Copula Gaussian Multiscale Graphical Models with Application to Geophysical Modeling

Hang Yu, Justin Dauwels, Xu Zhang, Shiyan Xu, and Wayne Isaac T. Uy
School of Electrical and Electronics Engineering, School of Physical and Mathematical Sciences
Nanyang Technological University, 50 Nanyang Avenue, Singapore 639798

Abstract—Gaussian Multiscale graphical models are powerful tools to describe high-dimensional spatial data; they capture long-range statistical dependencies among distant sites by introducing coarser scales. However, such models are only applicable to Gaussian data. In this paper, a new class of copula Gaussian multiscale graphical models is proposed which possesses rich modeling capabilities and computational efficiency while eliminating the Gaussian assumption. Numerical results are presented for synthetic data as well as data from a few applications in geophysics – models of sea surface temperature and Asian rainfall patterns.

I. INTRODUCTION

Multiscale or multiresolution models [1] have attracted tremendous interest in a wide variety of disciplines, especially for modeling phenomena exhibiting distinctive behavior over a range of scales or resolutions. They have been successfully applied to fields such as atmospheric studies [2], oceanography [3], and natural image processing [4]. One of the most common settings for representing multiscale models is the sparse graphical model which enables efficient representation and computation even for high-dimensional data.

For instance, global sea surface temperature can be modeled using a quadtree, a typical family of multiscale models, by associating the finest scale nodes with the original temperature measuring stations. Under this arrangement, the root node can be regarded as a kind of average global temperature while its children capture the deviations from this mean, and so forth at increasingly finer scales. The long-range statistical dependencies (e.g., between the Arctic Ocean and the Southern Ocean) can be captured at coarser scales. The resulting joint covariance matrix of all the sites in the finest scale is intractable due to the large number of entries. Nevertheless, multiscale graphical models provide a sparse description of the data, significantly improving the efficiency of representation and inference.

When the variables corresponding to all nodes in all scales are jointly Gaussian distributed, the structure of the graphical model can be defined by the precision matrix (inverse covariance matrix); a zero element corresponds to the absence of an edge in the graphical model or the conditionally independence between two variables. Recently, Choi et al. [5] proposed a class of Gaussian multiscale graphical models in which variables at each scale have a sparse conditional covariance structure conditioned on other scales. This particular Sparse In-scale conditional covariance Multiscale (SIM) model provides a good fit to the data with a small number of parameters.

While Gaussian models are powerful tools, they have well-known limitations when applied to some real-world problems where the data is manifestly non-Gaussian. For instance, in the context of geophysics, atmospheric and oceanic phenomena often exhibit non-Gaussian statistics. In this paper, we relax the assumption of Gaussianity by means of the Gaussian copula [6], making the proposed Copula Gaussian Sparse In-scale conditional covariance Multiscale (CSIM) graphical model applicable to a wider family of distributions than the original SIM model.

In our experiments on synthetic data, the CSIM model recovers the underlying structure of the data more precisely than the original SIM model. Next we use the problem of imputing missing data to benchmark CSIM with the SIM model, the Gaussian monoscale model [7], and the copula Gaussian monoscale model [8]. The CSIM model yields more accurate estimates, especially when the observed data is sparse. We also consider two data sets from geophysics, concerning sea surface temperature and Asian rainfall patterns. Results from these two real data sets show that the CSIM model provides an effective tool to estimate the missing data based on previous information.

The remainder of the paper is organized as follows. In Section II, we first review the Gaussian copula graphical model and the Gaussian multiscale graphical model with sparse in-scale conditional covariance. We then present the proposed copula Gaussian sparse in-scale conditional covariance multiscale graphical model. In Section III, we explain how to construct the model and how to estimate the missing data based on the learnt graphical model. In Section IV, we assess the proposed model and benchmark it with other models using the synthetic data. Application to real problems and analysis of the results are also presented in this section. Finally, we offer concluding remarks in Section V.

II. COPULA GAUSSIAN MULTISCALE GRAPHICAL MODEL

In the following, we briefly describe the Gaussian copula graphical model, the Gaussian multiscale graphical model, and the proposed copula Gaussian multiscale graphical model.

A. Gaussian Copula Graphical Model

We denote the observed non-Gaussian variables and hidden Gaussian variables as $Y_1, \ldots, Y_P$ and $Z_1, \ldots, Z_P$ respectively.
A Gaussian copula graphical model is defined as [6]:

\[
Z \sim N(0, K^{-1})
\]

\[
Y_k = F_k^{-1}(\Phi(Z_k)),
\]

where \( K \) is the precision matrix whose inverse (the covariance matrix) has normalized diagonal, \( \Phi \) is the CDF (cumulative distribution function) of the standard Gaussian distribution, and \( F_k \) is the CDF of \( Y_k \). The latter is often approximated by the empirical distribution \( \hat{F}_k \). Note that \( F_k^{-1} \) is the pseudo-inverse of \( F_k \), which is defined as:

\[
F^{-1}(y) = \inf_{x \in \mathcal{X}} \{ F(x) \geq y \}. \tag{3}
\]

### B. Gaussian Multiscale Models with Sparse In-scale Conditional Covariance (SIM Model)

The SIM model [5] has tree-structured connections between different scales and sparse conditional covariance structure at each scale. In detail, within each scale, a variable is correlated with only a few other variables in the same scale conditioned on variables at scales above and below. If we describe the adjacency graph corresponding to the sparse conditional covariance using a graph called conjugate graph (i.e., an edge in the conjugate graph if the corresponding entry in the conditional covariance is zero), the SIM model has a sparse graphical model for interscale structure and a sparse conjugate graph for in-scale structure.

As an illustration, a SIM model is shown in Fig. 1(a). The dashed lines within each scale correspond to the edges in the sparse conjugate graph. It is often the case that the inverse of a sparse matrix is not sparse any more. Therefore, the corresponding graphical model is dense within each scale. Note that the in-scale conditional precision matrix is the inverse of the sparse in-scale conditional covariance.

SIM model has an interesting structure. As an example of Fig. 1(a), its precision matrix \( K \) can be decomposed as:

\[
K = K_t + K_c, \tag{4}
\]

where

\[
K_t = \begin{bmatrix}
K_{[1]} & 0 \\
K_{[2,1]} & K_{[2,2]} \\
0 & K_{[3,2]} \\
0 & 0 & K_{[3,3]}
\end{bmatrix}
\]

\[
K_c = \begin{bmatrix}
0 & 0 \\
0 & K_{[2,2]} \\
0 & 0 & K_{[3,3]}
\end{bmatrix}. \tag{5}
\]

In the above, the submatrix \( K_{[m]} \) of \( K \) corresponds to the conditional precision matrix of variables at scale \( m \) conditioned on other scales. As illustrated in Fig. 1(b), a SIM model has a densely connected graphical model within each scale which prevents \( K_{[m]} \) from being a sparse matrix. However, \( K_{[m]}^{-1} \) is the conditional covariance of scale \( m \), which is sparse as shown in the conjugate graph for this scale (Fig. 1(a)). The submatrix \( K_{[m_1,m_2]} \) is sparse with respect to the graphical model structure connecting scale \( m_1 \) and \( m_2 \). Since each scale only has connections with its neighboring scale(s), \( K_{[m_1,m_2]} \) is a zero matrix if \(|m_1 - m_2| > 1 \). The matrix \( K \) can then be decomposed as a sum of \( K_t \), representing the hierarchical interscale tree structure, and \( K_c \), representing the conditional in-scale structure. By letting \( \Sigma_c = K_c^{-1} \), where \( \Sigma_c \) is sparse since \( K_c \) is a block-diagonal matrix, the precision matrix \( K \) of a SIM model can therefore be decomposed as a sum of a sparse matrix and the inverse of a sparse block-diagonal matrix:

\[
K = K_t + \Sigma_c^{-1}. \tag{7}
\]

### C. Multiscale Copula Gaussian Model with Sparse In-scale Conditional Covariance (CSIM Model)

A CSIM model with a quadtree as the interscale structure is shown in Fig. 2. Given i.i.d. samples of the observed non-Gaussian variables \( Y \), we first transform the non-Gaussian observed variables \( Y \) into Gaussian distributed latent variables \( Z \) (associated with the observed variables \( Y \)):

\[
Z_k = \Phi^{-1}(\hat{F}_k(Y_k)), \tag{8}
\]
where Φ is the CDF of the standard Gaussian distribution and \( F_k \) is the empirical CDF of \( Y_k \). We then regard \( Z \) as the finest scale variables in a Gaussian multiscale graphical model with sparse in-scale conditional covariance (SIM).

By considering the empirical covariance of the finest scale as the target matrix, the method of log-determinant maximization can be used to find the precision matrix \( K \) with the quadtree interscale structure and the sparse in-scale condition covariance. The problem can then be formulated as the following, cf. [5]:

\[
\hat{K} = \arg\max_{K>0} \sum_m \log \det K[m] + \sum_{(i,j) \in E_{\text{inter}}} |K(i,j)|
\]

\[
\text{s.t. } |K(i,j) - K^*(i,j)| \leq \gamma \quad \forall (i,j)
\]

(9)

where \( K[m] \) is the in-scale precision matrix at scale \( m \) and \( E_{\text{inter}} \) is the set of all possible interscale edges connecting successive neighboring scales. The objective function as well as the constraints can be decomposed into an interscale component and in-scale components, and can thus be solved using a two-step algorithm which will be elaborated in Section III-A. In brief, an expectation-maximization algorithm is first used to construct a multisresolution tree model so that the marginal covariance of the finest scale matches the target covariance \( \Sigma_F^* \). We then include correlations within each scale to further approximate the target covariance.

### III. Learning and Inference

In this section, we briefly introduce the learning method of the CSIM model given the i.i.d. non-Gaussian observed variables. Since our objective is to estimate the missing data in the testing samples, we further introduce the inference method for the multiscale copula Gaussian graphical model.

#### A. Learning the CSIM model

As mentioned in the last section, given i.i.d. non-Gaussian observed variables, we first apply the Gaussian copula (8) to obtain Gaussian distributed variables \( Z \), compute the empirical covariance \( \Sigma_F^* \) of \( Z \), and regard it as the target covariance that the marginal covariance corresponding to the finest scale of the multiscale model will match [5].

Second, we select a proper multiscale tree structure (e.g., quadtree), in which the transformed Gaussian latent variables \( Z \) are located at the finest scale and that additional hidden variables exist in the coarser scales. The expectation-maximization algorithm can then be used to estimate the parameters in precision matrix \( K \) corresponding to the tree graphical model. The resulting marginal covariance of the finest scale in the multiscale tree model has the minimum KL-divergence from the target covariance given the selected tree structure.

Third, since all the scales are connected by a tree model, the marginal covariance matrix at each scale can be represented in terms of the marginal covariance at the next finer scale as

\[
\Sigma[m] = A[m] \Sigma[m+1] A^T[m] + Q[m],
\]

where \( A[m] \) and \( Q[m] \) are determined by \( K_i \). In order to modify the tree model so that the marginal covariance at the finest scale becomes \( \Sigma_F^* \), we set \( \Sigma[M] = \Sigma_F^* \) for the finest scale \( M \) and compute the target marginal covariance for each scale in a fine-to-coarse manner [5].

Fourth, we aim to compute the target conditional precision matrix for each scale \( J^*[m] \). Specifically, we partition the joint precision matrix of the multiscale model as

\[
K = \begin{bmatrix}
K_c^* & K_{c,m}^* & 0 \\
K_{c,m}^* & K_{m}^* & K_{m,f}^* \\
0 & K_{m,f}^* & K_f^*
\end{bmatrix},
\]

with the marginal covariance at scale \( m \) being

\[
\Sigma[m] = (K_{m}^* - K_{m,c}^*(J_{c}^*[c])^{-1}K_{c,m}^*)^{-1}.
\]

(12)

By setting \( \Sigma[m] = \Sigma_F^* \), the target conditional precision matrix at scale \( m \) can be computed as [5]:

\[
K_{m}^* = (J_{c}^*[c]^{-1} + K_{m,c}^*(J_{c}^*[c])^{-1}K_{c,m}^* + K_{m,f}^*(J_{f}^*[f])^{-1}K_{f,m}^*^{-1}).
\]

(13)

Fifth, before computing the target conditional precision matrix for the next finer scale, we perform structure optimization at scale \( m \) to obtain a sparse in-scale conditional covariance approximation. This approximation can be performed by solving a simplified version of the convex optimization problem in (9):

\[
\hat{K}_m = \arg\max_{K>0} \log \det K[m]
\]

\[
\text{s.t. } |K(i,j) - K^*(i,j)| \leq \gamma \quad \forall (i,j) \in \mathcal{V}[m],
\]

(14)

where \( \mathcal{V}[m] \) is the set of nodes at scale \( m \) and the resulting \( K[m] \) has a sparse inverse. We replace \( K^*[m] \) with \( \hat{K}_m \) and proceed to the next finer scale. Modifying the conditional covariance structure in a coarse-to-fine manner, we can learn a CSIM model approximation of the given target marginal covariance of the finest scale \( \Sigma_F^* \).

#### B. Inferring Missing Data

After learning the multiscale graphical model using training data, we aim to estimate the missing components in the testing data. Given a testing sample vector with missing components, we first compute its corresponding value in the latent Gaussian layer using (8), where \( F_k \) is the empirical CDF of the training data. We then partition the transformed sample vector \( Z = (Z_O, Z_M) \), where \( Z_O \) are the observed components and \( Z_M \) are the missing components. It follows that

\[
\left( \begin{array}{c}
Z_M \\
Z_O
\end{array} \right) \sim \mathcal{N}\left( \begin{array}{c}
0 \\
0
\end{array} \right), \left( \begin{array}{c}
\Sigma_M, \\
\Sigma_{OM}, \\
\Sigma_O
\end{array} \right)
\]

(15)

and thus,

\[
Z_M|Z_O \sim \mathcal{N}(\Sigma_{MO}\Sigma_{O}^{-1}Z_O, \Sigma_M - \Sigma_{MO}\Sigma_{O}^{-1}\Sigma_{OM}).
\]

(16)

The maximum likelihood estimation of \( Z_M \) conditioned on \( Z_O \) is the conditional mean \( \Sigma_{MO}\Sigma_{O}^{-1}Z_O \). To further make use of the sparsity of the precision matrix \( K \), we compute
\( \hat{Z}_M \) as follows:

\[
\hat{Z}_M = -K_M^{-1}K_M^{O}Z_O.
\]  

(17)

We then obtain the missing components in the observed non-Gaussian sample vector using (2), where \( \hat{F}_k \) is the empirical CDF of training data.

IV. NUMERICAL RESULTS

In this section, we test the proposed CSIM model against existing graphical models (i.e., the Gaussian multiscale graphical model [5], the Gaussian monoscale graphical model [7], and the copula Gaussian monoscale graphical model [8]) on synthetic and real data sets.

A. Synthetic Data

We generate non-Gaussian synthetic data as follows:

1) Generate a \( 1024 \times 1024 \) covariance matrix for the finest scale. The 1024 variables are arranged spatially on a \( 32 \times 32 \) grid. Specifically, the diagonal elements \( \Sigma_{ii} \) equal to 1.5 while the off-diagonal elements are \( \Sigma_{ij} = d(i,j)^{-1/2} \), where \( d(i,j) \) is the spatial distance between nodes \( i \) and \( j \).

2) Generate a \( N \times P \) Gaussian dataset corresponding to the above covariance matrix, where \( N \) is the sample size (6000) and \( P \) is the number of dimensions (1024).

3) Apply different types of copula to each variable, including beta, exponential, and chi-square copula, transforming the Gaussian variables to continuous non-Gaussian variables.

4) Normalize the non-Gaussian data set to have zero mean and unit variance to make its empirical covariance more tractable.

We first proceed by learning the multiscale graphical model using 5000 samples. The results of the CSIM and the SIM model in every scale are shown in Fig. 3. The true structure is constructed by setting the target covariance of the SIM model to be the covariance given in the first step of synthetic data generation procedure. It is evident that the CSIM model outperforms the SIM model with a better recovery of the true structure.

A sample from the rest of the 1000 samples (testing data) is then randomly selected and some components (624) in the \( P \) dimensional sample vector are randomly removed. The contour map of \( 32 \times 32 \) grid corresponding to this sample vector without missing data is shown in Fig. 4(a). The observed and unobserved components in the grid are depicted by white and black pixels, respectively, as shown in Fig. 4(b). Our objective then is to estimate the unobserved components given the observed ones. By applying different graphical models, we obtain estimates shown in Fig. 4(c) to Fig. 4(f) with the corresponding mean absolute errors 0.4742, 0.4193, 0.4686, and 0.3805. We can see that the result of CSIM approximates the true value the most accurately, potentially since it captures long-range dependencies and is applicable to non-Gaussian data. The copula Gaussian monoscale model performs well but is outperformed by the CSIM model since the former fails to model the long-range dependencies. Interestingly, the estimates of the Gaussian multiscale model (SIM) and the Gaussian monoscale model are less accurate, probably because those models are designed for Gaussian data.

We further plot the mean absolute error as a function of the number of observed components averaged over 100 trials, which is shown in Fig. 5. As the number of observations decreases, both CSIM and SIM models outperform their monoscale counterparts, suggesting that modeling long-range dependencies is crucial to infer missing data from sparse observations.
Fig. 4. Results of different methods inferring the missing values of synthetic data: (a) True graph; (b) Observed components (white) and unobserved components (black) in the grid; (c) Estimate by the Gaussian monoscale graphical model; (d) Estimate by the copula Gaussian monoscale graphical model; (e) Estimate by the Gaussian multiscale graphical model; (f) Estimate by the copula Gaussian multiscale graphical model (The mean absolute errors corresponding to the 4 models are 0.4742, 0.4193, 0.4686, and 0.3805 respectively).

Fig. 5. Mean absolute error as a function of number of observed sites averaged over 100 trials for inferring missing data in synthetic dataset.

B. Sea Surface Temperature

The data set of global sea surface temperature is recorded daily since Aug. 31, 1981 until the present by the NASA GHRSSST L4 AVHRR OI satellite [9]. In practice, due to the limited satellite path or the presence of cloud cover, there are always missing data in satellite images, as illustrated by the grey area in Fig. 6(a). Note that the available data is limited – no data is available for large regions. From the satellite image, we wish to recover the temperature at the unobserved sites (Fig. 6(b)).

Our training data consists of 3000 samples corresponding to

Fig. 6. Sea surface temperature: (a) Satellite observations; (b) True temperature.
3000 randomly selected days in the period between 1981 and 2010 from 256 sites arranged on a $16 \times 16$ grid in the Pacific Ocean. This is represented by the white square as shown in Fig. 6(b). Records in the year 2011 are then used for the testing purposes. In Fig. 7, we show the histogram of the temperature distribution at 5 random sites. As can be seen from the figure, some histograms cannot be approximated effectively by Gaussian distributions. The underlying distributions are non-Gaussian.

The temperature estimates for one trial with 60 observed components are shown in Fig. 8. The mean absolute estimation error as a function of the number of observed sites averaged over 100 trials is shown in Fig. 9. As can be observed, the Gaussian copula does not yield a significant difference in this case due to the strong correlation (or anti-correlation) among the temperature changes for all the sites. Fig. 10 shows the observed non-Gaussian temperature time series and the latent Gaussian temperature time series reconstructed from copula Gaussian models using (8) at 3 sites. Strong correlation exists for both the observed and latent variables, making the empirical covariances similar for Gaussian models and copula Gaussian models.

C. Precipitation Data

In this section, we consider the GPCC Precipitation data provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA [10]. The data set covers the monthly global precipitation from 1951 to 2004. We select 256 sites arranged on a $16 \times 16$ grid located in east Asia, mainly in China; the 600 samples from 1951 to 2000 are used as training data and the rest are used as testing data. Fig. 11 shows the histogram of the precipitation distribution at 5 random sites. Since the data is non-Gaussian, it is expected that the copula Gaussian multiscale model will perform well on this data set.

Fig. 12 shows the result for one trial given 150 observed sites while Fig. 13 shows the mean absolute error as a function of the number of observed sites. Clearly, the CSIM model yields the lowest estimation error. As the number of observed sites increases, the monoscale copula Gaussian graphical model is also capable of inferring the missing data with high accuracy, since neighborhood dependencies only suffice to estimate the precipitation of an unobserved site. Interestingly, the Gaussian monoscale model performs better than the SIM model when the number of observations is less than 60, whereas the CSIM model outperforms the monoscale copula Gaussian model for a large range of observation number. This suggests that the Gaussian multiscale model may be more sensitive to the distribution of the data.
To compare these results with the sea surface temperature data, we plot the precipitation at 3 sites (two neighboring sites and two distant sites) with respect to time in Fig. 14. Evidently, correlation of precipitation is not as strong as that of the sea surface temperature. This implies that the target (empirical) covariance matrix for the CSIM model may be different from that for the SIM model, resulting in the distinct performances of the two models.

In this study, we introduced a novel class of copula Gaussian graphical models, called CSIM model. The model entails a tree structure between all the latent Gaussian scales and sparse in-scale conditional covariance, whereas the observed variables are non-Gaussian. This model is hence applicable to non-Gaussian data, making it useful for a wide variety of practical problems, e.g., modeling of non-Gaussian geophysical processes. Our numerical results for data sets of sea surface

V. CONCLUSION
Fig. 12. Results of different methods inferring the missing values for real datasets of Asian precipitation: (a) True graph; (b) Observed sites (white) and unobserved sites (black) in the grid; (c) Estimate by the Gaussian monoscale graphical model; (d) Estimate by the copula Gaussian monoscale graphical model; (e) Estimate by the Gaussian multiscale graphical model; (f) Estimate by the copula Gaussian multiscale graphical model (The mean absolute errors corresponding to the 4 models are 0.4686, 0.4193, 0.4742, and 0.3805 respectively).

Fig. 14. Precipitation at 3 sites as a function of time (Site 1 and Site 2 are neighbors; Site 1 and Site 256 locate at the opposite vertices on the grid): (a) Observed non-Gaussian time series; (b) Latent Gaussian time series.

temperature and Asian precipitation suggest that the CSIM model provides more accurate estimates than the Gaussian SIM model.

However, the selection of the regularization parameters in the proposed CSIM model is still a delicate issue. In future research, we will attempt to infer the regularization parameters from the data.

ACKNOWLEDGMENT

The authors would like to thank Dr. Myung Jin Choi for kindly sharing code for the SIM model.

REFERENCES