretable and capture true conditional dependencies.

Related publication: 4.

References

- [1] Nurudeen A. Adegoke, Abdaljbbar Dawod, Olatunde Adebayo Adeoti, Ridwan A. Sanusi, and Saddam Akber Abbasi. Monitoring multivariate coefficient of variation for high-dimensional processes. *Quality and Reliability Engineering International*, 38(5):2606–2621, 2022. doi:10.1002/qre.3094.
- [2] Michael Altenbuchinger, Antoine Weihs, John Quackenbush, Hans Jörgen Grabe,

- and Helena U. Zacharias. Gaussian and mixed graphical models as (multi-)omics data analysis tools. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1863(6):194418, 2020. Transcriptional Profiles and Regulatory Gene Networks. doi:<https://doi.org/10.1016/j.bbagr.2019.194418>.
- [3] Cherie Armour, Eiko I. Fried, Marie K. Deserno, Jack Tsai, and Robert H. Pietrzak. A network analysis of DSM-5 posttraumatic stress disorder symptoms and correlates in US military veterans. *Journal of Anxiety Disorders*, 45:49–59, 2017. doi:[10.31234/osf.io/p69m7](https://doi.org/10.31234/osf.io/p69m7).
 - [4] Vijendra Singh Bhati, Pravin Kumar Singh, Yogeshchandra J Parmar, and Pugazhenthir Godson Rokins. Foliar uptake and translocation of nanoparticles in the plants: a review. *Journal of Plant Nutrition*, pages 1–38, 2025.
 - [5] Marianne Skogbrott Birkeland, Talya Greene, and Tobias Raphael Spiller. The network approach to posttraumatic stress disorder: a systematic review. *European Journal of Psychotraumatology*, 11(1):1700614, 2020. PMID: 32002135. doi:[10.1080/20008198.2019.1700614](https://doi.org/10.1080/20008198.2019.1700614).
 - [6] Jim Blascovich and Wendy Berry Mendes. Social psychophysiology and embodiment. In Susan Fiske, Daniel Gilbert, and Gardner Lindzey, editors, *Handbook of Social Psychology*. Wiley, 5 edition, 2010.
 - [7] Juliette Bruin, Ivo V. Stuldreher, Paola Perone, Koen Hogenelst, Marnix Naber, Wim Kamphuis, and Anne-Marie Brouwer. Detection of arousal and valence from facial expressions and physiological responses evoked by different types of stressors. *Frontiers in Neuroergonomics*, Volume 5 - 2024, 2024. doi:[10.3389/fnrgo.2024.1338243](https://doi.org/10.3389/fnrgo.2024.1338243).
 - [8] Paulo Cortez, A. Cerdeira, F. Almeida, T. Matos, and J. Reis. Wine Quality. UCI Machine Learning Repository, 2009. DOI: <https://doi.org/10.24432/C56S3T>.
 - [9] Sacha Epskamp, Lourens J Waldorp, René Møttus, and Denny Borsboom. The gaussian graphical model in cross-sectional and time-series data. *Multivariate behavioral research*, 53(4):453–480, 2018.
 - [10] Michael Finegold and Mathias Drton. Robust graphical modeling of gene networks using classical and alternative t-distributions. *The Annals of Applied Statistics*, 5(2A):1057–1080, 2011.

- [11] Ryan Holliday, Nicholas Holder, Cynthia Yamokoski, Jennifer Burden, Lauren M. Sippel, Ilan Harpaz-Rotem, and Noelle B. Smith. A preliminary evaluation of the effectiveness of massed delivery of ptsd psychotherapy in va residential treatment. *General Hospital Psychiatry*, 97:126–129, 2025. doi:
[10.1016/j.genhosppsych.2025.10.004](https://doi.org/10.1016/j.genhosppsych.2025.10.004).
- [12] Zdeněk Kejík, Robert Kaplánek, Michal Masařík, Petr Babula, Adam Matkowski, Petr Filipenský, Kateřina Veselá, Jakub Gburek, David Šýkora, Pavel Martásek, et al. Iron complexes of flavonoids-antioxidant capacity and beyond. *International journal of molecular sciences*, 22(2):646, 2021.
- [13] D. Kelker. Distribution theory of spherical distributions and a location-scale parameter generalization. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 419–430, 1970.
- [14] Xuejiao Li, Ruili Wang, and Mincheol Whang. Designing light for emotion: A neurophysiological approach to modeling affective responses to the interplay of color and illuminance. *Biomimetics*, 10(10), 2025. doi:[10.3390/biomimetics10100696](https://doi.org/10.3390/biomimetics10100696).
- [15] Han Liu, Fang Han, Ming Yuan, John Lafferty, and Larry Wasserman. High-dimensional semiparametric gaussian copula graphical models. *Annals of Statistics*, 40(4):2293–2326, 2012. doi:[10.1214/12-AOS1037](https://doi.org/10.1214/12-AOS1037).
- [16] Zhengke Lu and Long Feng. Robust sparse precision matrix estimation and its application, 2025. URL: <https://arxiv.org/abs/2503.03575>, arXiv: 2503.03575.
- [17] Baba H. M. Propriétés asymptotiques d’un estimateur bi-spectral pour un processus p-adique strictement homogène. *C. R. Acad. Sci. Paris*, 326(1):987–990, 1998.
- [18] Richard J McNally, Donald J Robinaugh, Gwyneth WY Wu, Li Wang, Marie K Deserno, and Denny Borsboom. Mental disorders as causal systems: A network approach to posttraumatic stress disorder. *Clinical psychological science*, 3(6):836–849, 2015.
- [19] Quoc Thong Nguyen, Kim Phuc Tran, Philippe Castagliola, Giovanni Celano, and Salim Lardjane. One-sided synthetic control charts for monitoring the multivariate coefficient of variation. *Journal of Statistical Computation and Simulation*, 89(10):1841–1862, 2019. doi:[10.1080/00949655.2019.1600694](https://doi.org/10.1080/00949655.2019.1600694).

- [20] Mikhail Nikulin and Vassily Voinov. Unbiased estimators and their applications. In *International Encyclopedia of Statistical Science*, pages 2821–2824. Springer, 2025.
- [21] Takahiro Onizuka and Shintaro Hashimoto. Robust bayesian graphical modeling using γ -divergence. *Journal of Multivariate Analysis*, page 105461, 2025.
- [22] Renu Pandey. Mineral nutrition of plants. In *Plant Biology and Biotechnology: Volume I: Plant Diversity, Organization, Function and Improvement*, pages 499–538. Springer, 2015.
- [23] Ernest Uwiringiyimana, Hong-wei Lai, Ni Ni, Ren-yong Shi, Xiao-ying Pan, Jianing Gao, Md Romel Biswash, Jiu-yu Li, Xiu-min Cui, and Ren-kou Xu. Comparative efficacy of alkaline slag, biomass ash, and biochar application for the amelioration of different acidic soils. *Plant and Soil*, 504(1):47–61, 2024.
- [24] Joe Whittaker. *Graphical models in applied multivariate statistics*. Wiley Publishing, 2009.
- [25] Lingzhou Xue and Hui Zou. Regularized rank-based estimation of high-dimensional nonparanormal graphical models. *The Annals of Statistics*, 40(5):2541–2571, oct 2012. doi:[10.1214/12-AOS1041](https://doi.org/10.1214/12-AOS1041).
- [26] Wai Chung Yeong, Michael Boon Chong Khoo, Wei Lin Teoh, and Philippe Castagliola. A control chart for the multivariate coefficient of variation. *Quality and Reliability Engineering International*, 32(3):1213–1225, 2016. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/qre.1828>, doi: [10.1002/qre.1828](https://doi.org/10.1002/qre.1828).
- [27] Goran Šimić, Mladenka Tkalčić, Vana Vukić, Damir Mulc, Ena Španić, Marina Šagud, Francisco E. Olucha-Bordonau, Mario Vukšić, and Patrick R. Hof. Understanding emotions: Origins and roles of the amygdala. *Biomolecules*, 11(6), 2021. doi:[10.3390/biom11060823](https://doi.org/10.3390/biom11060823).



Registry number: DEENK/425/2026.PL
Subject: PhD Publication List

Candidate: Abdaljbbar Babiker Abdaljbbar Dawod
Doctoral School: Doctoral School of Informatics

List of publications related to the dissertation

Foreign language international book chapters (1)

1. **Dawod, A. B. A.**, Terdik, G.: Railway Track Quality Inspection via GLasso with Zerocrossing Segmentation.
In: Springer, New York, 244-250, 2026, (Lecture Notes in Networks and Systems, ISSN 2367-3389 ; 1768.) ISBN: 9783032138972

Foreign language scientific articles in Hungarian journals (1)

2. **Dawod, A. B. A.**, Terdik, G.: Use of Zero-crossings Segmentation for Track Quality Assessment.
Acta Tech. Jaurinensis. 17 (1), 8-21, 2024. EISSN: 2064-5228.
DOI: <http://dx.doi.org/10.14513/actatechjaur.00726>

Foreign language scientific articles in international journals (3)

3. Terdik, G., **Dawod, A. B. A.**: Extending GLASSO to Non-Gaussian Settings: Sparse Concentration Estimation via EM Algorithm.
Comput. Stat. "Accepted by Publisher", 1-33, 2026. ISSN: 0943-4062.
IF: 1.4 (2024)
4. **Dawod, A. B. A.**, Kovács, B., Eltayeb, O. A. M., Ahmed, A. E. M.: Semiparametric Gaussian Copula Graphical Modeling for Identifying Nutrient Interaction Networks in Baobab Leaves: Implications for Healthy Nutrient Source Analysis.
PLOS One. [Under review], 1-22, 2026. ISSN: 1932-6203.
IF: 2.6 (2024)
5. Adegoke, N. A., **Dawod, A. B. A.**, Adeoti, O. A., Sanusi, R. A., Abbasi, S. A.: Monitoring multivariate coefficient of variation for high-dimensional processes.
Qual. Reliab. Eng. Int. 38 (5), 2606-2621, 2022. ISSN: 0748-8017.
DOI: <http://dx.doi.org/10.1002/qre.3094>
IF: 2.3





Foreign language conference proceedings (1)

6. **Dawod, A. B. A.**, Terdik, G.: Mapping Cognitive-Behavioral Networks: A Graphical Lasso Approach to Dependence Structure Visualization.
In: 2025 IEEE 16th International Conference on Cognitive Infocommunications (CogInfoCom) /CogInfo (szerk.) ISBN: 9798350356922

List of other publications

Foreign language scientific articles in international journals (9)

7. Ahmed, A. E. M., Mozzon, M., **Dawod, A. B. A.**, Awad Mustafa Eltayeb, O., Shaikh, A. M., Kovács, B.: Potential of Baobab (*Adansonia digitata* L.) Dried Fruit Pulp in Contributing to Food and Nutrient Security in Sudan and Senegal.
J. Agric. Food Res. 19, 1-9, 2025. ISSN: 2666-1543.
DOI: <http://dx.doi.org/10.1016/j.jafr.2025.101681>
IF: 6.2 (2024)
8. **Dawod, A. B. A.**, Nsaif, M., Alshuwaili, D. G. H., Abdi, I. H., Neamah, H. A.: Value-at-Risk Estimation in the Tadawul Exchange: A Comparison of Recurrent Neural Networks and GARCH Models.
Systems and Soft Computing. 7, 1-14, 2025. EISSN: 2772-9419.
DOI: <https://doi.org/10.1016/j.sasc.2025.200411>
IF: 3.6 (2024)
9. Ahmed, A. E. M., Mozzon, M., **Dawod, A. B. A.**, Eltayeb, O. A. M., Shaikh, A. M., ElObeid, T., Mohammed, E. A. H., Kovács, B.: Evaluation of the Nutritional Impact of Baobab Leaves (*Adansonia digitata* L.) as a Dietary Intervention to Combat Nutrient Deficiencies and Poverty-Related Health Problems.
Nutrients. 16, 1-21, 2024. EISSN: 2072-6643.
DOI: <https://doi.org/10.3390/nu16244340>
IF: 5
10. Al, M. M., **Dawod, A. B. A.**: Model Selection and Post Selection to Improve the Estimation of the ARCH Model.
J. Risk Financ. Manag. 15 (4), 1-17, 2022. ISSN: 1911-8066.
DOI: <http://dx.doi.org/10.3390/jrfm15040174>
11. **Dawod, A. B. A.**, Adegoke, N. A., Abbasi, S. A.: Efficient linear profile schemes for monitoring bivariate correlated processes with applications in the pharmaceutical industry.
Chemometrics Intell. Lab. Syst. 206, 1-16, 2020. ISSN: 0169-7439.
DOI: <http://dx.doi.org/10.1016/j.chemolab.2020.104137>
IF: 3.491





12. Adegoke, N. A., Abbasi, S. A., **Dawod, A. B. A.**, Pawley, M. D.: Enhancing the performance of the EWMA control chart for monitoring the process mean using auxiliary information.
Qual. Reliab. Eng. Int. 35 (4), 920-933, 2018. ISSN: 0748-8017.
DOI: <http://dx.doi.org/10.1002/qre.2436>
IF: 1.409
13. **Dawod, A. B. A.**, Al, M. M., Abbasi, S. A.: On efficient estimation strategies in monitoring of linear profiles.
Int. J. Adv. Manuf. Technol. 96 (9-12), 3977-3991, 2018. ISSN: 0268-3768.
DOI: <http://dx.doi.org/10.1007/s00170-018-1835-y>
IF: 2.496
14. **Dawod, A. B. A.**, Abbasi, S. A., Al, M. M.: On the performance of coefficient of variation control charts in Phase I.
Qual. Reliab. Eng. Int. 34 (6), 1029-1040, 2018. ISSN: 0748-8017.
DOI: <http://dx.doi.org/10.1002/qre.2305>
IF: 1.409
15. **Dawod, A. B. A.**, Riaz, M., Abbasi, S. A.: On Model Selection for Autocorrelated Processes in Statistical Process Control.
Qual. Reliab. Eng. Int. 33 (4), 867-882, 2016. ISSN: 0748-8017.
DOI: <http://dx.doi.org/10.1002/qre.2063>
IF: 1.366

Foreign language abstracts (1)

16. **Dawod, A. B. A.**: Time series processes monitoring using statistical process control and machine learning approaches with applications on social network surveillance.
In: ICAI 2023 - 12th International Conference on Applied Informatics, Eszterházy Károly Egyetem, Eger, 1-2, 2023.

Total IF of journals (all publications): 31,271

Total IF of journals (publications related to the dissertation): 6,3

The Candidate's publication data submitted to the Tudóstér have been validated by DEENK on the basis of the Journal Citation Report (Impact Factor) database.

08 June, 2026

