

Article

An Expectation-Maximization Algorithm for Combining a Sample of Partially Overlapping Covariance Matrices

Deniz Akdemir ^{1,*}, Mohamed Somo ² and Julio Isidro-Sánchez ³¹ Center of International Bone Marrow Transplantation Research, Minneapolis, MN 55401, USA² Syngenta Seeds, Junction City, KS 66441, USA³ Centro de Biotecnología y Genómica de Plantas, Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, Universidad Politécnica de Madrid, 28223 Madrid, Spain

* Correspondence: dakdemir@nmdp.org

Abstract: The generation of unprecedented amounts of data brings new challenges in data management, but also an opportunity to accelerate the identification of processes of multiple science disciplines. One of these challenges is the harmonization of high-dimensional unbalanced and heterogeneous data. In this manuscript, we propose a statistical approach to combine incomplete and partially-overlapping pieces of covariance matrices that come from independent experiments. We assume that the data are a random sample of partial covariance matrices sampled from Wishart distributions and we derive an expectation-maximization algorithm for parameter estimation. We demonstrate the properties of our method by (i) using simulation studies and (ii) using empirical datasets. In general, being able to make inferences about the covariance of variables not observed in the same experiment is a valuable tool for data analysis since covariance estimation is an important step in many statistical applications, such as multivariate analysis, principal component analysis, factor analysis, and structural equation modeling.

Keywords: imputation; covariance estimation; expectation-maximization; multi-view data; heterogeneous databases



Citation: Akdemir, D.; Somo, M.; Isidro-Sánchez, J. An Expectation-Maximization Algorithm for Combining a Sample of Partially Overlapping Covariance Matrices. *Axioms* **2023**, *12*, 161. <https://doi.org/10.3390/axioms12020161>

Academic Editor: Anuradha Roy

Received: 28 December 2022

Revised: 24 January 2023

Accepted: 26 January 2023

Published: 4 February 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

High-throughput measurement technologies have revolutionized many scientific disciplines by decreasing the time and cost of analyzing multiple samples and generating huge amounts of data, which has created great opportunities but also brought new challenges, especially in data cleaning, normalization, dimension reduction, data harmonization, and data storage. The challenge we want to address in this manuscript is the integration of data from different sources by making inferences about the covariance of variables not observed in the same experiment. Covariance estimation is an important step in many statistical applications, such as multivariate analysis, principal component analysis, factor analysis, structural equation modeling, and statistical genomics.

Harmonization of the increasing amounts of datasets that accumulate in databases has great potential to accelerate our understanding of scientific processes. One of the challenges that still needs to be addressed is the incompleteness inherent in scientific data, in other words, each dataset that is the result of an experimental or observational study can address a limited number of variables with a limited number of samples. The lack of a large sample size reduces the reproducibility of the study [1], and the lack of a large number of measured variables narrows the scope of the study. The integration of information from different datasets is required to accelerate the identification of significant scientific outcomes.

Modern data integration approaches include feature imputation, conventional meta-analysis [2–5] and many other new methods based on machine learning and statistical meth-

ods [6–11]. Detailed reviews and classifications of some promising statistical approaches for data integration can be found in [12–14]. The results in these studies demonstrate the advantage of integrating multiple diverse datasets.

This work describes a method for learning the covariances among a set of variables by combining multiple covariance matrices from independent experiments. Our method is closely related to meta-analytic structural equation modeling (MASEM). MASEM is a two-stage approach to estimating the parameters of a structural equation model (SEM) using covariance matrices from different studies [15,16]. The first stage of MASEM is to pool the covariance matrices from the different studies, and the second stage is to estimate the parameters of the SEM using the pooled covariance matrix. Our method can be used in the first stage of MASEM to pool the covariance matrices from the different studies.

The main advantage of our method over the existing techniques for pooling covariances is that we use the expectation-maximization algorithm to maximize the likelihood function, leading to an analytical solution for each iteration of the algorithm. This makes our algorithm computationally efficient and suitable for combining covariance matrices involving many variables. In contrast, the existing MASEM methods for pooling covariances are based on iterative algorithms, which directly maximize the likelihood function over the parameters of the covariance matrix. Since the number of parameters in the covariance matrix is quadratic in the number of variables, these methods become more computationally demanding for large numbers of variables. We compared the performance of our method to the performance of the first stage of two-stage MASEM approaches in the Illustrations section with a simulated data set. The results of this comparison demonstrate that our method is computationally more efficient and more accurate than the existing MASEM approaches for pooling covariances. Our algorithm can be used to combine covariance matrices involving thousands of variables, unlike current implementations of the two-stage MASEM approaches, which are unable to handle such large data sets [17]. Another benefit of our method is that, when used with partially overlapping covariance matrices (i.e., when there are pairs of variables that were not observed within any one of the sample covariance matrices), the two-stage MASEM models resulted in a pooled covariance matrix where all these missing values were imputed with the same estimated mean population correlation value. This is not the case for our method, since our method does not make any assumptions about the covariance structure, as these methods do.

The covariance-based approach of combining data can be contrasted with feature imputation-based approaches, which are the preferred method for dealing with incomplete data sets. Popular approaches for imputation include random forest [18], expectation-maximization [19,20], and low-rank matrix factorization [21], among others. The main advantage of the covariance-based method is that it can infer the relationship of variables that are not observed in the same experiment without using feature imputation. This is useful in many applications. For example, the covariance-based approach can be used to infer the relationship between individuals by combining pedigree relationships with genomic relationship matrices calculated from other omics data [17]. A pedigree-based relationship matrix is calculated based on the known ancestral relationships between individuals; however, there is no straightforward way to impute this ancestral information for individuals not in the pedigree. Additionally, in many studies involving meta-analytical covariance estimation based on multiple datasets, the only data available are the covariance matrices, so the imputation of features is not an option.

Our method and results are unique to our knowledge, although it has been inspired by similar methods such as conditional iterative proportional fitting for the Gaussian distribution [22,23] and a method for updating a pedigree relationship matrix and a genotypic matrix relationship matrix that includes a subset of genotypes from the pedigree-based matrix [24] (the H-matrix).

In the following, we will first formally define and describe the statistical problem. We will provide a detailed description of the derivation of the algorithm for the case where all samples are assumed to have the same known degrees of freedom and the asymptotic standard errors for this case. Then, we generalize the algorithm to the case where the degrees of freedom are different for the different samples. We include a discussion on sparse covariance estimation and parametrized covariance models. Consequently, we present three simulation studies to study the properties of the algorithm such as convergence, dependence on initial estimates, and compare the performance against two MASEM methods. We also demonstrate the use of our method with an empirical case study where we combine data from 50 agricultural phenotypic trials to obtain the combined covariance matrix for 37 traits. This is followed by our conclusions and final comments.

2. Methods

In this section, we describe the use of the EM algorithm for the maximum likelihood estimation of parameters in a Wishart distribution for combining a sample of partially overlapping covariance matrices. The data are assumed to be a random sample of partial covariance matrices, and the EM algorithm is used to estimate the parameters of this distribution. EM is a popular iterative algorithm that alternates between two steps: the expectation step, which computes the expectation of the log-likelihood function with respect to the current estimates of the parameters, and the maximization step, which maximizes the expectation of the log-likelihood function with respect to the current estimates of the parameters. The EM algorithm is guaranteed to converge to a local maximum of the log-likelihood function, and this circumvents many of the difficulties related to the direct maximization of the observed data likelihood function. This is especially advantageous when the number of variables in the covariance matrix is large, as the number of parameters to estimate increases quadratically with the number of variables. In such cases, numerical methods such as the Gradient-Descent and Newton–Raphson methods are not effective.

2.1. Preliminary Results about Normal and Wishart Distributions

We use the following standard matrix notation: $tr(A)$ denotes the trace of matrix A and $|A|$ denotes the determinant of matrix A . For two matrices, A and B , $A \odot B$ denotes the Hadamard product (element-wise matrix product) of A and B and $A \otimes B$ denotes the Kronecker product of A and B .

We will write $\mathbf{x} \sim N_p(\boldsymbol{\mu}, \Psi)$ to say that a random vector \mathbf{x} has a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix Ψ . We will write $\mathbf{Y} \sim W_n(\nu, \Psi)$ to denote a Wishart distribution with degrees of freedom ν and scale matrix Ψ . The Wishart and the multivariate normal random variables are related by the following formula:

$$\mathbf{Y} = \mathbf{X}'\mathbf{X} \sim W_n(\nu, \Psi) \quad (1)$$

where \mathbf{X} is a $p \times n$ matrix of independent and identically distributed random variables with mean $\boldsymbol{\mu} = \mathbf{0}$ and covariance matrix Ψ . The Wishart distribution is the natural distribution for covariance matrices since the sample covariance matrix obtained from a multivariate normal sample has a Wishart distribution.

We use $\mathbf{X} \sim N_{p,n}(M, \Sigma, \Psi)$ to say that a random matrix \mathbf{X} has a matrix variate normal distribution with mean M , row covariance matrix Σ , and column covariance matrix Ψ . If $\mathbf{X} \sim N_{p,n}(M, \Sigma, \Psi)$, then $vec(\mathbf{X}) \sim N_{pn}(vec(M), \Psi \otimes \Sigma)$ where $vec(\mathbf{X})$ is the vectorization of \mathbf{X} by stacking the columns of \mathbf{X} on top of each other.

The following results about the normal and Wishart distributions and their derivations are given in classic multivariate statistics textbooks such as [25–27] and are used in the derivation of the EM-Algorithm, so we will not provide the proofs here.

Theorem 1. [27] (Theorem 2.2.9) Let $\mathbf{X} \sim N_{p,n}(M, \Sigma, \Psi)$. Then, $E[\mathbf{X}\mathbf{A}\mathbf{X}'] = tr(\Psi A)\Sigma + MAM'$.

Theorem 2. [27] (Theorem 2.4.12) Let $Y \sim W_n(\nu, \Psi)$ with Ψ and $\nu > n$. Then, the following holds:

1. The probability density function for Y is given by

$$p(Y) = \mathcal{W}_\nu(Y|\Psi) = \frac{|Y|^{(\nu-n-1)/2} \exp(-\frac{1}{2} \text{tr}(\Psi^{-1}Y))}{2^{\nu n/2} \pi^{n(n-1)/4} \prod_{i=1}^n \Gamma(\frac{\nu+1-i}{2}) |\Psi|^{nu/2}}$$

2. $E(Y) = \nu\Psi$, $\text{Cov}(Y) = 2\nu\Psi \otimes \Psi$.
3. If we assume Y is partitioned as

$$\begin{bmatrix} Y_a & Y_{ab} \\ Y_{ba} & Y_b \end{bmatrix},$$

where Y_a is $p \times p$ and Y_b is $q \times q$, $q + p = n$; and Ψ is partitioned as

$$\begin{bmatrix} \Psi_a & \Psi_{ab} \\ \Psi_{ba} & \Psi_b \end{bmatrix}$$

with the corresponding components to Y , then

- $Y_{a|b}$ is independent of (Y_{ab}, Y_b) ;
- $Y_b \sim W_q(\nu, \Psi_b)$;
- The conditional distribution of Y_{ab} given Y_b is multivariate Gaussian

$$N_{(n-q) \times q}(\Psi_{ab}\Psi_b^{-1}Y_b, \Lambda)$$

where $\Lambda_{ij,kl} = \text{Cov}(Y_{ij}, Y_{kl}|Y_{bb}) = \Psi_{ik}^{1|2}Y_{jl}$.

4. Similarly, by changing the order of the indices, we can show that

- $Y_{b|a}$ is independent of (Y_{ab}, Y_a) ;
- $Y_a \sim W_p(\nu, \Psi_a)$;
- The conditional distribution of Y_{ab} given Y_a is multivariate Gaussian

$$N_{(n-p) \times p}(\Psi_{ab}\Psi_a^{-1}Y_a, \Lambda)$$

where $\Lambda_{ij,kl} = \text{Cov}(Y_{ij}, Y_{kl}|Y_{aa}) = \Psi_{ik}^{1|2}Y_{jl}$

2.2. Combining Covariance Matrices with EM-Algorithm for the Wishart Distribution

2.2.1. Problem Definition

Let $A = \{a_1, a_2, \dots, a_m\}$ be the set of partially overlapping subsets of variables covering a set of K (i.e., $K = \cup_{i=1}^m a_i$) with total n variables. Let $G_{a_1}, G_{a_2}, \dots, G_{a_m}$ be the covariance matrices for variables in sets a_1, a_2, \dots, a_m , the sizes of the sets are given by n_1, n_2, \dots, n_m . We want to estimate the overall $n \times n$ covariance parameter Ψ using the sample $G_{a_1}, G_{a_2}, \dots, G_{a_m}$.

We will discuss two different algorithms for the estimation of Ψ using the sample covariance matrices $G_{a_1}, G_{a_2}, \dots, G_{a_m}$. One of the algorithms is based on the Wishart distribution with a single degree of freedom parameter and the other algorithm is based on the Wishart distribution with sample-specific degrees of freedom parameters. The choice between the two algorithms depends on the knowledge about the degrees of freedom of the sample covariance matrices. If all of the covariance matrices were obtained from similar experiments with similar precision, then the EM-Algorithm for the Wishart distribution with a single degree of freedom parameter should be preferred. If there are multiple degrees of freedom parameters, i.e., the precision of the sample covariance matrices varies significantly, then the EM-Algorithm for the Wishart distribution with sample-specific degrees of freedom values should be used.

2.2.2. EM Algorithm for Wishart Distribution with a Single Degree of Freedom

Given v , let $Y_{a_1} = vG_{a_1}, Y_{a_2} = vG_{a_2}, \dots, Y_{a_m} = vG_{a_m}$ be independent but partial realizations from a Wishart distribution with a known degree of freedom $v > n$ and a shape parameter Ψ . We want to estimate the overall covariance matrix Ψ after observing $G_{a_1}, G_{a_2}, \dots, G_{a_m}$. In the remainder, if we focus on a sample covariance matrix G_{a_i} , we drop the subscript and write G_a for notational economy (we perform the same with Y_{a_i}).

Theorem 3. *EM-Algorithm for the Wishart distribution with a single degree of freedom. Let v , let $Y_{a_1} = vG_{a_1}, Y_{a_2} = vG_{a_2}, \dots, Y_{a_m} = vG_{a_m}$ be independent but partial realizations from a Wishart distribution with a known degree of freedom $v > n$ and a shape parameter Ψ . Starting from an initial estimate of the covariance parameter matrix $\Psi^{(0)}$, the EM algorithm for Wishart distribution repeatedly updates the estimate of this matrix until convergence. The algorithm is given by:*

$$\Psi^{(t+1)} = \frac{1}{m} \sum_{a \in A} P_a \begin{bmatrix} G_a & G_a (B_{b|a}^{(t)})' \\ B_{b|a}^{(t)} G_a & \Psi_{b|a}^{(t)} + B_{b|a}^{(t)} G_a (B_{b|a}^{(t)})' \end{bmatrix} P_a', \quad (2)$$

where $B^{(t)}_{b|a} = \Psi_{ab}^{(t)'} (\Psi_a^{(t)})^{-1}$, $\Psi_{b|a}^{(t)} = \Psi_b^{(t)} - \Psi_{ab}^{(t)'} (\Psi_a^{(t)})^{-1} \Psi_{ba}^{(t)}$, a is the set of variables in the given partial covariance matrix, and b is the set difference of K and a . The matrices P_a are permutation matrices that put each completed covariance in the summation in the same order. The superscripts in parenthesis “ (t) ” denote the iteration number.

Proof. We write \bar{Y}_a for the completed version of Y_a obtained by complementing each of the observed data Y_a with the missing data components $Y_B = (Y_{ab}, Y_b)$ and assume \bar{Y}_a is partitioned as

$$\begin{bmatrix} Y_a & Y_{ab} \\ Y_{ba} & Y_b \end{bmatrix},$$

we partition $\Psi^{(t)}$ as

$$\begin{bmatrix} \Psi_a^{(t)} & \Psi_{ab}^{(t)} \\ \Psi_{ba}^{(t)} & \Psi_b^{(t)} \end{bmatrix},$$

where $\Psi_a^{(t)}$ is the part of the matrix that corresponds to the observed variables a , $\Psi_b^{(t)}$ is the part of the matrix that corresponds to the variables b , and $\Psi_{ab}^{(t)} = \Psi_{ba}^{(t)}$ is the part that corresponds to the covariance of the variables in a and b .

The likelihood function for the observed data can be written as

$$\begin{aligned} L(\Psi | v, Y_{a_1}, Y_{a_2}, \dots, Y_{a_m}) &= \prod_{i=1}^m W(Y_{a_i} | v, \Psi_{a_i}) \\ &= \prod_{i=1}^m \frac{|Y_{a_i}|^{(v-k_i-1)/2} \exp(-\frac{1}{2} \text{tr}(\Psi^{-1} Y_{a_i}))}{\left(2^{vk_i/2} \pi^{k_i(k_i-1)/4} \prod_{j=1}^{k_i} \Gamma(\frac{v+1-j}{2})\right) |\Psi_{a_i}|^{v/2}} \end{aligned}$$

The log-likelihood function with the constant terms combined in c is given by

$$l(\Psi | v, Y_{a_1}, Y_{a_2}, \dots, Y_{a_m}) = c - \frac{1}{2} \sum_{i=1}^m \left[\text{tr}(\Psi_{a_i}^{-1} Y_{a_i}) + v \log |\Psi_{a_i}| \right].$$

We can write the log-likelihood for the complete data up to a constant term as follows:

$$\begin{aligned}
\ell^c(\Psi | \nu, Y_{a_1}, Y_{a_2}, \dots, Y_{a_m}, Y_{B_1}, Y_{B_2}, \dots, Y_{B_m}) \\
= \frac{\nu - n - 1}{2} \left(\sum_{i=1}^m \log |Y_{a_i}| \right. \\
+ \sum_{i=1}^m |Y_{b_i} - Y'_{ab_i} Y_{a_i}^{-1} Y_{ab_i}|) \\
- \frac{\nu}{2} \left(\sum_{i=1}^m \log |\Psi_{a_i}| \right. \\
+ \sum_{i=1}^m \log |\Psi_{b_i} - \Psi'_{ab_i} \Psi_{a_i}^{-1} \Psi_{ab_i}|) \\
- \frac{1}{2} \text{tr}(\Psi^{-1} \sum_{i=1}^m \nu \bar{Y}_{a_i})
\end{aligned}$$

The expectation step of the EM-Algorithm involves calculating the expectation of the complete data log-likelihood conditional on the observed data and the value of Ψ at iteration t , which we denote by $\Psi^{(t)}$. We can write the expectation of the complete data log-likelihood up to a constant term as

$$\begin{aligned}
E \left[\ell^c(\Psi | \nu, Y_{a_1}, Y_{a_2}, \dots, Y_{a_m}, Y_{B_1}, Y_{B_2}, \dots, Y_{B_m}) | Y_{a_1}, Y_{a_2}, \dots, Y_{a_m}, \Psi^{(t)} \right] \\
= -\log |\Psi| - \text{tr}(\Psi^{-1} \frac{\sum_{i=1}^m E[\bar{Y}_{a_i} | Y_{a_i}, \Psi^{(t)}]}{\nu m}) \quad (3)
\end{aligned}$$

The maximization step of the EM algorithm updates $\Psi^{(t)}$ to $\Psi^{(t+1)}$ by finding the Ψ that maximizes the expected complete data log-likelihood. Using [25] (Lemma 3.3.2), the solution is given by:

$$\Psi^{(t+1)} = \frac{\sum_{i=1}^m E[\bar{Y}_{a_i} | Y_{a_i}, \Psi^{(t)}]}{\nu m}.$$

We need to calculate $E[\bar{Y}_{a_i} | G_{a_i}, \Psi^{(t)}]$ for each i , so we drop the index i in the remaining while deriving a formula for this term.

Firstly, $E[Y_a | Y_a, \Psi^{(t)}]$ is Y_a . Secondly, $Y_{ab} | Y_a, \Psi^{(t)}$ has a matrix-variate normal distribution with mean $Y_a \Psi_a^{(t)-1} \Psi_{ab}^{(t)}$. To calculate the expectation of Y_b , note that we can write this term as $Y_b = (Y_b - Y'_{ab} Y_a^{-1} Y_{ab}) + Y'_{ab} Y_a^{-1} Y_{ab}$. The distribution of the first term $(Y_b - Y'_{ab} Y_a^{-1} Y_{ab})$ is independent of Y_a and Y_{ab} , and is a Wishart distribution with degrees of freedom $\nu - n_a$ and covariance parameter $\Psi_b^{(t)} - \Psi_{ab}^{(t)\prime} \Psi_{ab}^{(t)-1} \Psi_a^{(t)}$. The second term is an inner product $(Y_a^{-\frac{1}{2}} Y_{ab})' (Y_a^{-\frac{1}{2}} Y_{ab})$. The distribution of $Y_a^{-\frac{1}{2}} Y_{ab}$ is a matrix-variate normal distribution with mean $Y_a^{-\frac{1}{2}} \Psi_a^{(t)-1} \Psi_{ab}^{(t)}$ and covariance structure given by $(\Psi_b^{(t)} - \Psi_{ab}^{(t)\prime} \Psi_a^{(t)-1} \Psi_{ab}^{(t)})$, I_{n_a} for the columns and rows, correspondingly. Therefore, the expectation of this inner-product is

$$(\Psi_b^{(t)\prime} \Psi_a^{(t)-1} Y_a \Psi_a^{(t)-1} \Psi_{ab}^{(t)} + n_a (\Psi_b^{(t)} - \Psi_{ab}^{(t)\prime} \Psi_a^{(t)-1} \Psi_{ab}^{(t)})).$$

This means that the expected value of Y_b given Y_a and $\Psi^{(t)}$ is

$$\begin{aligned}
& \Psi_{ab}^{(t)\prime} \Psi_a^{(t)-1} Y_a \Psi_a^{(t)-1} \Psi_{ab}^{(t)} + n_a (\Psi_b^{(t)} - \Psi_{ab}^{(t)\prime} \Psi_a^{(t)-1} \Psi_{ab}^{(t)}) \\
& + (\nu - n_a) (\Psi_b^{(t)} - \Psi_{ab}^{(t)\prime} \Psi_a^{(t)-1} \Psi_{ab}^{(t)}) \\
& = \nu (\Psi_b^{(t)} - \Psi_{ab}^{(t)\prime} \Psi_a^{(t)-1} \Psi_{ab}^{(t)}) + \Psi_{ab}^{(t)\prime} \Psi_a^{(t)-1} Y_a \Psi_a^{(t)-1} \Psi_{ab}^{(t)}.
\end{aligned}$$

Finally, putting $B^{(t)}_{b|a} = \Psi_{ab}^{(t)'} (\Psi_a^{(t)})^{-1}$, and $\Psi_{b|a}^{(t)} = \Psi_b^{(t)} - \Psi_{ab}^{(t)'} (\Psi_a^{(t)})^{-1} \Psi_{ba}^{(t)}$ leads to the iterative EM algorithm. \square

2.2.3. Asymptotic Standard Errors with a Single Degree of Freedom Parameter

Once the maximizer of $l(\Psi)$, $\hat{\Psi}$, has been found, the asymptotic standard errors can be calculated from the information matrix of Ψ evaluated at $\hat{\Psi}$. The log-likelihood is given by:

$$l(\Psi) = c - \frac{1}{2} \sum_{i=1}^m \left[\text{tr}(\Psi_{a_i}^{-1} Y_{a_i}) + v \log |\Psi_{a_i}| \right].$$

The first derivative with respect to the jk th element of Ψ is given by

$$\frac{\partial l(\Psi)}{\partial \psi_{jk}} = \frac{1}{2} \sum_{i=1}^m \left[\text{tr}(\Psi_{a_i}^{-1} \frac{\partial \Psi_{a_i}}{\partial \psi_{jk}} \Psi_{a_i}^{-1} Y_{a_i}) - v \text{tr}(\Psi_{a_i}^{-1} \frac{\partial \Psi_{a_i}}{\partial \psi_{jk}}) \right]$$

The derivative of the above for the lh th element of Ψ is given by

$$\frac{\partial^2 l(\Psi)}{\partial \psi_{jk} \partial \psi_{lh}} = \frac{1}{2} \sum_{i=1}^m \left[(-2 \text{tr}(\Psi_{a_i}^{-1} \frac{\partial \Psi_{a_i}}{\partial \psi_{jk}} \Psi_{a_i}^{-1} \frac{\partial \Psi_{a_i}}{\partial \psi_{lh}} \Psi_{a_i}^{-1} Y_{a_i}) + v \text{tr}(\Psi_{a_i}^{-1} \frac{\partial \Psi_{a_i}}{\partial \psi_{jk}} \Psi_{a_i}^{-1} \frac{\partial \Psi_{a_i}}{\partial \psi_{lh}})) \right]$$

The expected value of the second derivative is given by

$$\begin{aligned} E\left(\frac{\partial^2 l(\Psi)}{\partial \psi_{jk} \partial \psi_{lh}} \mid \Psi = \hat{\Psi}\right) &= \frac{1}{2} \sum_{i=1}^m \left[(-2 \text{tr}(\hat{\Psi}_{a_i}^{-1} \frac{\partial \Psi_{a_i}}{\partial \psi_{jk}} \hat{\Psi}_{a_i}^{-1} \frac{\partial \Psi_{a_i}}{\partial \psi_{lh}} \hat{\Psi}_{a_i}^{-1} E(Y_{a_i} \mid \Psi = \hat{\Psi})) + v \text{tr}(\hat{\Psi}_{a_i}^{-1} \frac{\partial \Psi_{a_i}}{\partial \psi_{jk}} \hat{\Psi}_{a_i}^{-1} \frac{\partial \Psi_{a_i}}{\partial \psi_{lh}})) \right] \\ &= -\frac{v}{2} \sum_{i=1}^m \left[\text{tr}(\hat{\Psi}_{a_i}^{-1} \frac{\partial \Psi_{a_i}}{\partial \psi_{jk}} \hat{\Psi}_{a_i}^{-1} \frac{\partial \Psi_{a_i}}{\partial \psi_{lh}}) \right] \end{aligned}$$

Therefore, the information matrix is given by

$$\{I(\Psi)\}_{jk,lh} = \{-E\left(\frac{\partial^2 l(\Psi)}{\partial \psi_{jk} \partial \psi_{lh}} \mid \Psi = \hat{\Psi}\right)\}_{jk,lh} = \frac{v}{2} \sum_{i=1}^m \left[\text{tr}(\hat{\Psi}_{a_i}^{-1} \frac{\partial \Psi_{a_i}}{\partial \psi_{jk}} \hat{\Psi}_{a_i}^{-1} \frac{\partial \Psi_{a_i}}{\partial \psi_{lh}}) \right].$$

The asymptotic variance–covariance matrix for $\hat{\Psi}$ is given by

$$V(\hat{\Psi}) = I(\hat{\Psi})^{-1}.$$

2.2.4. EM-Algorithm for the Wishart Distribution with Sample-Specific Degrees of Freedom Values

Theorem 4. *EM-Algorithm for the Wishart distribution with sample-specific degrees of freedom values. Assume the degrees for the sample of covariance matrices $G_{a_1}, G_{a_2}, \dots, G_{a_m}$ are given by v_1, v_2, \dots, v_m . Let $Y_{a_i} = v_i G_{a_i}$. Starting from an initial estimate of the genetic relationship matrix $\Psi^{(0)}$, the EM-Algorithm repeatedly updates the estimate of the genetic relationship matrix until convergence:*

$$\Psi^{(t+1)} = \frac{1}{\sum_{i=1}^m v_i} \sum_{a \in A} v_i P_a \begin{bmatrix} G_a & G_a (B_{b|a}^{(t)})' \\ B_{b|a}^{(t)} G_a & \Psi_{b|a}^{(t)} + B_{b|a}^{(t)} G_a (B_{b|a}^{(t)})' \end{bmatrix} P_a'. \quad (4)$$

Proof. The proof is similar to the proof of Theorem 3. The main difference is that the degrees of freedoms for the Wishart distributions are now sample-specific.

We can write the log-likelihood for the complete data up to a constant term as follows:

$$\begin{aligned}
& \ell^c(\Psi | \nu_1, \nu_2, \dots, \nu_m, Y_{a_1}, Y_{a_2}, \dots, Y_{a_m}, Y_{B_1}, Y_{B_2}, \dots, Y_{B_m}) \\
&= \sum_{i=1}^m \frac{\nu_i - n - 1}{2} (\log|Y_{a_i}| + |Y_{b_i} - Y'_{ab_i} Y_{a_i}^{-1} Y_{ab_i}|) \\
&\quad - \sum_{i=1}^m \frac{\nu_i}{2} (\log|\Psi_{a_i}| + \log|\Psi_{b_i} - \Psi'_{ab_i} \Psi_{a_i}^{-1} \Psi_{ab_i}|) \\
&\quad - \frac{1}{2} \text{tr}(\Psi^{-1} \sum_{i=1}^m \nu_i \bar{Y}_{a_i})
\end{aligned}$$

We can write the expectation of the complete data log-likelihood up to a constant term as

$$\begin{aligned}
& E \left[\ell^c(\Psi | \nu_1, \nu_2, \dots, \nu_m, Y_{a_1}, Y_{a_2}, \dots, Y_{a_m}, Y_{B_1}, Y_{B_2}, \dots, Y_{B_m}) | Y_{a_1}, Y_{a_2}, \dots, Y_{a_m}, \Psi^{(t)} \right] \\
&= -\log|\Psi| - \text{tr}(\Psi^{-1} \frac{\sum_{i=1}^m E[\bar{Y}_{a_i} | Y_{a_i}, \Psi^{(t)}]}{\sum_{i=1}^m \nu_i}). \tag{5}
\end{aligned}$$

Taking the derivative of the above expression with respect to Ψ and setting it to zero, we obtain the following update equation:

$$\Psi^{(t+1)} = \frac{\sum_{i=1}^m E[\bar{Y}_{a_i} | Y_{a_i}, \Psi^{(t)}]}{\sum_{i=1}^m \nu_i}.$$

The components of $E[\bar{Y}_{a_i} | Y_{a_i}, \Psi^{(t)}]$ can be calculated as before using the same methods as in the proof of Theorem 3. This completes the proof. \square

2.2.5. Asymptotic Standard Errors with Sample-Specific Degrees of Freedom Values

The information matrix for the case of sample-specific degrees of freedom values is obtained in a similar fashion as the Wishart distribution with a common degrees of freedom value. The information matrix is given by

$$\{I(\Psi)\}_{jk,lh} = \{-E(\frac{\partial^2 l(\Psi)}{\partial \psi_{jk} \partial \psi_{lh}} | \Psi = \hat{\Psi})\}_{jk,lh} = \frac{1}{2} \sum_{i=1}^m \left[\nu_i \text{tr}(\hat{\Psi}_{a_i}^{-1} \frac{\partial \Psi_{a_i}}{\partial \psi_{jk}} \hat{\Psi}_{a_i}^{-1} \frac{\partial \Psi_{a_i}}{\partial \psi_{lh}}) \right].$$

2.3. Sparse Estimation of Pooled Covariance Matrices

It is often useful to study the sparsity pattern of covariance and precision matrices. For a multivariate normal random variable, zeros in the covariance matrix correspond to marginal independence between variables, while zeros in the inverse covariance matrix (precision matrix) indicate a conditional independence between variables. An ℓ_1 -penalized maximum likelihood approach is a commonly used method for estimating these sparse matrices. This involves adding a term $p(\lambda) = -\lambda ||O \odot \Psi||_1$ (for sparsity in the covariance matrix) or $p(\lambda) = -\lambda ||O \odot \Psi^{-1}||_1$ (for sparsity in the precision matrix) to the likelihood function [28–31] for a nonnegative scalar value λ . A frequent choice for O is the matrix of all ones, although an alternative is to set $O_{ij} = 1$, for $i \neq j$, and $O_{ij} = 0$ otherwise, which shrinks the off-diagonal elements to zero.

We can incorporate the sparsity in our algorithm by adding the term $p(\lambda)$ to the expectation of the observed data log-likelihood function [32]. The ℓ_1 -penalized function that must be maximized in the EM maximization steps becomes

$$f(\Psi, \Psi^{(t)}) = -\log|\Psi| - \text{tr}(\Psi^{-1} \frac{\sum_{i=1}^m E[\bar{Y}_{a_i} | Y_{a_i}, \Psi^{(t)}]}{\sum_{i=1}^m \nu_i}) + p(\lambda).$$

At each iteration of the EM algorithm, this function can be maximized with respect to Ψ iteratively by using the methods in [29,31]. For instance, the following iterative algorithm can be used to maximize the ℓ_1 -penalized function with $p(\lambda) = \lambda ||O \odot \Psi||_1$ (see [31] for more details and caveats):

1. Set $S = \frac{\sum_{i=1}^m E[\bar{Y}_{a_i} | Y_{a_i}, \Psi^{(t)}]}{\sum_{i=1}^m v_i}$.
2. Set $\Psi_{(0)} = \Psi^{(t-1)}$.
3. For each iteration $j = 1, 2, \dots$, until convergence
 - set $\Psi_{(j)} = h(\Psi_{(j-1)} - v(\Psi_{(0)}^{-1} - \Psi_{(j-1)}^{-1} S \Psi_{(j-1)}^{-1}), \lambda t O)$, where $h(x, y)$ is the soft thresholding function defined as $h(x, y) = \text{sgn}(x)(|x| - y)_+$, $\text{sgn}(x)$ is a sign of the value x , $(x)_+ = \max(x, 0)$, and $v > 0$ is the learning rate parameter.
4. Set $\Psi^{(t)} = \Psi_{(j)}$.

2.4. Parametrized Covariance Matrices

In many covariance prediction problems such as SEM, the covariance matrix is assumed to have a certain parametric structure, i.e., the covariance matrix can be written as $\Psi(\theta)$ for a vector of parameters θ . In this case, the EM algorithm can still be used to estimate the parameters θ of the covariance matrix; however, the maximization step of the EM algorithm needs to be modified. The modified version of the expected likelihood function for the Wishart distribution with sample-specific degrees of freedom values is now expressed as

$$\begin{aligned} E\left[\ell^c(\theta | v_1, v_2, \dots, v_m, Y_{a_1}, Y_{a_2}, \dots, Y_{a_m}, Y_{B_1}, Y_{B_2}, \dots, Y_{B_m}) | Y_{a_1}, Y_{a_2}, \dots, Y_{a_m}, \Psi(\theta^{(t)})\right] \\ = -\log|\Psi(\theta)| - \text{tr}(\Psi(\theta)^{-1} \frac{\sum_{i=1}^m E[\bar{Y}_{a_i} | Y_{a_i}, \Psi(\theta^{(t)})]}{\sum_{i=1}^m v_i}). \end{aligned}$$

The maximization step of the EM algorithm now involves the maximization of the expected likelihood function with respect to the parameters θ , and numerical methods such as the Newton–Raphson method can be used for this purpose.

3. Illustrations

Illustration 1—Simulation study: Inferring the combined covariance matrix from its parts

To establish that a combined covariance can be inferred from realizations of its parts, we have conducted the following simulation study: In each round of the simulation, the true parameter value of the covariance matrix was generated as $\Sigma = \text{diag}(r_1, r_2, \dots, r_{N_{Total}}) + 0.3 * \mathbf{1}_{N_{Total} \times N_{Total}}$, where r_i were independently generated as $1 + 0.7 * u_i$ with u_i being a realization from the uniform distribution over $(0, 1)$. Σ was then adjusted by dividing its elements by the mean value of its diagonal elements. This parameter was taken as the covariance parameter of a Wishart distribution with 300 degrees of freedom, and N_{kernel} samples from this distribution are generated. After that, each of the realized covariance matrices was made partial by leaving a random sample of 10 to 40 (this number was also selected from the discrete uniform distribution for integers 10 to 40) variables in it. These partial kernel matrices were combined using the EM-Algorithm for the Wishart distribution iterated for 50 rounds (each round cycles through the partial covariance matrices in random order). The resultant combined covariance matrix $\hat{\Sigma}$ was compared with the corresponding parts of the parameter Σ . In certain instances, the union of the variables in the parts did not recover all of the N_{Total} variables; therefore, this calculation was based on the recovered part of the full covariance matrix by calculating the mean squared error between the upper diagonal elements of these matrices. This experiment was replicated 10 times for each value of $N_{Total} \in \{50, 75, 100, 200\}$ and $N_{kernel} \in \{10, 20, 30, 100\}$.

The results of this simulation study are summarized in Figure 1. For each covariance size, the MSEs decreased as the number of incomplete samples increased. On the other hand, as the size of the covariance matrix increased, the MSEs also increased.

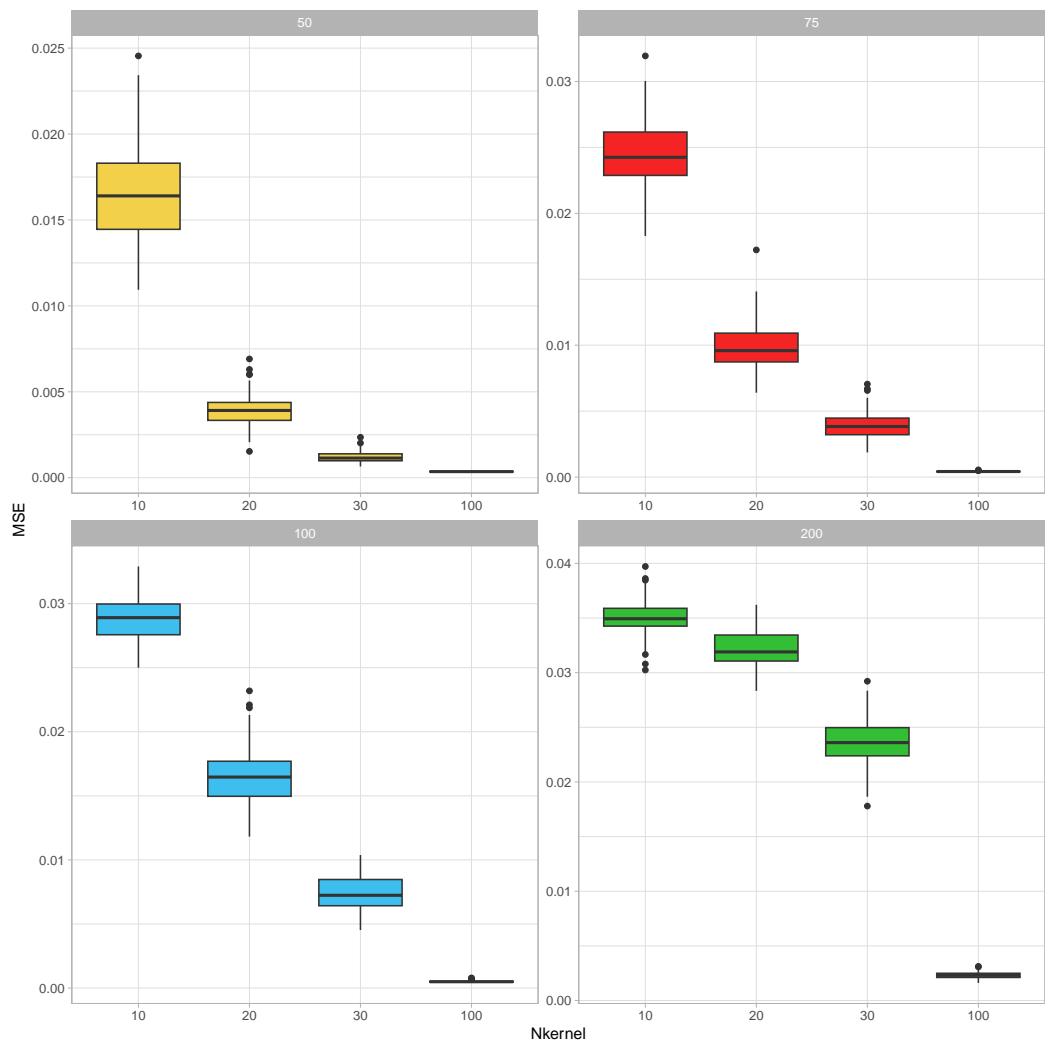


Figure 1. Illustration 1—Mean square errors (MSE) for estimating correlation parameters based on partial samples for $N_{Total} \in \{50, 75, 100, 200\}$ (number of variables in the covariance matrix) and $N_{kernel} \in \{10, 20, 30, 100\}$ (number of incomplete covariance matrix samples). Each incomplete covariance matrix had a random size between 10 to 40. The MSEs are calculated over 10 replications of the experiment.

Illustration 2—Simulation study: Likelihood Convergence

The EM-Algorithm for the Wishart distribution maximizes the likelihood function for a random sample of incomplete observations from a Wishart distribution. In this application, we explore the convergence of the algorithm for several instances, starting from several different initial estimates.

The example is composed of 10 experiments, each of which starts with a slightly different assumed Wishart covariance parameter $\Sigma = diag(\mathbf{b} + 1) + 0.21_{n \times n}$, where b_i for $i = 1, 2, \dots, n$ are i.i.d. uniform between 0 and 1. For each true assumed covariance matrix, we have generated 10 partial samples including between n_{min} and n_{max} variables (random at discrete uniform from n_{min} to n_{max}) each using the Wishart distribution. n , the total number of variables in the assumed covariance matrix was taken to be 100 or 1000. Corresponding to these two matrix sizes, the n_{min} and n_{max} are taken as 10 and 25 or 100 and 250. These 10 matrices are combined using the EM-Algorithm for the Wishart distribution 10 different times, each time using a slightly different initial estimate of the covariance parameter $\Sigma_0 = diag(0.5\mathbf{b} + 1) + 0.3 * b_0 \mathbf{1}_{n \times n}$ where b_i for $i = 0, 2, \dots, n$ are i.i.d. uniform between 0 and 1. We record the path of the log-likelihood function for all these examples.

At each instance of the parameter and a particular sample, the likelihood functions converged to the same point (See Figure 2). We have not observed any abnormalities in convergence according to these graphs.

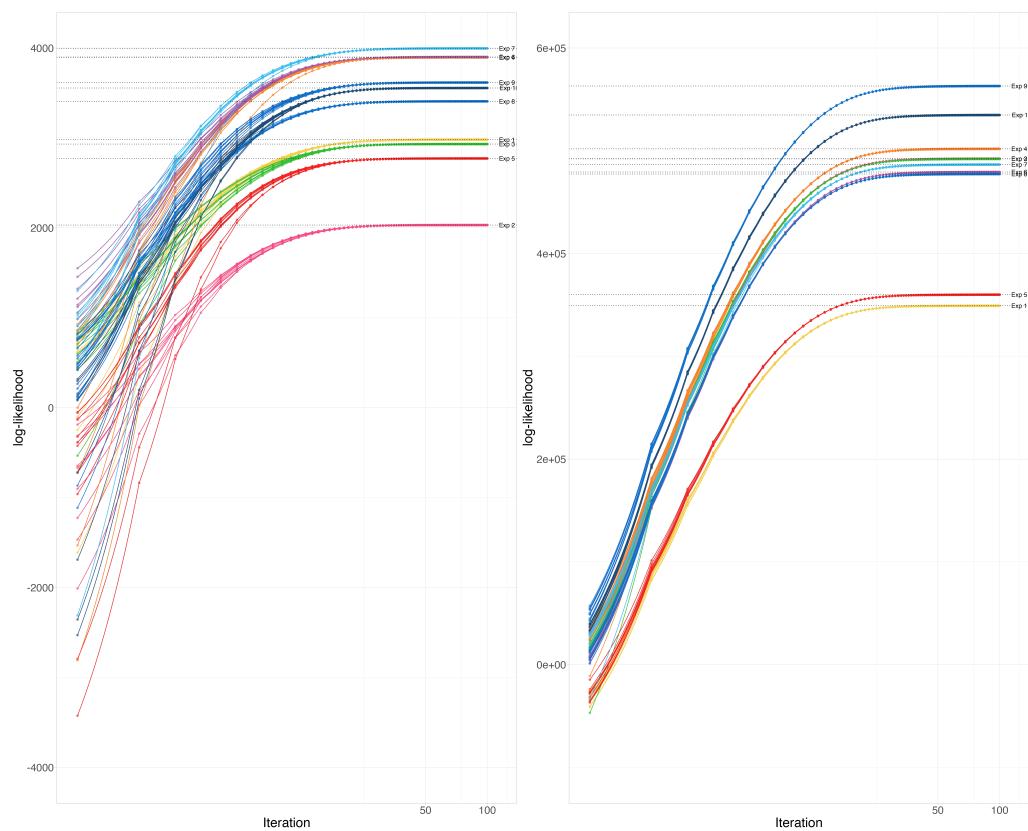


Figure 2. Illustration 2—Convergence of the log-likelihood function: Each color represents a different experiment. In each experiment, a sample of incomplete covariance matrices from a Wishart distribution was combined using the EM-Algorithm for the Wishart distribution starting from 10 slightly different random initial estimates. n , the total number of variables in the assumed covariance matrix, was taken to be 100 (A) or 1000 (B).

Illustration 3—Simulation study: Comparing the performance with MASEM

In this example, we compare our EM-based approach with the correlation pooling methods from the multivariate MASEM approach. We use the "tssem1" function contained within the "metaSEM" [33] R package for pooling covariance matrices using the MASEM approach. We use two different options for the random error structure in the MASEM approach: a diagonal error covariance matrix or a zero covariance matrix (since the variance component of the random effects is zero, the model becomes a fixed-effects model that is equivalent to the Generalized Least Squares (GLS) approach proposed by Becker [34]). When applying our method, we use the correlation matrices as covariance matrices, and before comparing it with the true parameter, we convert this matrix into a correlation matrix. The results are obtained by repeating the following simulation scenario 10 times. A random correlation matrix for 20 variables is generated. Using the random covariance matrix as the covariance parameter of a Wishart distribution with 100 degrees of freedom, we have generated three random covariance matrices. The first of these matrices contained the variables from 1 to 8, the second covariance matrix contained the variables 6 to 15 and the third matrix contained the variables 13 to 20. We have used our method as well as three forms of the MASEM approaches to obtain a complete covariance matrix estimate. We then compared the mean absolute error between the predicted and the true covariance matrix. The results, which are summarized in Figure 3, show that the EM-Algorithm for the Wishart distribution performs much better than the MASEM approaches in terms of the mean absolute errors and especially in terms of computation time. We also note that we could not make the MASEM algorithms produce any results for cases where the number of variables in the covariance matrix was larger than 30 in a reasonable time. This is due to the fact that the MASEM algorithms use a direct numerical optimization of the likelihood function, which becomes very difficult for unstructured covariance matrices. The EM-Algorithm for the Wishart distribution does not have this problem since each iteration of the algorithm depends solely on matrix

algebra operations. More importantly, both of the MASEM approaches we have tried resulted in the imputation of previously unobserved correlation values with the same estimated common covariance parameter value.

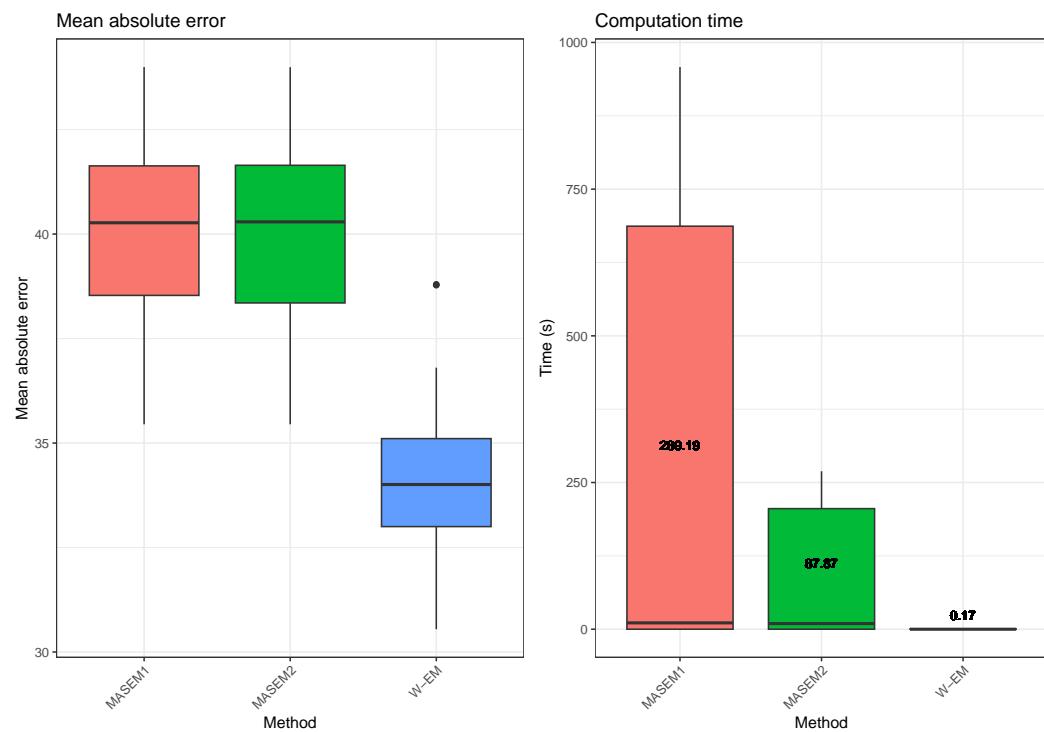


Figure 3. Illustration 3—Mean absolute errors and computation times for estimating the covariance matrix based on partial samples for two stage MASEM approaches and our EM-Algorithm for the Wishart distribution. The results are calculated over 30 replications of the experiment.

Illustration 4—Empirical study: Cassava data

The need to exploit genomic and statistical tools to harness the total effects of all the genes in the crop genome is gaining traction in most crops. In our illustrations, we used the world's largest and most updated public database for cassava (CassavaBase) from the Nextgen cassava funded project (<http://www.nextgencassava.org>). It is estimated that close to a billion people depend on cassava for their dietary needs, particularly in tropical regions of the world. We have accessed the data on 16 November 2019.

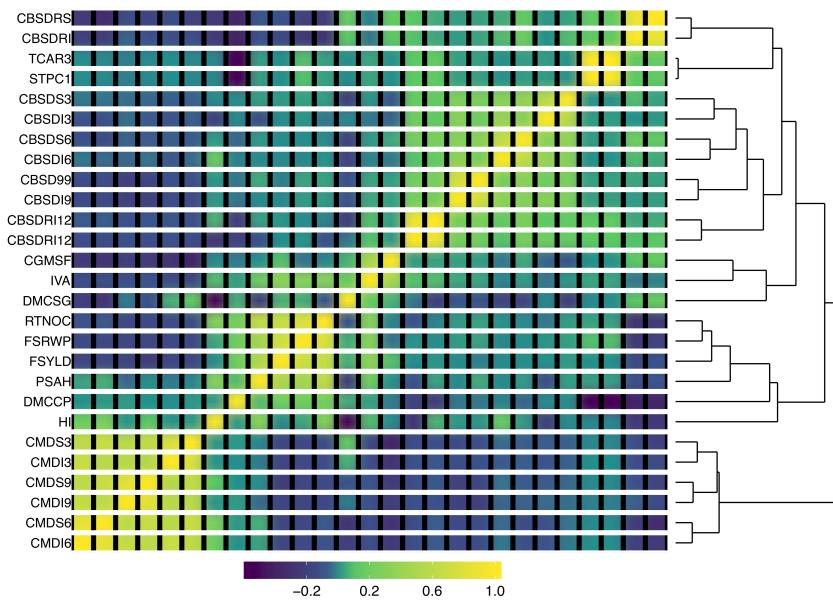
The initial data come from 135 phenotypic experiments performed by the East Africa cassava plant breeding programs. The dataset covers 81 traits and contains more than half a million phenotypic records. After filtering the outlier trait values, filtering the traits based on the number of records (at least 200 records per experiment), and trait-trial combinations also based on the number of records (at least 200 records for each trait in a given trial), a subset of 50 of these trials and a total of 37 traits were identified and used for this application (See Figure 4).



Figure 4. Illustration 4—This figure represents the availability of very unbalanced and heterogeneous phenotypic data for 37 traits in 50 trials in East Africa. The number of phenotypic records in each trial is represented with a colored box at the intersection of the trait (rows) and trial (columns). Gray squares represent unobserved traits and trial combinations. The mean number of traits observed in these studies was 10.38. Certain trait combinations were not observed together in these experiments; for example, root weight in water and dry matter content percentage, root number, and total carotenoid content.

Due to the relatively high cost of phenotypic experiments, they typically focus on a limited set of key traits. As a result, when observing multiple phenotypic datasets, the data is typically heterogeneous and incomplete, with certain trait combinations not appearing together in any of the experiments (e.g., root weight in water and dry matter content percentage, root number and total carotenoid content).

Figure 5 includes a heatmap of the resulting covariance matrix for 37 traits was obtained from combining the sample covariance matrices from 50 phenotypic trials. The heatmap indicates that there are three clusters of traits that appear to be positively correlated within each group, but little to no correlation between the groups. Two of these clusters correspond to disease-related traits, and the other is composed primarily of agronomic traits related to yield. The cluster related to cassava mosaic disease is found to be negatively correlated with the other disease traits, which are related to brown streak disease.



Abbrev	Trait	Abbrev	Trait
CBSDS3	Brown streak disease leaf severity 3 month	CMD13	Mosaic disease incidence 3 month
CBSDS6	Brown streak disease leaf severity 6 month	CMD16	Mosaic disease incidence 6 month
CBSD99	Brown streak disease leaf severity 9 month	CMD19	Mosaic disease incidence 9 month
CBSDR12	Brown streak disease root severity 12 month	SPRC3	Sprout count at three month
CMDS3	Mosaic disease severity 3 month	RTNOC	Root number
CMDS6	Mosaic disease severity 6 month	DMCCP	Dry matter content percentage
CMDS9	Mosaic disease severity 9 month	CBSDRS	Brown streak disease root severity
FSYLD	Fresh shoot weight measurement in kg per plot	WTFL	Spiraling whitefly severity visual 1-5
FSRWP	Fresh storage root weight per plot	DMCSG	Dry matter content by specific gravity method
HI	Harvest index	DMCSG2	Specific gravity
IVA	Initial vigor assessment 1-7	CGMSF	Green mite severity first
PSAH	Plant stands harvested	DMCRP	Dry matter content of fibrous root percentage
SOFTR15	Softness of boiled roots in newtons minute 15 COMP 0000076	RTWTPP	Fresh fibrous root weight per plant
SOFTR30	Softness of boiled roots in newtons minute 30 COMP 0000077	TCAR1	Total carotenoid by chart 1-8
SOFTR45	Softness of boiled roots in newtons minute 45 COMP 0000078	RTWT	Fresh storage root weight per plant
CBSDI3	Brown streak disease leaf incidence 3 month	STPC1	Root color visual rating 1-3
CBSDI6	Brown streak disease leaf incidence 6 month	TCAR3	Total carotenoid content in ug g
CBSDI9	Brown streak disease leaf incidence 9 month		
CBSDR12	Brown streak disease root incidence 12 months		

Figure 5. Illustration 4—Cassava data: The resulting combined covariance matrix for 37 traits obtained by combining the sample covariance matrices from 50 phenotypic trials downloaded from CassavaBase. Abbrev: Abbreviations.

We have used the R Package qgraph [35] to introduce sparsity to the off-diagonal elements of the estimated covariance matrix and graphically present this in Figure 6.

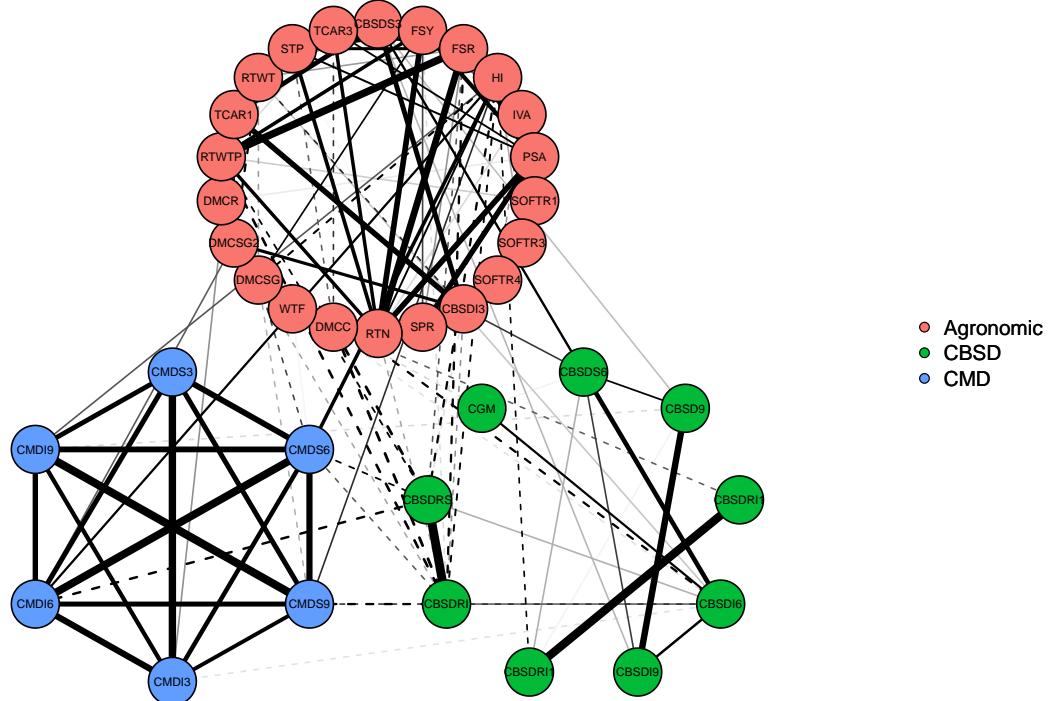


Figure 6. Illustration 4—Cassava data: A covariance graph for the covariance matrix resulting in a combined covariance matrix for 37 traits obtained by combining the sample covariance matrices from 50 phenotypic trials downloaded from CassavaBase. The green colored lines represent positive trait correlations, red colored lines represent negative trait correlations. The thickness of the line is an indication of the strength of the correlation. The traits that are not connected with any lines are estimated to be marginally independent. The abbreviations for the traits are the same as those given in Figure 5. Similar traits are grouped and have the same color nodes.

4. Conclusions

Analyzing data in large and heterogeneous databases remains a challenge due to the need for new statistical methodologies and tools to make inferences. The EM-Algorithm for the Wishart distribution is one such tool that can be used to solve a very specific problem: combining datasets using covariance matrices (similarly combining relationship or similarity matrices). Our approach is highly beneficial in terms of its statistical formalism and computational efficiency; to the best of our knowledge, this is the first time the EM procedure for pooling covariance matrices has been described, although it has been inspired by similar methods such as (conditional) iterative proportional fitting for the Gaussian distribution [22,23] and a method for combining a pedigree relationship matrix and a genotypic relationship matrix, which includes a subset of genotypes from the pedigree-based matrix [24] (namely, the H-matrix).

Despite the benefits of the proposed framework for combining heterogeneous datasets, certain limitations should be taken into account. Specifically, when combining data using covariance matrices, the original features are not imputed. It is known that the nature of missingness in data can significantly influence the performance of imputation and inference. Consequently, any approaches that disregard the missing data mechanisms are only applicable to data that is missing completely at random (MCAR) or missing at random (MAR). However, such techniques cannot be utilized for data not missing at random (NMAR) [36,37]. Additionally, there could be heterogeneity in covariance matrices to some extent. This can be addressed with a hierarchical distribution (see, e.g., [38]). Furthermore, this structural misspecification can also be accounted for by the method in [39].

Overall, the combination of heterogeneous datasets via covariances matrices and the EM-Algorithm for the Wishart distribution is novel, and we expect it to be beneficial in a variety of fields, such as physics, engineering, biology, neuroscience, finance, genomics, and other -omics disciplines.

Author Contributions: Conceptualization, D.A.; methodology, D.A.; software, D.A.; validation, D.A., M.S. and J.I.-S.; formal analysis, D.A.; data curation, D.A., M.S.; writing—original draft preparation, D.A.; writing—review and editing, D.A., M.S. and J.I.-S.; visualization, D.A. and J.I.-S. All authors have read and agreed to the published version of the manuscript.

Funding: D.A.: Support comes from the National Marrow Donor Program/Be The Match, part of a research collaboration that is CIBMTR. CIBMTR support includes funding from Public Health Service U24CA076518 from the National Cancer Institute and N00014-21-1-2954 from the Office of Naval Research. J.I.-S.: Support comes from the Beatriz GalindoProgram (BEAGAL18/00115) from the Ministerio de Educació n y Formació n Profesional of Spain and the Severo Ochoa Program for Centres of Excellence in R&D from the Agencia Estatal de Investigación of Spain, grant SEV-2016-0672 (2017–2021) to the CBGP.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All of the data used in this study is available from cassavabase.org. We have accessed the data on 16 November 2019.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Gomez-Cabrero, D.; Abugessaisa, I.; Maier, D.; Teschendorff, A.; Merkenschlager, M.; Gisel, A.; Ballestar, E.; Bongcam-Rudloff, E.; Conesa, A.; Tegnér, J. Data integration in the era of omics: Current and future challenges. *BMC* **2014**, *8*, 1–10. [[CrossRef](#)]
2. Rhodes, D.R.; Chinnaiyan, A.M. Integrative analysis of the cancer transcriptome. *Nat. Genet.* **2005**, *37*, S31. [[CrossRef](#)] [[PubMed](#)]
3. Evangelou, E.; Ioannidis, J.P. Meta-analysis methods for genome-wide association studies and beyond. *Nat. Rev. Genet.* **2013**, *14*, 379–389. [[CrossRef](#)]
4. Huttenhower, C.; Hibbs, M.; Myers, C.; Troyanskaya, O.G. A scalable method for integration and functional analysis of multiple microarray datasets. *Bioinformatics* **2006**, *22*, 2890–2897. [[CrossRef](#)]
5. Liu, J.; Huang, J.; Xie, Y.; Ma, S. Sparse group penalized integrative analysis of multiple cancer prognosis datasets. *Genet. Res.* **2013**, *95*, 68–77. [[CrossRef](#)] [[PubMed](#)]
6. Zhang, K.; Gray, J.W.; Parvin, B. Sparse multitask regression for identifying common mechanism of response to therapeutic targets. *Bioinformatics* **2010**, *26*, i97–i105. [[CrossRef](#)]
7. Kirk, P.; Griffin, J.E.; Savage, R.S.; Ghahramani, Z.; Wild, D.L. Bayesian correlated clustering to integrate multiple datasets. *Bioinformatics* **2012**, *28*, 3290–3297. [[CrossRef](#)]
8. Xiong, Q.; Ancona, N.; Hauser, E.R.; Mukherjee, S.; Furey, T.S. Integrating genetic and gene expression evidence into genome-wide association analysis of gene sets. *Genome Res.* **2012**, *22*, 386–397. [[CrossRef](#)] [[PubMed](#)]
9. Chen, Y.; Wu, X.; Jiang, R. Integrating human omics data to prioritize candidate genes. *BMC Med. Genom.* **2013**, *6*, 57. [[CrossRef](#)]
10. Troyanskaya, O.G.; Dolinski, K.; Owen, A.B.; Altman, R.B.; Botstein, D. A Bayesian framework for combining heterogeneous data sources for gene function prediction (in *Saccharomyces cerevisiae*). *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 8348–8353. [[CrossRef](#)]
11. Balbin, O.A.; Prensner, J.R.; Sahu, A.; Yocom, A.; Shankar, S.; Malik, R.; Fermin, D.; Dhanasekaran, S.M.; Chandler, B.; Thomas, D.; et al. Reconstructing targetable pathways in lung cancer by integrating diverse omics data. *Nat. Commun.* **2013**, *4*, 1–13. [[CrossRef](#)] [[PubMed](#)]
12. Bersanelli, M.; Mosca, E.; Remondini, D.; Giampieri, E.; Sala, C.; Castellani, G.; Milanesi, L. Methods for the integration of multi-omics data: Mathematical aspects. *BMC Bioinform.* **2016**, *17*, S15. [[CrossRef](#)]
13. Subramanian, I.; Verma, S.; Kumar, S.; Jere, A.; Anamika, K. Multi-omics data integration, interpretation, and its application. *Bioinform. Biol. Insights* **2020**, *14*, 1177932219899051. [[CrossRef](#)]
14. Jiang, D.; Armour, C.R.; Hu, C.; Mei, M.; Tian, C.; Sharpton, T.J.; Jiang, Y. Microbiome multi-omics network analysis: Statistical considerations, limitations, and opportunities. *Front. Genet.* **2019**, *10*, 995. [[CrossRef](#)] [[PubMed](#)]
15. Cheung, M.W.L.; Chan, W. Meta-analytic structural equation modeling: A two-stage approach. *Psychol. Methods* **2005**, *10*, 40. [[CrossRef](#)] [[PubMed](#)]
16. Jak, S. *Meta-Analytic Structural Equation Modelling*; Springer: Berlin, Germany, 2015.
17. Akdemir, D.; Knox, R.; Isidro y Sánchez, J. Combining partially overlapping multi-omics data in databases using relationship matrices. *Front. Plant Sci.* **2020**, *11*, 947. [[CrossRef](#)]
18. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
19. Dempster, A.P.; Rubin, D.B.; Tsutakawa, R.K. Estimation in covariance components models. *J. Am. Stat. Assoc.* **1981**, *76*, 341–353. [[CrossRef](#)]
20. Dempster, A.; Laird, N.; Rubin, D. Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B* **1977**, *39*, 1–38.

21. Hastie, T.; Mazumder, R.; Lee, J.D.; Zadeh, R. Matrix completion and low-rank SVD via fast alternating least squares. *J. Mach. Learn. Res.* **2015**, *16*, 3367–3402.
22. Cramer, E. Conditional iterative proportional fitting for Gaussian distributions. *J. Multivar. Anal.* **1998**, *65*, 261–276. [[CrossRef](#)]
23. Cramer, E. Probability measure with given marginals and conditionals: I-projections and conditional iterative proportional fitting. *Stat. Risk Model.* **2000**, *18*, 311–330.
24. Legarra, A.; Aguilar, I.; Misztal, I. A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* **2009**, *92*, 4656–4663. [[CrossRef](#)]
25. Anderson, T. *An Introduction to Multivariate*; Wiley & Sons: Hoboken, NJ, USA, 1984.
26. Gupta, A.; Nagar, D. *Matrix Variate Distributions*; Chapman and Hall/CRC Monographs and Surveys in Pure and Applied Mathematics; Chapman and Hall: Boca Raton, FL, USA, 2000.
27. Kollo, T.; von Rosen, D. *Advanced Multivariate Statistics with Matrices*; Springer Science & Business Media: Berlin, Germany, 2006; Volume 579.
28. Bickel, P.J.; Levina, E. Covariance regularization by thresholding. *Ann. Stat.* **2008**, *36*, 2577–2604. [[CrossRef](#)] [[PubMed](#)]
29. Friedman, J.; Hastie, T.; Tibshirani, R. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* **2008**, *9*, 432–441. [[CrossRef](#)]
30. Lam, C.; Fan, J. Sparsistency and rates of convergence in large covariance matrix estimation. *Ann. Stat.* **2009**, *37*, 4254. [[CrossRef](#)]
31. Bien, J.; Tibshirani, R.J. Sparse estimation of a covariance matrix. *Biometrika* **2011**, *98*, 807–820. [[CrossRef](#)] [[PubMed](#)]
32. Green, P.J. On use of the EM algorithm for penalized likelihood estimation. *J. R. Stat. Soc. Ser. B* **1990**, *52*, 443–452. [[CrossRef](#)]
33. Cheung, M.W.L. metaSEM: An R package for meta-analysis using structural equation modeling. *Front. Psychol.* **2015**, *5*, 1521. [[CrossRef](#)]
34. Becker, B.J. Missing Data and the Synthesis of Correlation Matrices. In Proceedings of the Annual Meeting of the American Educational Research Association, San Francisco, CA, USA, 20–24 April 1992.
35. Epskamp, S.; Cramer, A.O.J.; Waldorp, L.J.; Schmittmann, V.D.; Borsboom, D. qgraph: Network Visualizations of Relationships in Psychometric Data. *J. Stat. Softw.* **2012**, *48*, 1–18. [[CrossRef](#)]
36. Little, R.; Rubin, D. *Statistical Analysis with Missing Data*; Wiley: New York, NY, USA, 2002.
37. Rubin, D.B. Inference and missing data. *Biometrika* **1976**, *63*, 581–592. [[CrossRef](#)]
38. Ke, Z.; Zhang, Q.; Tong, X. Bayesian meta-analytic SEM: A one-stage approach to modeling between-studies heterogeneity in structural parameters. *Struct. Equ. Model. Multidiscip. J.* **2019**, *26*, 348–370. [[CrossRef](#)]
39. Chen, C.F. Bayesian inference for a normal dispersion matrix and its application to stochastic multiple regression analysis. *J. R. Stat. Soc. Ser. B* **1979**, *41*, 235–248. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.