Modeling Spatiotemporal Multimodal Data with Kernel Graph Regression Models and Copulas

Jeffrey Wu* and Gareth W. Peters†
* University of California, Santa Barbara
E-mail: jeffreywu@pstat.ucsb.edu Tel: +1-352-672-5164
† University of California, Santa Barbara
E-mail: garethpeters@pstat.ucsb.edu

***Abstract*—Multimodal spatiotemporal data frequently arises in environmental monitoring, sensor networks, and public health, where multiple data sources with distinct statistical properties are collected over space and time. Traditional models such as multiple-output Gaussian Processes impose restrictive assumptions on marginal distributions and linear dependence structures, limiting their ability to capture complex real-world interactions. This work introduces a novel framework for modeling high-dimensional multimodal spatiotemporal data by combining graph signal processing, approximate Bayesian inference, and copula theory. The proposed approach decouples marginal estimation from dependency modeling, enabling scalable inference while capturing nonlinear cross-modal dependencies. Empirical results demonstrate that incorporating the copula structure improves predictive accuracy, particularly for modalities poorly modeled in isolation. Anonymized code available at: https://anonymous.4open.science/r/multimodal-KGR-2DE0**

I. INTRODUCTION

Multimodal data refers to information that combines multiple types of data sources and formats. Sometimes, modeling different types of data together results in a more complete and nuanced understanding of a complex phenomenon or network compared to modeling them separately. Building a model that can jointly represent the information from a multimodal dataset is key when there are interactions between the different modalities [1]. Most of the recent advancements in modeling multimodal data come from machine learning methods, in which the objective is to integrate data from different inputs, e.g., audio, visual, text, to give a computer or large language model more contextual information and help it complete its task [2], [3]. See [1], [4] for a survey of this area.

Another context in which multimodal data problems arise is the analysis of data from sensor networks. Sensors can simultaneously measure many different variables. This results in a dataset that exhibits not only spatiotemporal dependencies, but also potentially dependencies between the different modalities. A classic approach to account for correlations across multiple dimensions is multiple-output Gaussian Processes (GPs) [5], [6]. However, this approach assumes marginal distributions that are Gaussian and does not allow for nonlinear dependence structures, which makes it suboptimal for modeling many real world problems. As a result, alternative approaches such as Copula models have been proposed to incorporate non-Gaussian marginals and more complex, multimodal dependence structures. Copula models separate marginal distributions from the dependence structure of multivariate distributions, allowing non-linear dependence structures to be captured. See [7], [8] for a discussion of the properties of copulas and when they are useful for capturing multivariate measures of association/concordance. For example, in Zhang et al., a hierarchical model is proposed in which a multiple GP is used as a latent process, while the marginal distribution can take any form, and the dependence between marginal distributions is captured with a copula [9].

This paper builds upon previous work by proposing a novel methodology for modeling spatiotemporal multimodal data, where multimodality in this case refers to multiple types of data (modalities) observed over the same spatial units and time points. Kernel graph regression (KGR) models are used to model each spatiotemporal modality marginally because they combine a temporal kernel and a learned graph structure via a Kronecker product to estimate a flexible and parsimonious spatiotemporal dependence structure, as demonstrated in [10]. KGR models are particularly useful for modeling sensor networks and other types of spatiotemporal data where complex spatiotemporal dependencies exist but are difficult to capture with fully parametric models. To capture cross-modality dependencies, a copula is used to model the joint distribution. Copulas provide flexibility in modeling complex cross-dependencies, including tail dependencies that standard joint Gaussian models cannot capture, while maintaining the marginal distributions estimated with the KGR models.

The approach is motivated by an important real world application: modeling excess mortality due to heat in England. It is well understood that episodes of extremely hot or cold temperatures are associated with increased mortality [11]–[13]. Thus, instead of modeling mortality separately, the case study of this paper models daily excess mortality due to heat and the average maximum temperature for that day jointly across the nine regions of England during June-September of 2018.

Contributions. To facilitate the modeling of spatiotemporal multimodal data, this work proposes a novel framework for efficiently estimating a high-dimensional copula. First, KGR models are used to estimate the spatiotemporal covariance matrices and marginal posterior distributions for each spatiotemporal observation. These are then used to estimate said copula. This combination of graph regression models and copulas to fit multimodal data is novel and results in improved forecasting performance for modalities that may not be estimated well by a marginal model. This paper focuses on a two modality problem, but the proposed framework can easily be extended to more modalities.

II. BACKGROUND STATISTICAL CONCEPTS FOR MULTIMODAL DATA MODELING

This section presents important definitions of some key components used in the proposed modeling approach for multimodal data, namely GPs, linearly dependent GPs, copula models, and KGR models.

Definition 1 (Gaussian Process): A Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution. A Gaussian process is completely specified by its mean function, $\mu(x)$, and covariance function $k(x, x')$, and denoted by

$$f(x) \sim \mathcal{GP}(\mu(x), k(x, x')).$$

Definition 2 (Linearly Dependent Gaussian Processes): Given two Gaussian processes f_1 and f_2, if the correlation between $f_1(x)$, $\forall x$ in the domain of f_1, and $f_2(x')$, $\forall x'$ in the domain of f_2, is:

$$k_{f_1(x),f_2(x')} = E\left[(f_1(x) - \mu_1(x))(f_2(x') - \mu_2(x'))\right],$$

then the two random processes are said to be linearly dependent.

The dependency structure of the two dependent Gaussian processes is captured via a kernel matrix K:

$$K = \begin{bmatrix} K_1 & K_{12} \\ K_{21} & K_2 \end{bmatrix},$$

where K_1 and K_2 are the correlation matrices within process 1 and process 2, respectively, and K_{12}, K_{21} are the cross-correlation matrices capturing dependency between the processes.

Furthermore, it will be useful to define a copula distribution for a multivariate random vector as it provides a means to study dependence structures which are scale-free measures of dependence or concordance.

Definition 3 (Copula Distribution): A copula is a function $C : [0, 1]^d \to [0, 1]$ that satisfies the following properties:

- **Grounded:** $C(u_1, \ldots, u_d) = 0$ if at least one $u_j = 0$.
- **Uniform Marginals:** For all $j \in \{1, \ldots, d\}$ and $u_j \in [0, 1]$,
$$C(1, \ldots, 1, u_j, 1, \ldots, 1) = u_j,$$
where u_j appears in the j-th position.
- **d-increasing:** For all vectors $x, y \in [0, 1]^d$ such that $x_j \leq y_j$ for all j, it holds that
$$\sum_{i_1=1}^{2} \cdots \sum_{i_d=1}^{2} (-1)^{i_1+\cdots+i_d} C(u_{1i_1}, \ldots, u_{di_d}) \geq 0,$$
where $u_{j1} = x_j$, $u_{j2} = y_j$ for all $j \in \{1, \ldots, d\}$.

Let F_Y be a joint cumulative distribution function (CDF) of the random vector $Y = (Y_1, \ldots, Y_d)$ with marginal CDFs $F_{Y_1}, \ldots, F_{Y_d}$ and $U_j := F_{Y_j}(Y_j) \in [0, 1]$ for $j = 1, \ldots, d$.

Theorem 1 (Sklar's Theorem): There exists a copula $C : [0, 1]^d \to [0, 1]$ such that for all $(y_1, \ldots, y_d) \in R^d$,

$$\begin{aligned} F_Y(y_1, \ldots, y_d) &= F_Y\left(F_{Y_1}^{-1}(u_1), \ldots, F_{Y_d}^{-1}(u_d)\right) \\ &= C\left(u_1, \ldots, u_d\right). \end{aligned}$$

If all F_{Y_j} are continuous, then the copula C is unique; otherwise, it is uniquely determined on $\text{Ran}(F_{Y_1}) \times \cdots \times \text{Ran}(F_{Y_d})$.

A. Kernel graph regression models

KGR models use techniques from graph signal processing, namely the graph Laplacian, to estimate the spatial dependence structure and kernel functions to model temporal dependence. Graph Laplacians (see [10] for discussion) are calculated from a known or estimated graph structure and used to obtain a smooth spatial dependence structure over the spatial units of the graph, known as a graph filter

$$\begin{aligned} L &= V \mathbf{\Lambda} V^T \\ H^2 &= V\, \eta^2(\Lambda)\, V^T, \end{aligned} \tag{1}$$

where the graph Laplacian L and graph filter H^2 both have dimension $N \times N$, V and $\mathbf{\Lambda}$ are the eigenvectors and eigenvalues of the graph Laplacian respectively, and $\eta^2(\Lambda)$ is a decreasing function such as the cutoff function with given probability threshold $q \in [0, 1]$ used in this paper:

$$\widetilde{\lambda}_i = \eta(\lambda_i) = \begin{cases} \lambda_i, & \text{if } i/N \leq q \\ 0, & \text{otherwise.} \end{cases}$$

This graph filter is combined with a temporal kernel K, such as the locally periodic radial basis function (RBF) kernel defined below, via a Kronecker product to produce a parsimonious and interpretable spatiotemporal dependence structure

$$\begin{aligned} k_{lp}\left(x, x'; \theta\right) &:= k_p\left(x, x'; \sigma^2, \rho_p\right) k_{rbf}\left(x, x'; \sigma^2, \rho_{rbf}\right) \\ &= \sigma^2 e^{-\frac{2sin^2(\pi|x-x'|/12)}{\rho_p}} e^{-\frac{\sum_{i=1}^{d}|x-x'|^2}{2\rho_{rbf}}}. \end{aligned}$$

III. SYSTEM AND MODELS

Consider two modalities $Y_A(s, t)$ and $Y_B(s, t)$ observed at locations indexed by $s = 1, \ldots, N$ and times $t = 1, \ldots, T$, then denote $Y = (Y_A(1, 1), \ldots, Y_A(N, T), Y_B(1, 1), \ldots, Y_B(N, T)) = (Y_1, \ldots, Y_d)$ to be a multivariate random vector with CDF F_Y with marginals at (s, t) given by exponential family distributions:

$$\pi(y_i) = e^{\sum_{j=1}^{k} \eta_j(\theta) T_j(y_i) - \psi(\theta)} h(y_i).$$

Furthermore, assume modality $Y_A \in R^{N \times T}$ is Poisson distributed with a stochastic intensity process that produces a Log Gaussian Cox Process (LGCP)

$$\begin{aligned} & Y_A | \mathbf{\Lambda}, F_A \sim Poisson(\mathbf{\Lambda}), \\ & \quad \Lambda_{s,t} | F = e^{\sum_{i=1}^{N} \alpha_i I[i \in \mathcal{S}_i] + \sum_{m=6}^{9} \beta_m I[t \in m] + F_{A,(s,t)}}, \\ & \quad F_A \sim \mathcal{GP}(0, \mathbf{\Sigma_A} = K_A \otimes {H_A}^2), \end{aligned}$$

and modality $Y_B \in R^{N \times T}$ is modeled directly with a GP

$$Y_B = F_B \sim \mathcal{GP}(\mu, \mathbf{\Sigma_B} = K_B \otimes {H_B}^2) \ ,$$
$$\mu_{s,t} = \sum_{i=1}^{N} \alpha_i I \left[i \in \mathcal{S}_i \right] + \sum_{m=6}^{9} \beta_m I \left[t \text{ in month } m \right] \ ,$$

with latent intensity model parameters $\alpha \in R^N$ and $\beta \in R^m$ for m monthly fixed effects, let $\psi = \{\alpha, \beta\}$. Each process is modeled with a KGR model, which is based on the assumption that the latent process follows a GP (F_A and F_B). The dependence structure for each GP is a separable spatiotemporal covariance matrix comprised of a unique temporal kernel $K_j \in R^{T \times T}$ and graph filter $H_j \in R^{N \times N}$ for each modality. The same locally periodic RBF is used for K_A and K_B, but different hyperparameters are learned for each modality. The prior kernel hyperparameters are given generically by $\theta = \{\theta_A, \theta_B\}$ with the vector of kernel parameters given for the locally periodic RBF kernel corresponding to each modality by: $\theta_j = \{\rho_{rbf,j}, \rho_{p,j}, \sigma_j^2\}$. Separate graphs are estimated based on spatial correlations in the response Y_j to produce different graph filters H_A and H_B for each modality.

Modeling Y_A and Y_B separately ignores any potential interdependencies between these two modalities. Thus, estimating the joint distribution $\pi(Y_A, Y_B)$ in addition to the marginals can lead to improved predictions/forecasts. Since the likelihood of each modality may vary, the proposed modeling approach of this paper focuses on modeling the joint distribution of the latent processes $\pi(F_A, F_B)$ with a multivariate copula. This paper considers a d-dimensional Student's t-copula, defined as

$$C_t(u; \hat{\boldsymbol{\rho}}, \nu) := T(t_\nu^{-1}(u_1), \ldots, t_\nu^{-1}(u_d); \hat{\boldsymbol{\rho}}, \nu) \ , \tag{2}$$

where $\hat{\boldsymbol{\rho}}$ is a $d \times d$ correlation matrix, T is the joint CDF of $F \sim T_d(\nu, 0, \hat{\boldsymbol{\rho}})$, and t_ν is the standard univariate CDF of a t-distribution with ν degrees of freedom. Once again, u is obtained from the marginal CDFs of F_A and F_B. The correlation matrix is constructed as

$$\hat{\boldsymbol{\rho}} = \begin{bmatrix} \hat{\boldsymbol{\rho}}_{\boldsymbol{A}} & \hat{\boldsymbol{\rho}}_{\boldsymbol{AB}} \\ \hat{\boldsymbol{\rho}}_{\boldsymbol{BA}} & \hat{\boldsymbol{\rho}}_{\boldsymbol{B}} \end{bmatrix} \ ,$$

where $\hat{\boldsymbol{\rho}}_{\boldsymbol{A}}$ is calculated from $\hat{\mathbf{\Sigma}}_{\boldsymbol{A}}$, $\hat{\boldsymbol{\rho}}_{\boldsymbol{B}}$ is calculated from $\hat{\mathbf{\Sigma}}_{\boldsymbol{B}}$, and $\hat{\boldsymbol{\rho}}_{\boldsymbol{AB}}$ and its transpose are estimated as a diagonal cross-correlation matrix between the two modalities using the following kernel function:

$$k(y_A(s,t), y_B(s,t); \rho_{AB}) = e^{-\frac{(y_A(s,t) - y_B(s,t))^2}{2\rho_{AB}^2}} \ .$$

This diagonal correlation matrix provides a simple representation of the instantaneous cross-correlations between modalities A and B. Denser cross-correlation matrices that account for lagged spatial or temporal dependencies can be used instead, if deemed necessary, at the cost of a greater computational burden. Using (2), the joint distribution of the latent processes F_A and F_B can be written as the product of a multivariate t-distribution and two multivariate normal distributions:

$$\pi(F_A, F_B) = c(U_A, U_B; \hat{\boldsymbol{\rho}}, \hat{\nu}) \pi_A(F_A) \pi_B(F_A) =$$
$$\left[\frac{\Gamma(\frac{\nu+d}{2})}{\Gamma(\frac{\nu}{2})(\nu\pi)^{\frac{d}{2}} |\hat{\boldsymbol{\rho}}|^{\frac{1}{2}}} \right] \left(1 + \frac{1}{\nu} u^T \hat{\boldsymbol{\rho}}^{-1} u \right)^{-\frac{\nu+d}{2}}$$
$$\times \frac{1}{(2\pi)^{\frac{d}{2}} |\hat{\mathbf{\Sigma}}_A|^{\frac{1}{2}}} exp\left(-\frac{1}{2} (f_A - \mu_A)^T \hat{\mathbf{\Sigma}}_A^{-1} (f_A - \mu_A) \right)$$
$$\times \frac{1}{(2\pi)^{\frac{d}{2}} |\hat{\mathbf{\Sigma}}_B|^{\frac{1}{2}}} exp\left(-\frac{1}{2} (f_B - \mu_B)^T \hat{\mathbf{\Sigma}}_B^{-1} (f_B - \mu_B) \right) \ .$$

Thus,

$$\pi(F_A, F_B | Y; \psi, \hat{\boldsymbol{\rho}}, \hat{\nu}) \propto$$
$$\pi(Y_A | F_A; \psi_A, \hat{\mathbf{\Sigma}}_A) \pi(Y_B | F_B; \psi_B, \hat{\mathbf{\Sigma}}_B)$$
$$c(U_A, U_B; \hat{\boldsymbol{\rho}}, \hat{\nu}) \pi_A(F_A) \pi_B(F_B) \pi(\psi) \ ,$$

where U_A and U_B are each vectors of length NT defined as $U_j = F_{Y_j}(Y_j)$ for $j = 1, \ldots, d$, the likelihood is the product of two exponential family distributions (Poisson and Gaussian in this case), and the prior is the joint distribution $\pi(F_A, F_B)$ times the prior for the fixed effects of the linear predictor. Let $\psi \sim MVN(0, diag(\sigma^2))$.

IV. Estimation

The goal is to derive a low-complexity algorithm to perform multimodal forecasting at multiple locations given marginal forecasts made by KGR models fit on each modality separately. The proposed procedure follows from Inference Functions for Margins (IFM) by [14]:

1) Estimate spatiotemporal covariance matrix $\hat{\mathbf{\Sigma}} = K \otimes H^2$ for each modality. Using graphical lasso, implemented by the R package *huge* (see [15]), different graphs are estimated for each modality based on the response and subsequently transformed into graph filters via (1). The length and scale parameters for each temporal kernel are estimated with a grid search and used to calculate the gram matrix.
2) Estimate marginal posterior distribution $\pi(f_j(s,t); \hat{\psi}, \hat{\theta}, \hat{\mathbf{\Sigma}})$. The $j = 1, \ldots, NT$ marginal distributions $f_j(s,t)$ are estimated with a KGR model. To expedite the estimation of these models, the Integrated Nested Laplace Approximation (INLA), an approximate Bayesian inference method, is used. See [16] for details.
3) Use these marginal posterior distributions to get pseudo-observations $\hat{u}_j(s,t) = F_{f_j}(f_j(s,t); \hat{\psi}, \hat{\theta}, \hat{\mathbf{\Sigma}})$. Concatenate $\hat{u}_j(s,t) \forall j$ from each modality to get $u = \{\hat{u}_1, \ldots, \hat{u}_d\}$
4) Construct diagonal blocks for $\hat{\boldsymbol{\rho}}$ from $\hat{\boldsymbol{\rho}}_A$ and $\hat{\boldsymbol{\rho}}_B$. Then, define kernel function for off diagonal blocks $\hat{\boldsymbol{\rho}}_{AB}$ and $\hat{\boldsymbol{\rho}}_{BA}$. A diagonal or dense cross-correlation can be used, depending on computational resources.
5) Evaluate copulas of the form in (2) with fixed correlation matrix $\hat{\boldsymbol{\rho}}$ and a grid of ρ hyperparameters $\rho_{AB} = \{\rho_{AB}^{(1)}, \ldots, \rho_{AB}^{(g)}\}$ for $\hat{\boldsymbol{\rho}}_{\boldsymbol{AB}}$ and degrees of freedom values

$\nu = \{\nu^{(1)}, \ldots, \nu^{(g)}\}$ to find the pair that maximizes the likelihood of the copula

$$(\hat{\rho}_{AB}, \hat{\nu}) = \underset{\rho_{AB}^{(i)}, \nu^{(i)}}{\text{argmax}} \; l_C(t_\nu^{-1}(\hat{u}_1), \ldots, t_\nu^{-1}(\hat{u}_d); \boldsymbol{\rho}, \nu^{(i)}) \; .$$

Note: for high-dimensional datasets, using a copula or vine copula package may be intractable, so directly evaluating the copulas as a function is more reliable.

A. Forecasting

To forecast h steps ahead from time t, let

$$\begin{aligned}
&\pi(F_{A,t+h}, F_{B,t+h}, F_{A,1:t}, F_{B,1:t}|Y) \\
&= \pi(F_{A,t+h}, F_{B,t+h}|F_{A,1:t}, F_{B,1:t}, Y)\pi(F_{A,1:t}, F_{B,1:t}|Y) \\
&= \pi(F_{A,t+h}, F_{B,t+h}|F_{A,1:t}, F_{B,1:t})\pi(F_{A,1:t}, F_{B,1:t}|Y) \\
&= \pi(F_{A,t+h}|F_{A,1:t})\pi(F_{B,t+h}|F_{B,1:t}) \\
&\quad c(U_{A,t+h}, U_{B,t+h}; \hat{\boldsymbol{\rho}}_{t+h}, \hat{\nu})\pi(F_{A,1:t}, F_{B,1:t}|Y) \; .
\end{aligned}$$

Forecasts $Y_{A,t+h}$ and $Y_{B,t+h}$ are obtained by sampling from the posterior predictive distributions of the latent processes of the two modalities as follows:

1) Estimate $\hat{\boldsymbol{\rho}}_{t+h}$ based on a fixed length in sample window of $t+h$ (for the temporal kernels), assuming stationarity, $\hat{\nu}$ is held constant
2) Evaluate copula in (2) with $\hat{\boldsymbol{\rho}}_{t+h}, \hat{\nu}$ and then sample pseudo-observations from it $(U^*_{A,t+h}, U^*_{B,t+h}) \sim C(U_{A,t+h}, U_{B,t+h}; \hat{\boldsymbol{\rho}}_{t+h}, \hat{\nu})$.
3) Apply inverse marginal CDF to each pseudo-observation with the corresponding discretized marginal posterior distribution that INLA outputs. Say for $U_A(s, t+h)$, it provides values $\{f_A^{(1)}(s, t+h), \ldots, f_A^{(m)}(s, t+h)\}$ and probabilities $\{p_A^{(1)}(s, t+h), \ldots, p_A^{(m)}(s, t+h)\}$. Find j such that
$$P_A^{(j-1)}(s, t+h) < U_A^{(1)}(s, t+h) \leq P_A^{(j)}(s, t+h) \, ,$$
$$\text{where } P_A^{(j)}(s, t+h) = \sum_{k=1}^{j} p_A^{(k)}(s, t+h) \, ,$$
$$\Rightarrow F_A^*(s, t+h) = f_A^{(j)}(s, t+h) \; .$$
4) Repeat steps 2 and 3 to get samples of $F_A^*(s, t+h)$ and take the mean to get forecast $\hat{F}_A(s, t+h)$ $\Rightarrow$ do the same for $U^*_{B,t+h}$.
5) If necessary, like for modality A in this case, transform $\hat{F}_A(s, t+h)$ into $\hat{Y}_A(s, t+h)$.
6) If a rolling one-step-ahead forecast is being performed, like in the case study, i.e., $h = 1$, shift the in-sample window forward by 1 and repeat steps 1-5

V. Case Study

In this section, the results of a multimodal modeling problem are presented in which daily excess mortalities due to heat (Y_A) is modeled in conjunction with the corresponding average maximum air temperature for that day (Y_B) across the regions of England ($N = 9$) during the summer months (June-September) of 2018 ($T = 122$ days).

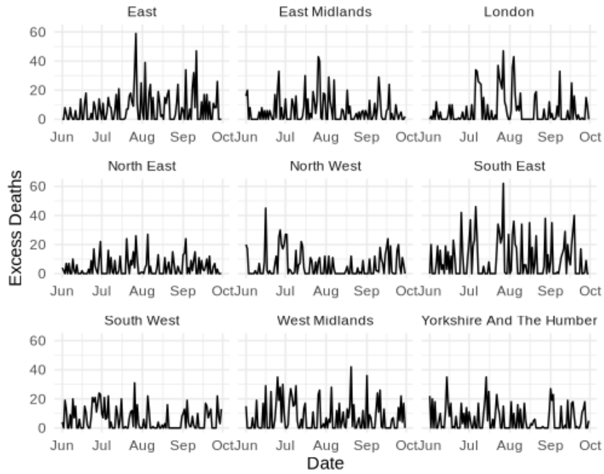

Fig. 1. Modality A: Excess deaths by region

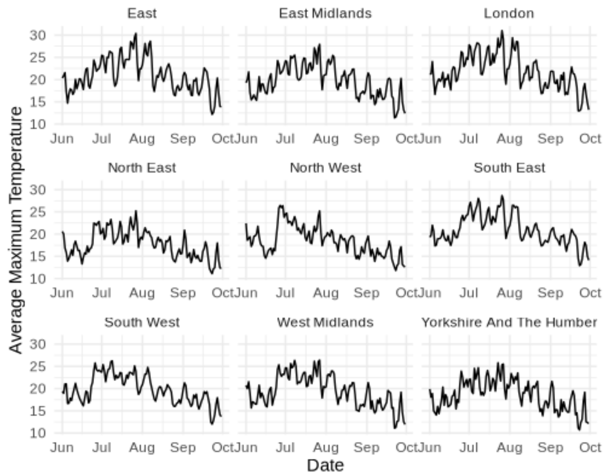

Fig. 2. Modality B: Average maximum temperature by region

The goal is to make better predictions by modeling the dependence between the two modalities with the inclusion of a copula function, compared to modeling the two modalities with separate KGR models.

A. Data

The time series for Y_A and Y_B are shown above. Data for Y_A comes from a report by the Office for National Statistics (ONS) [13]. Excess deaths are defined as the number of observed deaths compared to the five-year average. Let X_A represent this raw data, which may be negative. To make $Y_A(s, t)$ Poisson distributed, any negative values were replaced with zero i.e., $Y_A(s, t) = X_A(s, t) \wedge 0$.

Data for Y_B comes from the Met Office Integrated Data Archive System (MIDAS), which contains land surface station data starting from 1853 [17]. This collection comprises of hourly and daily weather measurements and observations of parameters relating to temperature, rainfall, and more. For each county within a given region, data from three monitoring stations were collected (if available), resulting in about ten measurements of the maximum air temperature for each day for each region. These measurements were averaged together to get a time series of daily average maximum temperatures for each region.

To evaluate the utility of modeling the interdependence between Y_A and Y_B with a copula, a rolling one-step-ahead forecasting exercise was first conducted on each modality separately. The first 115 observations for all nine regions were used for the in-sample window ($t = 115$), and then the next day's value $Y_{i,t+1}$ was forecasted for all nine regions simultaneously. This was repeated with the in-sample window sliding forward by one day until the last seven days are forecasted ($h = 1, \ldots, 7$). The predictions for each modality are then combined to estimate the copula presented in (2), which leads to better forecasts as illustrated below:

B. Results

Using the steps described in Section IV-A, more accurate forecasts were obtained, as expected, by estimating and sampling from the multivariate t-copula representing the joint distribution between the latent processes of the two modalities.

Note that in all of the plots in this section, the dotted lines represent the forecasts made marginally with the KGR models and the dashed lines represent the forecasts made with the inclusion of the t-copula. For Y_A, the KGR model severely overestimates the observations for each out-of-sample day and across all regions. The copula approach reduces mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean squared error (RMSE) by more than 50% for most forecasts as shown in Table I:

h	Marginal			Copula		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
1	14.81	2.39	15.85	8.15	1.36	8.32
2	22.81	3.65	24.37	8.34	1.39	8.49
3	23.70	3.81	25.21	8.76	1.46	8.90
4	25.59	4.10	27.03	9.27	1.54	9.42
5	35.14	5.56	37.13	9.59	1.60	9.73
6	34.59	5.50	36.59	9.88	1.65	10.04
7	42.26	6.71	44.57	9.49	1.58	9.75

TABLE I
FORECAST PERFORMANCE WITHOUT/WITH COPULA FOR Y_A

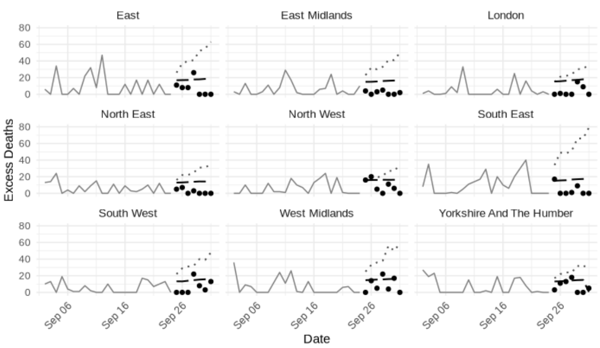

Fig. 3. Rolling one step ahead forecasts for Y_A

h	Marginal			Copula		
	MAE	MAPE	RMSE	MAE	MAPE	RMSE
1	1.976	0.158	2.478	1.965	0.157	2.448
2	1.845	0.140	2.253	1.883	0.143	2.285
3	2.059	0.119	2.422	2.016	0.117	2.377
4	3.023	0.157	3.474	3.032	0.158	3.479
5	2.026	0.145	2.542	1.963	0.140	2.444
6	2.962	0.229	3.490	2.940	0.228	3.512
7	3.513	0.278	3.851	3.535	0.280	3.886

TABLE II
FORECAST PERFORMANCE WITHOUT/WITH COPULA FOR Y_B

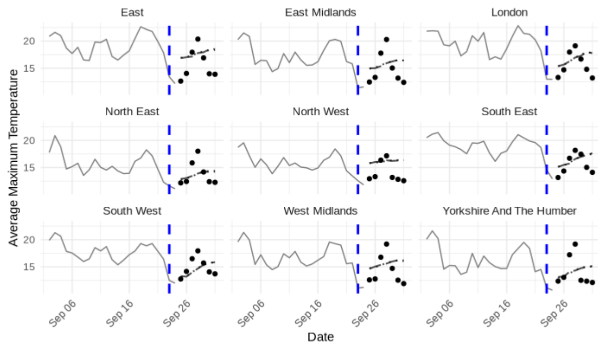

Fig. 4. Rolling one step ahead forecasts for Y_B

For Y_B, the t-copula leads to slight improvements in forecast performance. The KGR model produces forecasts that are hard to beat; hence, the addition of the copula does not make much of a difference to the predictions. In this case, the forecasted values and error metrics are virtually identical between the marginal model and the copula model.

VI. CONCLUSIONS

In this paper, a novel approach to modeling spatiotemporal multimodal data is introduced, synthesizing ideas from graph signal processing, approximate Bayesian inference, and copula models. This approach allows for the estimation of copulas even for high-dimensional data. In the case study, the inclusion of a copula function is demonstrated to improve the accuracy of forecasts (in terms of MAE, MAPE, and RMSE) over those made by marginal models, mainly for the modality that was fit poorly. This supports the notion that copulas can capture complex dependence between modalities. Potential avenues for future works include performing uncertainty quantification for these copula-based predictions, comparing with alternative multimodal modeling approaches, such as multi-output GPs and deep learning based fusion models, and exploring the theoretical properties of the model, such as identifiability and convergence guarantees.

References

[1] P. P. Liang, A. Zadeh, and L.-P. Morency, "Foundations and trends in multimodal machine learning: Principles, challenges, and open questions," no. arXiv:2209.03430, Feb. 20, 2023. DOI: 10.48550/arXiv.2209.03430. arXiv: 2209.03430[cs]. [Online]. Available: http://arxiv.org/abs/2209.03430 (visited on 04/25/2025).

[2] Y.-H. H. Tsai, S. Bai, P. P. Liang, J. Z. Kolter, L.-P. Morency, and R. Salakhutdinov, "Multimodal transformer for unaligned multimodal language sequences," no. arXiv:1906.00295, Jun. 1, 2019. DOI: 10.48550/arXiv.1906.00295. arXiv: 1906.00295[cs]. [Online]. Available: http://arxiv.org/abs/1906.00295 (visited on 04/25/2025).

[3] J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, "Multimodal deep learning," in *Proceedings of the 28th International Conference on Machine Learning, ICML 2011, Bellevue, Washington, USA, June 28 - July 2, 2011*, T. S. Lise Getoor, Ed., Omnipress, 2011, pp. 689–696. [Online]. Available: https://icml.cc/2011/papers/399%5C_icmlpaper.pdf.

[4] T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, Feb. 2019, ISSN: 1939-3539. DOI: 10.1109/TPAMI.2018.2798607. [Online]. Available: https://ieeexplore.ieee.org/document/8269806 (visited on 04/25/2025).

[5] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*. The MIT Press, Nov. 23, 2005, ISBN: 978-0-262-25683-4. DOI: 10.7551/mitpress/3206.001.0001. [Online]. Available: https://doi.org/10.7551/mitpress/3206.001.0001 (visited on 04/25/2025).

[6] M. A. Osborne, S. J. Roberts, A. Rogers, and N. R. Jennings, "Real-time information processing of environmental sensor network data using bayesian gaussian processes," *ACM Trans. Sen. Netw.*, vol. 9, no. 1, 1:1–1:32, Nov. 30, 2012, ISSN: 1550-4859. DOI: 10.1145/2379799.2379800. [Online]. Available: https://dl.acm.org/doi/10.1145/2379799.2379800 (visited on 04/25/2025).

[7] R. Nelsen, *An Introduction to Copulas* (Springer Series in Statistics). New York, NY: Springer, 2006, ISBN: 978-0-387-28659-4. DOI: 10.1007/0-387-28678-0. [Online]. Available: http://link.springer.com/10.1007/0-387-28678-0 (visited on 04/25/2025).

[8] M. D. Taylor, "Multivariate measures of concordance," *Annals of the Institute of Statistical Mathematics*, vol. 59, no. 4, pp. 789–806, Dec. 1, 2007, ISSN: 1572-9052. DOI: 10.1007/s10463-006-0076-2. [Online]. Available: https://doi.org/10.1007/s10463-006-0076-2 (visited on 04/25/2025).

[9] P. Zhang, G. W. Peters, I. Nevat, K. B. Teo, and Y. Wang, "Multimodal data fusion of non-gaussian spatial fields in sensor networks," no. arXiv:1906.03772, Jun. 10, 2019. DOI: 10.48550/arXiv.1906.03772. arXiv: 1906.03772[stat]. [Online]. Available: http://arxiv.org/abs/1906.03772 (visited on 04/25/2025).

[10] E. Antonian, G. Peters, M. J. Chantler, and H. Yan, "GLS kernel regression for network-structured data," no. 3901694, Aug. 9, 2021. DOI: 10.2139/ssrn.3901694. [Online]. Available: https://papers.ssrn.com/abstract=3901694 (visited on 04/02/2024).

[11] F. C. Curriero, K. S. Heiner, J. M. Samet, S. L. Zeger, L. Strug, and J. A. Patz, "Temperature and mortality in 11 cities of the eastern united states," *American Journal of Epidemiology*, vol. 155, no. 1, pp. 80–87, Jan. 1, 2002, ISSN: 0002-9262. DOI: 10.1093/aje/155.1.80.

[12] A. Gasparrini, Y. Guo, and M. e. a. Hashizume, "Mortality risk attributable to high and low ambient temperature: A multicountry observational study," *The Lancet*, vol. 386, no. 9991, pp. 369–375, Jul. 25, 2015, Publisher: Elsevier, ISSN: 0140-6736, 1474-547X. DOI: 10.1016/S0140-6736(14)62114-0. [Online]. Available: https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(14)62114-0/fulltext (visited on 04/25/2025).

[13] Office for National Statistics (ONS) and UK Health Security Agency (UKHSA), "Excess mortality during heat-periods: 1 june to 31 august 2022," Office for National Statistics, Statistical Article, Oct. 2022, Released 7 October 2022. [Online]. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/causesofdeath/articles/excessmortalityduringheatperiods/1juneto31august2022 (visited on 04/18/2025).

[14] H. Joe and J. J. Xu, "The estimation method of inference functions for margins for multivariate models," Oct. 1996. DOI: 10.14288/1.0225985. [Online]. Available: https://open.library.ubc.ca/soa/cIRcle/collections/facultyresearchandpublications/52383/items/1.0225985 (visited on 04/18/2025).

[15] T. Zhao, H. Liu, K. Roeder, J. Lafferty, and L. Wasserman, "The huge package for high-dimensional undirected graph estimation in r," *Journal of machine learning research: JMLR*, vol. 13, pp. 1059–1062, Apr. 2012, ISSN: 1532-4435.

[16] H. Rue, S. Martino, and N. Chopin, "Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 71, no. 2, pp. 319–392, Apr. 6, 2009, ISSN: 1467-9868. DOI: 10.1111/j.1467-9868.2008.00700.x.

[17] Met Office, *Met office midas open: Uk land surface stations data (1853–current)*, Accessed via CEDA Archive, Centre for Environmental Data Analysis, 2019. [Online]. Available: http://catalogue.ceda.ac.uk/uuid/dbd451271eb04662beade68da43546e1/ (visited on 04/18/2025).