UTS UNIVERSITY OF TECHNOLOGY SYDNEY

Deep Probabilistic Modeling for Coupled Multivariate Time Series

by Yang Yang

Thesis submitted in fulfilment of the requirements for the degree of

Master of Analytics (Research)

under the supervision of Longbing Cao

University of Technology Sydney
Faculty of Faculty of Engineering and Information Technology

April 2023

Certificate of Original Authorship

CERTIFICATE OF ORIGINAL AUTHORSHIP

I, Yang Yang, declare that this thesis is submitted in fulfilment of the requirements for the award of Master of Analytics (Research), in the School of Computer Science/FEIT at the University of Technology Sydney.

This thesis is wholly my own work unless otherwise referenced or acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

This document has not been submitted for qualifications at any other academic institution.

This research is supported by the Australian Government Research Training Program.

Signature: Production Note:
Signature removed prior to publication.

Date: 10/04/2023

Deep Probabilistic Modeling for Coupled Multivariate Time Series

A thesis submitted in partial fulfilment of the requirements for the degree of

Master of Analytics

by

Yang Yang

to

School of Computer Science
Faculty of Engineering and Information Technology
University of Technology Sydney

April 2023

ABSTRACT

Some complex intelligent systems such as for tackling the COVID-19 pandemic involve coupled Multivariate Time Series (MTSs), where both target variables (such as COVID-19 infected, confirmed, and recovered cases) and external factors (such as virus mutation and infectivity, vaccination, and government intervention influence) are coupled. Forecasting such MTSs with multiple external factors needs to model interactions within and between MTSs and handle their uncertainty, heterogeneity, and dynamics. However, existing shallow to deep MTSs modelers, including regressors, deep recurrent neural networks such as DeepAR, deep state space models, and deep factor models cannot jointly characterize these issues in a probabilistic manner across coupled MTSs. Therefore, it raises two main research problems: (1) How to conduct robust probabilistic forecasting for COVID-19 using coupled MTSs with multiple external factors? (2) how to explicitly model intra- and inter-MTS couplings and effectively handle volatile covariates of coupled MTSs?

To tackle the first problem, Chapter 3 proposes an end-to-end deep probabilistic cross-MTS learning network (MTSNet). It incorporates a tensor input consisting of scaled targeting and external MTSs. It then vertically and horizontally stacks long-short memory networks for encoding and decoding target MTSs and enhances uncertainty modeling, generalization, and forecasting robustness by residual connection, variational zoneout, and probabilistic forecasting. The tensor input is projected to a probability distribution for target MTS forecasting. MTSNet outperforms the State-of-the-Art (SOTA) deep probabilistic MTS networks in forecasting COVID-19 confirmed cases and Intensive Care Unit (ICU) patient numbers for six countries by involving virus mutation, vaccination, government interventions, and infectivity.

To tackle the second problem, Chapter 4 proposes Deep Spectral Copula Mechanisms (DSCM). Specifically, DSCM incorporates a Singular Spectral Analysis (SSA) module to reduce the volatility of multiple covariates. It applies an intra-MTS coupling module to explicitly model the temporal couplings within a single set of multivariate time series and transforms target variables into joint probability distributions via Gaussian copula transformation to establish inter-MTS couplings across multiple multivariate time series. Substantial experiments on COVID-19 time-series data from multiple countries indicate the superiority of DSCM over state-of-the-art approaches.

DEDICATION

This thesis is dedicated to the most important people in my life: my wife Shengnan and my parents, Xiaoming and Hongzhen.

To Shengnan, thank you for your unwavering support, love, and encouragement throughout my academic journey. You have been my rock and my partner in every sense of the word.

To my parents, Xiaoming and Hongzhen, who have always believed in me and encouraged me to pursue my dreams. Their unwavering support and sacrifices have been the driving force behind my academic success, and I am forever grateful for their love and guidance.

To my loyal and beloved pets, Brown and Chewy, who have been by my side through the long hours of research and writing, patiently waiting for me to take breaks and play with them.

This work is dedicated to you all with deep appreciation and love.

ACKNOWLEDGMENTS

I would like to express my deepest appreciation to my academic supervisor, Prof. Longbing Cao, for his unwavering support, guidance, and mentorship throughout my thesis journey. His expertise, dedication, and patience have been invaluable in shaping my research and improving my writing.

I am grateful for the countless hours that Prof. Cao has spent reading, reviewing, and providing feedback on my drafts, as well as for his encouragement, motivation, and wisdom. His insightful comments, constructive criticism, and timely advice have pushed me to be more thoughtful, critical, and rigorous in my research.

I am also grateful to my family and friends for their love, support, and understanding during this challenging but rewarding journey. Their encouragement and belief in me have kept me going even in the toughest times.

Thank you all for your contributions and support. This work is a product of your collective effort and dedication.

List of Publications

Submitted Papers Related to the Thesis :

1. Paper: Deep Probabilistic Cross-multivariate Time Series Modeling with External Factors for COVID-19.
   (Accepted on 03-02-2023, Conference: IJCNN 2023)

2. Paper: Deep Spectral Copula Mechanisms for Coupled and Volatile Multivariate Time Series.
   (Submitted on 22-05-2023, Conference: DSAA research track 2023)

TABLE OF CONTENTS

List of Figures

List of Tables

CHAPTER 1

INTRODUCTION

1.1 Background

1.1.1 Significance of Coupled MTSs Modeling

Time series analysis and prediction have been significant in various research topics such as signal processing, weather forecasting, COVID-19 prediction, finance, transportation, and any field involving temporal random variables [1, 2]. Based on the number of temporal observed random variables, time series analysis can be further divided into univariate and multivariate [3]. Multivariate time series (MTSs) are composed of more than one set of temporal variables, and each set of temporal variables is not only related to their historical variables but also has dependencies on each other. The coupled multivariate time series refers to a set of two or more multivariate time series that are interdependent and affect each other over time. In other words, the pattern of one multivariate time series in the set is influenced by the other multivariate time series in the same set. This coupling can occur either within the same MTSs over time, known as intra-MTS couplings, or between different MTSs, known as inter-MTS couplings. Analyzing MTS interactions and handling their uncertainty, heterogeneity, and dynamics present important challenges and make it critical for understanding complex systems and making predictions [4, 5].

Coupled MTSs are widely seen in businesses and applications and may present significant challenges to modeling multivariate time series. COVID-19 presents an

imperative challenging use case, which involves two major sets of MTSs: (1) the COVID-19 case time series, such as confirmed, death, and recovered numbers across multiple countries (2) the external factors affiliated with COVID-19 case MTSs, such as non-pharmaceutical interventions (NPIs), vaccination rates, virus mutation rates, virus infectivity (e.g., by reproduction number), mobility, and weather conditions. These sets of MTSs could be further united into different forms of MTSs, for example, MTSs of multiple states or cities in a country, or MTSs from multiple countries with their external temporal variables. This requires *cross-MTSs modeling* (also, *coupled MTSs modeling*), i.e., modeling the interactions and couplings both within heterogeneous MTSs (e.g., case MTSs, or external factor MTSs) and between multiple heterogeneous MTSs (e.g., between case and external factor MTSs). Compared to univariate time series analysis, multivariate time series analysis studies are more challenging [4, 5].

1.1.2 Motivation of Deep Probabilistic Modeling for Coupled MTSs

Time series forecasting involved either univariate or multivariate needs to develop a model that predicts the future based on the available observations. Coupled MTSs modeling involves numerous applications such as finance, economics, meteorology, and COVID-19 prediction. For example, multivariate time series models are widely used for forecasting stock prices, portfolio optimization, and risk management, macroeconomic variables such as GDP, inflation, and unemployment rates. Moreover, coupled multivariate time series models can also be used for weather forecasting, analyzing climate patterns, and COVID-19 predictions. The basic time series model is based on regression analysis, which constructs a model by considering historical time-series variables to predict the variables at future time steps. However, the majority of the multivariate time series regression models applied for the above applications were based on point forecasting, which means generating a deterministic series of values for the future period. These typical methods include the most recurrent neural network (RNN) based models, autoregressive integrated moving average (ARIMA) based models, etc [6] [7]. Compared with point forecasting, deep probabilistic forecasting for coupled MTSs aims to predict the probability distribution which can show extra quantitative information using the extracted non-linear temporal representations [8]. As the prediction of COVID-19 cases has high levels of uncertainty and complexity, deep probabilistic forecasting using multiple exogenous variables and historical data is more realistic and practical than

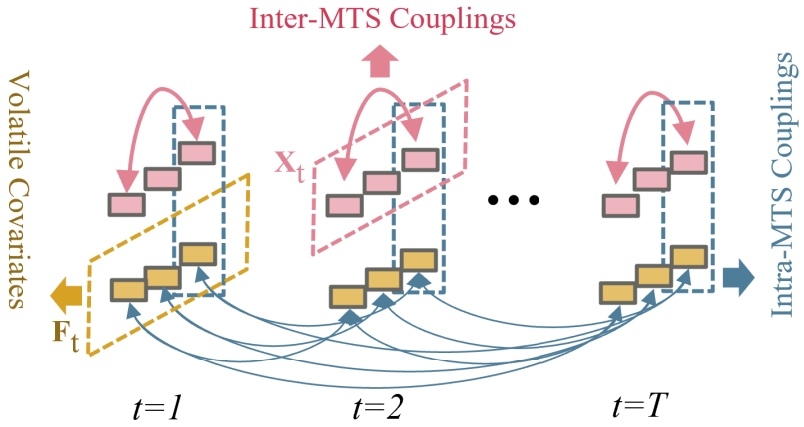


Figure 1.1: Coupled and volatile multivariate time series (CVMTS) have both intra- and inter-MTS couplings. The red dashed box denotes a target vector composed of the values of multivariate time series at time step t, and the yellow dashed box represents the volatile multiple covariates $\mathbf{F}_t$ corresponding to the target multivariate time series. The blue and red arrows indicate the intra- and inter-MTS couplings across time steps.

point forecasting for the objectives of the task. Therefore, it is necessary to explore and develop a deep probabilistic forecasting model for coupled MTSs.

1.1.3 Challenges of Couple Multivariate Time Series for COVID-19

This section discusses the challenges of modeling coupled multivariate time series for COVID-19 transmission prediction.

(1) Incorporating multiple external factors: Modeling the interactions between multiple factors and their interconnected influence on COVID-19 transmission and case movement is crucial. The interdependence among external factors can vary in strength, which must be taken into account. For example, it shows that NPIs, virus mutations, and vaccination may interact with each other and jointly influence the case numbers [9]. However, it is still unclear how multiple factors interact with each other and how they jointly affect the case movement since they may involve various implicit interactions and couplings [10]. Modeling their interactions and couplings and joint influence on case time series is difficult since external factors are highly random, diversified, separately or inconsistently enforced, influence each other differently, and distinctively affect the cases.

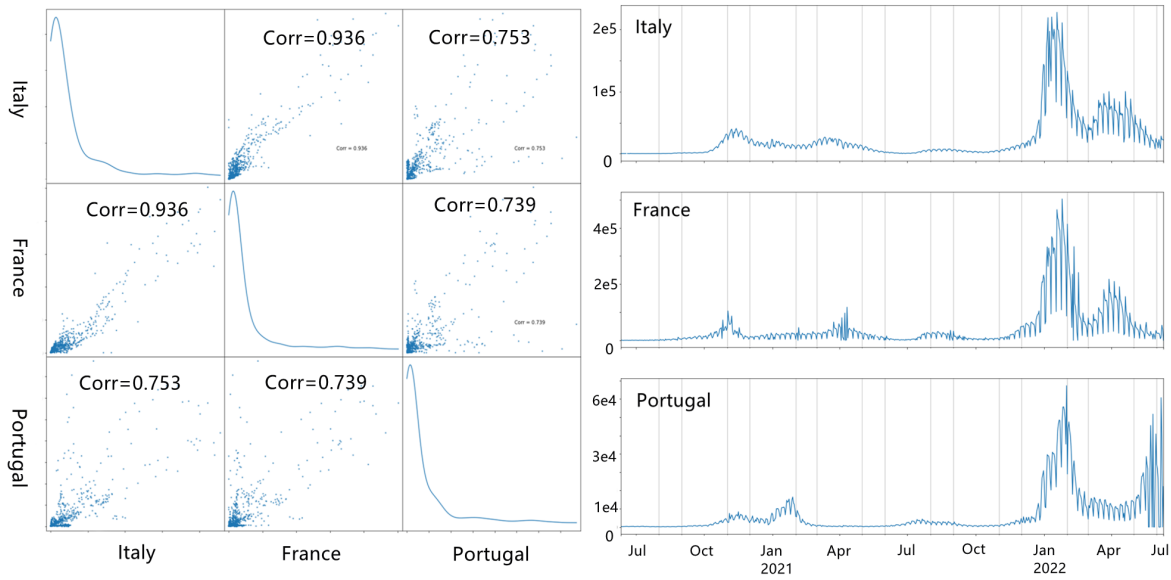


Figure 1.2: Italy, France, and Portugal have strong inter-MTS couplings in COVID-19 confirmed cases from 10/06/2020 to 10/06/2022, where the Pearson correlation coefficients reach 0.94, 0.75, and 0.74, respectively. The scatter plot indicates a positive correlation between the data from the two countries. The right plot visualizes the patterns of these three countries.

For example, intervention policies can be regarded as discretely sampled at non-equal time intervals. In addition, the timing of the COVID-19 outbreak and their corresponding interventions, and their duration periods are also random. Similar randomness also applies to the mutation of the coronavirus and the vaccination rate.

(2) Modeling uncertainty with multiple external factors: The COVID-19 case reporting, NPIs, vaccination, and virus mutation all take place at random. Their data is point-based, and stochastic, incurring high uncertainties of cases and multiple factors, and challenges in modeling such uncertainties. The existing work mainly makes point estimations to predict COVID-19 cases by estimating a deterministic series of values for future case development. Such point estimation methods include various recurrent neural networks (RNNs), autoregressive integrated moving average (ARIMA), and compartmental models in epidemiology [6, 7, 9]. They focus on data point fitting and often result in undesired results in capturing complex uncertainties. In contrast, probabilistic models estimate the probability distributions of case movements. Typical methods in probabilistic case modeling include Bayesian networks [11] and integrating SIR and statistical models [11]. Few methods involve multiple factors and model the uncertainties of cases, multiple factors, and their couplings. Although probabilistic forecasting has

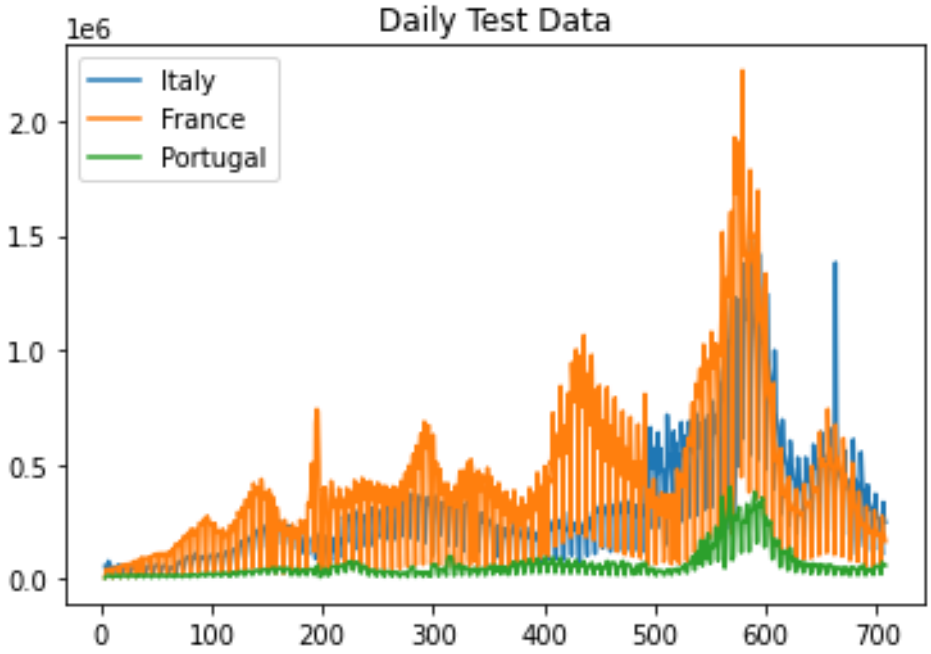


Figure 1.3: High levels of volatilities caused by missing samples in daily test data (700 days) of Italy, France, and Portugal. The Missing samples make many observations of the time series to be zero, which can affect the statistical properties of the data and the estimation of parameters. This could be due to the absence of the phenomenon or event that the random variable is measuring on the weekend.

shown its strength in capturing uncertainties, thus highly suitable for the very small and volatile COVID-19 data, it is analytically intractable to directly involve them in classic probabilistic models. It is even more challenging to model multiple factors and their coupled effect on the case movement. The work in [9] is the only one jointly modeling NPIs, virus mutation, and vaccination effect on case movement, which builds an extended SIR model.

(3) Small COVID-19 datasets lead to overfitting issues: COVID-19 data also challenges the SOTA neural MTS networks despite their significant progress. DeepAR [12], Deep State Space Model (DSM) [13], and Deep Factor (DF) [14] represent the SOTA on neural MTS forecasting. No specific deep probabilistic neural networks are available to model cross-MTSs with case MTSs and external factor MTSs across countries for COVID-19 modeling. In addition, MTSs involved in COVID-19 could be a small dataset that contains less number of samples. There are inherent limitations when using deep neural networks to fit these MTSs. As the training sets are smaller, this tends to lead to overfitting issues due to learning from fewer samples.

(4) Exploration for the sensitivity to distribution assumptions: Probabilistic models of time series are usually sensitive to the assumption of different probability distributions, for example, DeepAR [12] assumes that continuous time variables are based on Gaussian distributions while discrete-time variables apply binomial distributions. However, for a global model involving MTSs from COVID-19, there is still a need to explore and apply an appropriate probability distribution assumption to characterize the uncertainty of the temporal target variables of MTSs.

(5) Inter-MTS couplings between CVMTS: Cross-country COVID-19 time series data own complex inter-MTS couplings between multiple multivariate time series. Specifically, the countries with similar NPIs, vaccination rates, and population density likely have similar patterns of COVID-19 transmission during the outbreak. Furthermore, the peaks in the epidemics of those countries may occur at the same time. They could have consistent trends in the epidemic transmission as shown in Figure 1.2. Accordingly, at each particular time point, there are strong inter-MTS couplings between the COVID-19 data from different countries. Most of the existing deep probabilistic frameworks ignore the inter-MTS couplings. For example, DeepAR [12] can model the COVID-19 data since it can process multiple multivariate time series. However, it assumes the target variables of the COVID-19 time series from different countries are independent and follow a univariate probability distribution during the training and forecasting process. The parameters of the probability distribution are learned by a global neural network. Similarly, the Deep State Space Model [13] follows the same assumption and treats time series independently. DeepVAR [15] partially addresses this issue by implementing the Gaussian copula assumption to connect with each marginal probability distribution for multivariate time series. However, DeepVAR cannot adapt to multiple volatile multivariate time series.

(6) Intra-MTS couplings between CVMTS: For nonstationary COVID-19 case data, patterns at the previous time step far away from the current time step may have strong intra-MTS couplings with the present states. Specifically, the random variables at the current time step are coupled with the variables from multiple previous steps at different strengths, and the couplings exist in each multivariate time series of CVMTS. The existing deep probabilistic frameworks apply the RNN mechanism [16–18] to consider the variable dependence, where the current states are conditioned upon the previous states. However, these frameworks could diminish the influence of multiple previous steps on the current time step. Specifically, DeepAR [12], Deep State Space Model [13], and DeepVAR [15] merely apply the RNN mechanism to establish limited temporal

couplings and cannot consider the impacts from multiple previous steps on the current time step.

(7) Handling Volatile Covariates: COVID-19 time series data could contain volatile covariates. They are mainly caused by the complexity and inconsistency of multiple external factors, unequal sampling time intervals, and missing sampling information. Figure 1.3 visualizes the daily test numbers over 700 days in three countries as an example, where the absence of weekend or holiday information results in values close to zero, causing volatilities in those time series. Complex volatility in multidimensional covariates have a negative effect on the task of modeling the target variable. Fluctuations in the multiple covariates may add unnecessary noises and weaken the trend features, which would decline the modeling performance for target variables. The previous deep probabilistic frameworks, include DeepAR, merely normalize the covariates and fed them into RNNs [12, 13], which limits to extract the joint features from the multiple volatile covariates.

1.2 Research Problems, Objectives, and Contributions

1.2.1 Research Problems

Here, we summarize two main research problems with associated challenges in modeling coupled MTSs. These research questions are directly linked to our main topic. Firstly, accurate forecasting of COVID-19 transmission using coupled MTSs with multiple external factors is a crucial problem to solve. It involves designing a trainable deep probabilistic architecture to jointly process the coupled MTSs with multiple external factors that can enhance the accuracy and performance of forecasting models for various applications. Secondly, explicitly modeling intra- and inter-MTS couplings and effectively handling volatile covariates of coupled MTSs is also a significant problem related to our main topic. Addressing these research questions can provide meaningful insights into the development of deep probabilistic models for coupled MTSs that can have significant implications in various areas.

Research Problem (1): How to conduct robust probabilistic forecasting using coupled MTSs with multiple external factors (COVID-19 data) ? COVID-19 data with multiple external factors including virus mutations, vaccination, government interventions, and infectivity constitute coupled MTSs. The specific challenges are sum-

marized as the following:

- How to incorporate multiple external factors for COVID-19 prediction.
- How to design a trainable deep probabilistic architecture to jointly process the coupled MTSs with multiple external factors?
- How to tackle the overfitting and gradient vanishing issues caused by limited MTSs dataset?
- How to model the diverse uncertainties of MTSs and perform probabilistic forecasting?

Research Problem (2): How to explicitly model intra- and inter-MTS couplings and effectively handle volatile covariates of coupled MTSs? The following summarizes the challenges associated with this research problem:

- How to effectively mitigate the interference caused by volatile covariates and improve the robustness of probabilistic prediction of coupled and volatile multivariate time series (CVMTS)?
- How to design a trainable deep probabilistic architecture to explicitly model the intra-MTS couplings of the CVMTS?
- How to establish the inter-MTS correlations of CVMTS?

1.2.2 Objectives

The two primary objectives of the thesis are intended to address the two main research questions that have been introduced in the previous section.

The first research question motivates us to explore a deep probabilistic framework using coupled MTSs for COVID-19 prediction. It aims to propose an MTSs-based global model for probabilistic forecasting of COVID-19 daily confirmed cases and the number of ICU patients from multiple countries with exogenous variables. And this approach is flexible and compatible with different assumed probability distributions for robust prediction. Moreover, the proposed method can not only outperform the SOTA approaches but also demonstrate the effectiveness of incorporating the exogenous variables.

For the second research question, we aim to propose a deep probabilistic mechanism to improve the robustness of the probabilistic prediction for CVMTS and explicitly capture the inter- and intra-MTS couplings of CVMTS. Moreover, the proposed approach can

not only outperform the SOTA approaches but also demonstrate the effectiveness of the designed module in probabilistic forecasting for CVMTS.

1.2.3 Contributions

This paper presents two major contributions that relate to the proposed research questions about deep probabilistic modeling for coupled multivariate time series.

To tackle the first research problem, we propose an end-to-end deep probabilistic cross-MTS network (MTSNet), designed to utilize the historical time series data and multiple exogenous variables for probabilistic forecasting. The main contribution of MTSNet is addressing the four challenges (1) to (4) in section 1.1.3. Specifically, we jointly model multiple exogenous variables by concatenating them to a normalized second-order tensor. We design a global model capable of processing multiple time series from different countries via vertically and horizontally stacked recurrent neural networks to map hidden states to the assumed probability distribution. To improve the accuracy and robustness of the probabilistic forecasting, we incorporate the Residual Connection and Variational Dropout into the MTSnet. To examine the diverse uncertainties caused by various statistical assumptions, we test different probability distributions including Laplace distribution, Gaussian distribution, and Student-T distribution. Moreover, we conduct massive experiments on the data from six countries (UK, US, Canada, France, Spain, and Italy). The results not only indicate that the proposed method can outperform the state-of-art models but also compares the difference between various distribution assumptions. The ablation experiments show the effectiveness of incorporating the exogenous variables and Variational Dropout.

For the second research problem, we introduce the deep spectral copula mechanisms (DSCM), which incorporates the Singular Spectral Analysis (SSA) module to reduce the volatilities of multiple covariates and use an intra-MTS Coupling module to explicitly model the temporal correlation within a single set of multivariate time series. The main contribution of DSCM is to tackle the challenges (5) to (7) in section 1.1.3. Particularly, to establish inter-MTS correlations across multiple multivariate time series, we transform the target variables into joint probability distributions via Gaussian copula transformation. The contributions of DSCM are summarized as the following: To reduce the volatility of covariates in CVMTS and enhance the modeling robustness, DSCM incorporates a singular spectral analysis (SSA) module to decompose and reconstruct multiple covariates. DSCM applies an intra-MTS coupling module to capture the temporal couplings between hidden representations generated by a deep neural network within

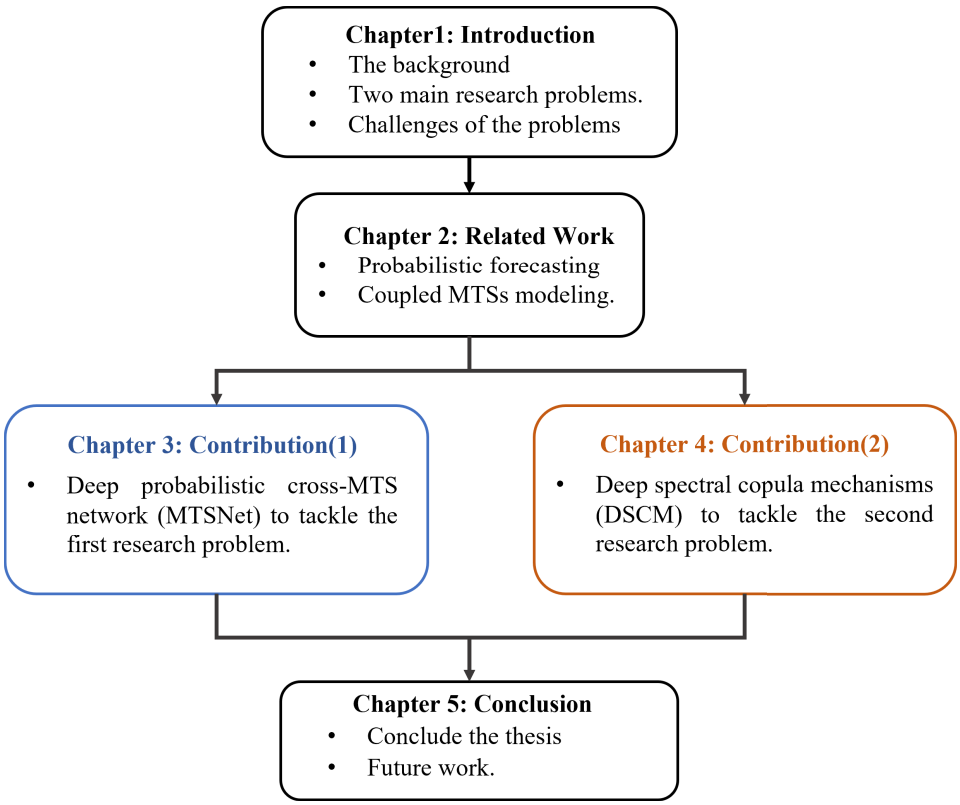


Figure 1.4: The diagram provides an overall structure of the thesis and a brief explanation of each chapter with its relevance to each other.

multivariate time series. To capture the inter-MTS couplings across the coupled multivariate time series, DSCM utilizes a Gaussian copula function to connect the various marginal probability distributions of temporal target variables. The massive experiments on COVID-19 time-series data from multiple countries indicate the superiority of the proposed framework over the deep probabilistic SOTA approaches in modeling CVMTS.

1.2.4 Thesis Structure

As shown in Figure1.4, the thesis comprises five chapters, which are connected and serve a distinct purpose in advancing the research. Chapter 1 provides an introduction to the research by presenting the background and context of the study, outlining the research objectives, and identifying the two research problems that the study aims to address. Additionally, this chapter highlights the challenges that need to be overcome to successfully address these research problems in section 1.1.3. Chapters 3 and 4 respectively describe the two primary contributions in tackling the proposed research problems. Those

chapters provide a detailed description of the proposed algorithms, experimental design, data sources, and result analysis to address the challenges associated with our research problems. Finally, Chapter 5 presents a comprehensive summary of the main findings of the study, compares the results with the existing literature, and discusses the implications of the findings for future research. This chapter concludes with recommendations for future work.

CHAPTER 2

RELATED WORK AND RESEARCH GAPS

2.1 Work Related to COVID-19 Forecasting

Since the primary research problems refer to the deep probabilistic modeling for coupled multivariate time series using COVID-19 data, we review the work related to COVID-19 forecasting, COVID-19 factor modeling, and MTS probabilistic modeling, respectively.

First, COVID-19 forecasting has been intensively studied. They involve topics and tasks including predicting case movement, transmission, infection, and impact on hospitalization and mental health [44]. Most of the methods are built on regression models without exogenous variables [19–22, 25–27, 29–33, 35–38, 41, 42]. Table 2.1 in Appendix summarizes the MTS modeling for COVID-19 forecasting and indicates this research gap. On the other hand, most studies focus on point-based fitting and trend forecasting, including regression, SIR compartmental models, and DNNs. They thus suffer from the highly uncertain, noisy, and fluctuating observations in COVID-19 data [45].

Second, COVID-19 transmission is influenced by multiple external factors [44]. They include NPIs such as travel restrictions, social distancing, and school closures, and other factors such as healthcare interventions, vaccination effect, and economic response. Existing work includes modeling the NPI effect such as government interventions by statistical methods [46, 47], NPIs by multivariate modeling [48], mobility [49, 50], and other timed interventions [51]. These studies illustrate the necessity of examining the influence of exogenous variables on COVID-19 transmission. However, most methods involve none to up to two factors such as weather or mobility on case movement. Few

models take into account multiple external factors. For Example, In [23], CNNs extract the features about diabetes, smoking, and population information and then feed them to LSTM networks. In [24], long-short memory neural networks (LSTM) serve as an encoder of multivariate time series of temperature, rainfall value, population, and locations. In [51], a heterogeneous autoregression model (HAR) based on multiple linear regressions predicts COVID-19 confirmed cases with growth rate and vaccination rate. In [34], ordinary least squares analyze the influence of travel history and contacts. A graph neural network [43] incorporates the factor of mobility into modeling. In [9], a new compartmental model integrates multiple factors including NPIs, vaccination, and mutation to model movements of infected, confirmed, and recovered cases, which shows significant modeling challenges in MTS forecasting with multifactor context.

Third, typically, MTS forecasting is made by regressors, probabilistic models, and DNNs. Probabilistic time-series forecasting estimates the probability distribution of underlying observations over time by involving certain density hypotheses. The parameters of density distributions are optimized to fit the entire samples and capture their uncertainty and development. Bayesian statistics are commonly used in COVID-19 case estimation [11]. A recent focus is on integrating RNNs with temporal probability distributions to characterize case uncertainties as well as temporal dependencies. For example, DeepAR [12] does probabilistic MTS forecasting using two layers of RNNs to extract features and then map the hidden states of the RNN cells to the parameters of an assumed distribution. In [13], DSM combines a linear state space model with RNNs. By introducing Gaussian noise, the marginal likelihood function of a linear state space model forms an analytically tractable loss function to train the neural networks. The joint probability distribution is obtained after generating the parameters of the state space model by the RNNs. In addition, DF [14] uses a global time series to represent the latent representation of each time series and is applied to the state space model and Gaussian process. However, so far no studies are exploring the SOTA deep probabilistic methods in predicting the COVID-19 epidemic with multiple exogenous variables.

Lastly, two pieces of work are mostly relevant to the methodology proposed in chapter 3. The work in [9] involves various factors but in a differential compartmental framework. The second one is DeepAR since our network also uses RNNs as the encoder and maps the hidden states of RNN cells to the parameters of the probability distribution function through fully connected layers. The difference lies in (1) DeepAR has not been applied to COVID-19, while ours is developed for this application; (2) we jointly model multiple exogenous variables by concatenating them to form the covariates; (3) we incorporate a

variational zoneout layer and residual connection into a stacked architecture to make the model fit the COVID-19 uncertainty better. Our model thus shows better capacity in characterizing four external factors including virus mutation, vaccination, government interventions, and virus infectivity and probabilistically forecasting COVID-19 confirmed cases and the number of ICU patients.

2.2 Work Related to MTSs Modeling and our study

Modeling coupled and volatile multivariate time series (CVMTSs) is challenging because of not only the complicated inter- and intra-MTS couplings inherent in the CVMTS but also the volatile covariates affecting the model performance. Here, we review the related work about multivariate time series modeling, deep probabilistic time-series forecasting, and COVID-19 case series forecasting. Then, we examine the work relevant to our study and discuss their research gaps.

First, Multivariate time series analysis is a discipline that examines multiple time-related variables. The objective is to comprehend the connections between these variables and make predictions about future values. It is used in various areas such as financial forecasting, climate research, and healthcare data examination. The approaches used in multivariate time series analysis include vector autoregression (VAR) [52], vector error correction models (VECM) [53], and dynamic factor models (DFM) [54]. One of the challenges of this field is the sheer amount of data generated and the need for specialized software and techniques to handle it. Another challenge is dealing with non-stationary data which can have an impact on the accuracy of predictions. To address this, researchers use pre-processing techniques such as detrending or normalization to stabilize the data. However, all of these traditional approaches could not capture the nonlinear representations and correlations of multivariate time series.

Second, deep probabilistic time-series forecasting combines statistical methods with deep neural networks. A typical framework is DeepAR [12]. It assumes target variables follow a univariate probabilistic distribution at each time point and utilize RNNs to learn the assumed distribution. In contrast to the traditional point estimation of time series, deep probabilistic models aim to learn the probabilistic distribution of the target variables by incorporating the maximum likelihood estimation into backpropagation algorithms. However, DeepAR processes multiple time series independently in a global framework and ignores the inter-couplings across target variables. Similarly, the Deep State Space model [13] combines a linear state space model with deep neural

networks. The independent temporal probability distribution is obtained by optimizing the parameters of the state space model in recurrent neural networks. Those frameworks cannot capture the dependencies between temporal sequential observations. Furthermore, DeepVAR [15] combines an RNN with a Gaussian copula process by incorporating the low-rank covariance structure. This method uses the Copula function to model the dependencies between target variables from different distributions. However, it ignores the intra-couplings within a single multivariate time series and does not incorporate multiple time series covariates.

Third, intensive research has been done on modeling COVID-19 time series [44]. Most of the sequential models are based on regression using historical data [19–22, 25–27, 29–33, 35–38, 41, 42]. Those model apply point estimation without considering probabilistic forecasting. Furthermore, none of them models the CVMTS scenario by considering the dependencies of multiple covariates on target variables across different countries [45].

Lastly, there are two papers most relevant to our study. From the application perspective, the work in [9] involves multiple sequential COVID-19 data but in a differential compartmental framework. From the modeling perspective, DeepVAR [15] is close to our paper since we also utilize a similar deep Gaussian copula process to handle the dependencies between the time-series data from various distributions. Our work differs from DeepVAR significantly: (1) We use the SSA module to decompose and reconstruct multiple covariates in CVMTS to reduce the volatility and enhance the robustness of the framework. (2) We design an intra-MTS coupling module to capture the temporal correlation of the hidden representations generated by the deep neural network within the correlated multivariate time series. (3) We add statistical features to the time series embedding to improve the learning ability of neural networks for a certain target time series of CVMTS.

In Chapter 4, we introduce deep spectral copula mechanisms (DSCM) to conduct probabilistic forecasting of coupled and volatile multivariate time series. DSCM consists of the SSA module, the intra-MTS coupling module, and the inter-MTS coupling module. Furthermore, it also conducts massive comparison experiments and ablation studies to demonstrate its superiority. Then, Chapter 4 concludes the work and discusses its limitations.

2.3 Research Gaps in Coupled Multivariate Time Series for COVID-19

For COVID-19, with the uncertain spread of COVID-19 worldwide, accurately predicting their transmissions is critical for informed intervention and healthcare policymaking. However, COVID-19 modeling [44] is challenging as it involves significant uncertainties and the influence of multiple external factors such as NPIs, vaccination, and virus mutation on the virus transmission and case series movement [9, 44]. So far, thousands of references have been reported on engaging epidemiological, statistical, and shallow to deep learning methods in modeling COVID-19 case movement [44, 45]. Nevertheless, most of them focus on point-based case data only, and very limited work involves comprehensive external factors [47, 49]. Cross-MTSs modeling raises three important yet challenging perspectives for COVID-19 modeling. One is to couple multifactors with case MTSs and then jointly model their interactions and coupled effect on case movement. The second is to extend the point-based case fitting to process-based modeling of COVID-19 cases with multifactor context for better uncertainty modeling of often small, volatile case data. The third is to jointly model case movement across multiple countries, i.e., developing a global model for cross-country MTS modeling.

On the other hand, coupled and volatile multivariate time series (CVMTS) is one type of multivariate time series, which contains multiple coupled multivariate time series with volatile covariates at the same length. Specifically, CVMTS is characterized by intricate inter-MTS couplings that exist between multiple multivariate time series. These inter-MTS couplings entail complex relationships and dependencies among different sets of multivariate time series data. Moreover, CVMTS encompasses the analysis of intra-MTS couplings, which captures the interconnections and interdependencies within each individual multivariate time series. These intra-MTS couplings represent the internal associations and dependencies among the variables within a specific multivariate time series. Each set of multivariate time series is composed of target time-series variables and volatile covariates. The components of CVMTS are illustrated in Figure 1.1. Accordingly, the challenges of CVMTS include exploring inter- and intra-MTS couplings and dealing with volatile covariates. A typical example of CVMTS is the COVID-19 case time series across multiple countries with volatile external factors. Its interested target variables may include confirmed cases, deaths, and hospital patients. The corresponding covariates may include daily test number, stringency index, the share of virus variants, vaccination rate, non-pharmaceutical interventions (NPIs), and virus infectivity. How-

ever, existing state-of-the-art approaches either model each set of multivariate time series independently (such as DeepAR, Deep State Space Model, etc.) or process the multivariate time series without considering their volatile covariates (e.g., DeepVAR, GPVAR). It is significant to explore a deep multivariate time series modeling approach for CVMTS to explicitly intra-and inter-MTS correlations.

As the characteristics of CVMTS are widely present in the COVID-19 datasets. it is urgent to explore a robust model to effectively alleviate the issues in the probabilistic prediction of CVMTS. In recent years, numerous epidemiological, statistical, and deep learning approaches have been applied for COVID-19 prediction tasks [44, 45, 47, 49, 55, 56] and examine the importance of considering the covariates in modeling the COVID-19 transmission [48, 50, 51]. However, most of these studies are based on point estimation and cannot generate a probability distribution to depict the uncertainty in predicting the future.

Though some deep probabilistic frameworks [12–15] conduct probabilistic forecasting by combining deep neural networks with statistical approaches, these deep probabilistic frameworks cannot explicitly model the inter-MTS couplings between and intra-MTS couplings within multiple countries with similar epidemiological trends. Furthermore, they cannot effectively process volatile covariates. And in COVID-19 data, incorporating the covariates of volatile time series may affect the robustness and accuracy of the model. It is significant to explore an approach to tackle issues of incorporating the volatile time series into deep multivariate time series modeling. The above reasons motivate us to conduct this research.

Approaches	**Time Series Type**	**Prediction Type**	**Involved Factors**
RNN and LSTM [19]	Univariate	Point Forecasting	Historical Cases
Stacked LSTM Networks [20]	Univariate	Point Forecasting	Historical Cases
ANN, LSTM and RNN [21]	Univariate	Point Forecasting	Historical Cases
LSTM-based models [22]	Univariate	Point Forecasting	Historical Cases
Hybrid CNN-LSTM model [23]	Multivariate	Point Forecasting	Historical Cases; Diabetic patients; Smokers; Gender; Age
Multivariate stacked LSTM [24]	Multivariate	Point Forecasting	Historical Cases; Temperature; Rainfall Values; Population; Area
Hybrid deep learning models, CNN, LSTM [25]	Univariate	Point Forecasting	Historical Cases
LSTM, GRU and Bi-LSTM [26]	Univariate	Point Forecasting	Historical Cases
LSTM networks [27]	Univariate	Point Forecasting	Historical Cases
LSTM model[28]	Univariate	Point Forecasting	Historical Cases
LSTM networks [29]	Univariate	Point Forecasting	Historical Cases
ARIMA, NARNN and LSTM [30]	Univariate	Point Forecasting	Historical Cases
LSTM networks [31]	Univariate	Point Forecasting	Historical Cases
[32]	Univariate	Point Forecasting	Historical Cases
ARIMA models [33]	Univariate	Point Forecasting	Historical Cases
linear regression models [34]	Multivariate	Point Forecasting	Historical Cases; Travelling History; Contacts
SVR and Stacking-ensemble learning, ARIMA, CUBIST, RIDGE, and RF models [35]	Univariate	Point Forecasting	Historical Cases
LSTM models [36]	Univariate	Point Forecasting	Historical Cases
Hybrid ML and SI method [37]	Univariate	Point Forecasting	Historical Cases
Non-linear growth models (Gompertz, Verhulst) and exponential model (SIR) [38]	Univariate	Point Forecasting	Historical Cases
Heterogeneous autoregression models (HAR) [39]	Multivariate	Point Forecasting	Historical Cases; Growth Rates; Vaccination
Bayesian LSTM[40]	Multivariate	Point Forecasting	Historical Cases; Population; Mobility; Location
Weighted combining methods [41]	Univariate	Probabilistic Forecasting	Historical Cases
Statistical Analysis [42]	Univariate	Probabilistic Forecasting	Historical Cases
Auto-regressive Mixed Density Dynamic Diffusion Network (ARM3Dnet) [43]	Multivariate	Probabilistic Forecasting	Historical Cases; Mobility; Location
Proposed Model (MTSNet)	Multiple Multivariate	**Probabilistic Forecasting**	**Historical Cases; Virus Mutation; Vaccination; Government Interventions;Virus Infectivity)**

Table 2.1: Related Works about MTS Modeling for COVID-19 forecasting: The majority of research papers focus on point-based forecasting instead of probabilistic forecasting and most deep models for COVID-19 prediction involved limited historical factors.

CHAPTER 3

CONTRIBUTION- PROBABILISTIC FORECASTING USING COUPLED MTSS WITH MULTIPLE EXTERNAL FACTORS

3.1 Deep Probabilistic Cross-MTS Network

3.1.1 Introduction

Given the examination of background and research questions related to COVID-19 modeling in the previous chapters, it is evident that COVID-19 forecasting is a complex challenge due to high levels of uncertainty caused by various external factors. Although many algorithms have been proposed to predict confirmed cases, most previous studies focused on point forecasting rather than probabilistic forecasting. Moreover, none of these studies have investigated the probabilistic prediction using coupled MTSs with multiple external factors simultaneously, including virus mutations, vaccination, government interventions, and infectivity.

To tackle the first research questions, this chapter proposes an end-to-end model, namely MTSNet, designed to utilize the above four types of exogenous variables and historical time-series data for probabilistic forecasting. This approach is a global model capable of processing multiple time series from different countries via stacked recurrent neural networks to learn a normalized and concatenated tensor and map hidden states to an assumed probability distribution. In addition, Residual Connection and Variational Dropout are implemented in the architecture to improve the accuracy and robustness. Massive experiments on the data from six countries (UK, US, Canada, France, Spain, and

Italy) indicate that the proposed method can outperform the state-of-art models for probabilistic forecasting. The ablation experiments show the effectiveness of incorporating the exogenous variables.

3.1.2 The MTSNet Architecture

This section introduces the architecture of MTSNet for modeling coupled MTSs with multiple external MTS factors. MTSNet stacks an encoder-decoder sequence-to-sequence learning framework, as shown in Fig. 3.1. In a single sequence-to-sequence network [57], an encoder aims to map the input sequence to their associated hidden states which will then be fed into a decoder to generate the output sequence. MTSNet consists of an encoder network that learns the interactions and representations of target and external MTSs over history (i.e., $t = 1$ to $t = t_0 - 1$) and then learns the embedding of the last time step ($t = t_0$). The decoder network incorporates the external MTSs and sampled target of the last time to decode and produce the embedding of every next time step (t) progressively to forecast the target variables over the next time steps (i.e., $t = t_0 + 1$ to $t = t_0 + m$).

MTSNet has residual connections and variational zoneout to capture sequential time-series dependencies and cross-MTS couplings both within and between target MTSs and multifactor MTSs and the MTS uncertainty. In addition, MTSNet builds upon the DeepAR model [12] to learn probabilistic projection mechanisms and encoding process. Differing from the traditional sequence-to-sequence networks and DeepAR, MTSNet has a vertically and horizontally stacked LSTM architecture with multiple layers for residual connection and variational zoneout. Its second layer applies residual connection to modify its output for robust forecasting and preventing gradient vanishing. The variational zoneout is implemented in the third layer to extend modeling uncertainty and alleviate overfitting. The projection layer of MTSNet projects the hidden states to the assumed probability distribution.

MTSNet has stacked LSTM layers. To avoid repeating their architectures and module formulation, we assume that readers are familiar with LSTM. Otherwise, interested readers can refer to the basic feed-forward neural networks [16], simple RNN [17], and LSTM [18] for more information. DNNs such as RNNs hold nonlinear and deterministic characteristics to fit time-series data [58]. LSTM further alleviates the gradient vanishing and exploding problems during its optimization process.

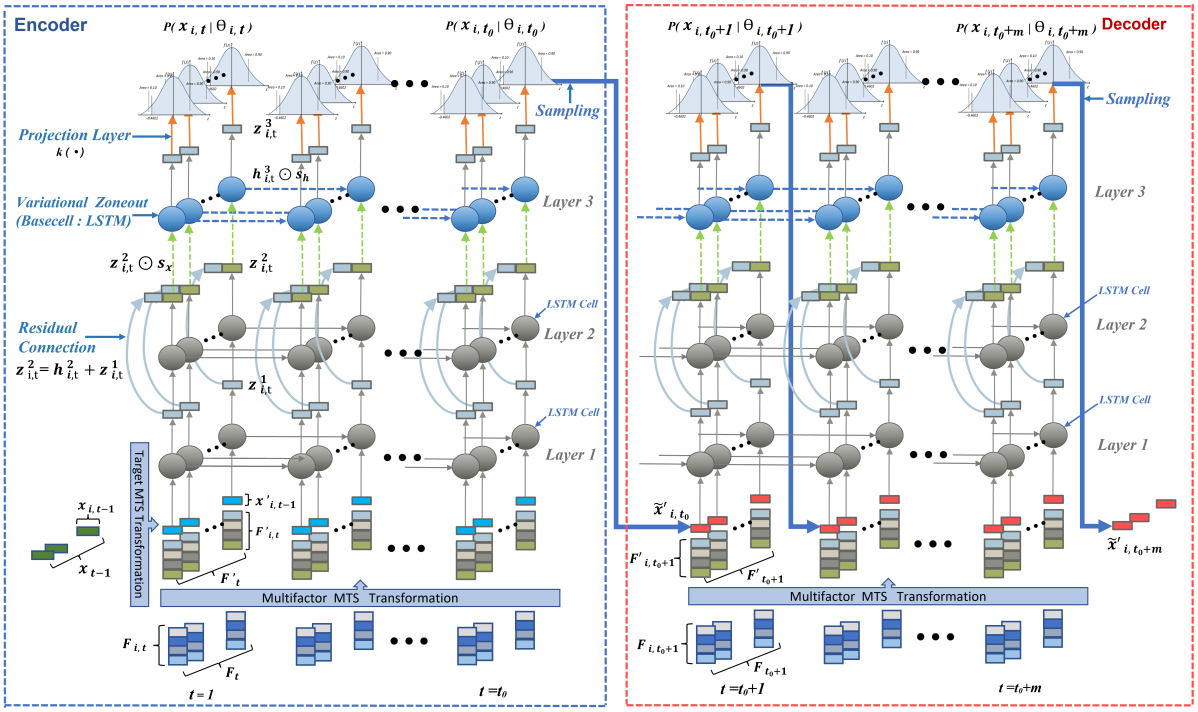


Figure 3.1: Deep probabilistic cross-MTS network with vertically and horizontally stacked LSTMs, residual connection, variational zoneout, and probabilistic forecasting over the target and multifactor MTSs. During the training process, both target observations $x_{i,t}$ and external factors $\mathbb{F}_{i,t}$ $(1 \le t \le t_0+m)$ are available at each time step to train the encoder and decoder. At the prediction stage, the target value $x_{i,t}$ $(t_0 < t \le t_0+m)$ for the decoder will be sampled from the predicted distribution of the previous time step (red dashed box).

3.1.3 Scaled Multifactor and Target MTS Tensor

MTSs for target input and external factors are often diversified with heterogeneous value ranges and follow different distributions. To fit them to a global network well and neutralize their contributions in the neural transformations, an appropriate transformation could improve the computational efficiency and accuracy of prediction results [59]. Here, we apply the min-max scaling to normalize all time-series data into the range of [0, 1]. $\mathbb{F}_{i,j,t} \in \mathbb{R}$ is a single value of factor j (e.g., reproduction rate) at time step t for a source i (e.g., a country). Let vector $\mathbb{F}_{i,j}$ denote a time series from source i for factor j and of length t, then the min-max normalization is given in Eq. (3.1):

$$\mathbb{F}'_{i,j,t} = \frac{\mathbb{F}_{i,j,t} - \min(\mathbb{F}_{i,j})}{\max(\mathbb{F}_{i,j}) - \min(\mathbb{F}_{i,j})} \tag{3.1}$$

We obtain the normalized value $\mathbb{F}'_{i,j,t}$ for each input and then combine these values into a vector $\mathbb{F}'_{i,t}$, which represents the integration of all factors j at time t from source i.

$$\begin{aligned} \mathbb{F}'_{i,t} &= [\mathbb{F}'_{i,1,t}, \mathbb{F}'_{i,2,t}, \ldots, \mathbb{F}'_{i,j,t}]^T \\ \mathbb{F}'_{t} &= [\mathbb{F}'_{1,t}, \mathbb{F}'_{2,t}, \ldots, \mathbb{F}'_{i,t}]^T \end{aligned} \tag{3.2}$$

To accelerate the convergence of gradient descent and improve the computational efficiency, we also use the Min-Max scaling to process all target time series [59]. Another reason is that Min-Max scaling can alleviate the issue caused by the heterogeneous value ranges between MTSs.

$$\begin{aligned} x'_{i,t} &= \frac{x_{i,t} - \min(\mathbf{x_i})}{\max(\mathbf{x_i}) - \min(\mathbf{x_i})} \\ \mathbf{x}'_t &= [x'_{1,t}, x'_{2,t}, \ldots, x'_{i,t}]^T \end{aligned} \tag{3.3}$$

where $x'_{i,t}$ and $x_{i,t}$ represent the normalized value and original value of the target time series. Let vector $\mathbf{x}_i$ denote a time series from source i, then it uses the normalized target value $x'_{i,t}$ to form a new vector $\mathbf{x}'_t$.

To model the joint effect of multiple external factors $\mathbb{F}'_{i,t}$ on the underlying time series $\mathbf{x}'_t$, we concatenate the vectors $\mathbb{F}'_{i,t}$ of all sources to a second-order tensor $\mathbb{F}'_t$ as shown in Fig. 3.1. $\mathbb{F}'_t$ is the global representation of temporal exogenous variables with the information from all sources, which is fed to the MTSNet network as input. The interactions between the concatenated multifactors $\mathbb{F}'_t$ and between the multifactors and their underlying time series $\mathbf{x}'_t$ are captured by multi-layer nonlinear neural transformations in MTSNet. The hidden representations are further characterized by the parameters of prior probability distribution in probabilistic forecasting.

3.1.4 Residual Connection

To enhance the stability of nonlinear transformations in MTSNet, we incorporate residual connection [60] into the second layer of MTSNet. As shown in Fig. 3.1, the updated output of the LSTM cells in layer 2 is equal to the input element-wisely added to the original output of this cell through residual connection. Let LSTM_2 and LSTM_3 denote the neural operations of layers 2 and 3 in Fig. 3.1, $\mathbf{h}^2_{i,t}$, $\mathbf{c}^2_{i,t}$ and $\mathbf{z}^1_{i,t}$ represent the hidden states, memory states, and input of LSTM cells in layer 2, respectively. The residual connection

applied in MTSNet is formulated below in Eq. (3.4).

$$
\begin{aligned}
\mathbf{h}^2_{i,t}, \mathbf{c}^2_{i,t} &= \mathrm{LSTM}_2\left(\mathbf{h}^2_{i,t-1}, \mathbf{c}^2_{i,t-1}, \mathbf{z}^1_{i,t}; \mathbf{W}^2\right) \\
\mathbf{z}^2_{i,t} &= \mathbf{h}^2_{i,t} + \mathbf{z}^1_{i,t} \\
\mathbf{h}^3_{i,t}, \mathbf{c}^3_{i,t} &= \mathrm{LSTM}_3\left(\mathbf{h}^3_{i,t-1}, \mathbf{c}^3_{i,t-1}, \mathbf{z}^2_{i,t}; \mathbf{W}^3\right)
\end{aligned} \tag{3.4}
$$

$\mathbf{z}^2_{i,t}$ is the input to layer 3, equalling $\mathbf{z}^1_{i,t}$ element-wisely added to $\mathbf{h}^2_{i,t}$. According to [60], the residual connection can lift neural learning robustness and mitigate gradient vanishing.

3.1.5 Variational Zoneout

To tackle the challenges of small data on network overfitting and enhance network generalization, zoneout techniques are used as a regularization mechanism. However, traditional naive dropout in RNNs could weaken the significant information when processing time series and result in deteriorated performance because these techniques apply different masking mechanisms to inputs and outputs at each time step. Moreover, the traditional naive zoneout cannot incorporate the masks into recurrent connections [61, 62]. To alleviate these issues in traditional methods, in MTSNet, we further add variational zoneout [63] as a regularization mechanism in the architecture. As shown in Fig. 3.1, we apply variational zoneout in Layer 3, where $\mathbf{z}^2_{i,t} \odot \mathbf{s}_z$ and $\mathbf{h}^3_{i,t-1} \odot \mathbf{s}_h$ represent the zoneout processes illustrated by the dashed green and blue arrows across all time steps. $\mathbf{s}_x$ and $\mathbf{s}_h$ denote the random initialized masks applied on the input $\mathbf{z}^2_{i,t}$ and hidden states $\mathbf{h}^3_{i,t-1}$, respectively. Then Eq. (3.5) illustrates the gating mechanism of the LTSM cell in layer 3.

$$
\begin{aligned}
\mathbf{I}_t &= sigmoid\left(\mathbf{W}_I\left(\mathbf{z}^2_{i,t} \odot \mathbf{s}_z\right) + \mathbf{U}_I\left(\mathbf{h}^3_{i,t-1} \odot \mathbf{s}_h\right) + b_I\right) \\
\mathbf{f}_t &= sigmoid\left(\mathbf{W}_f\left(\mathbf{z}^2_{i,t} \odot \mathbf{s}_z\right) + \mathbf{U}_f\left(\mathbf{h}^3_{i,t-1} \odot \mathbf{s}_h\right) + b_f\right) \\
\mathbf{o}_t &= sigmoid\left(\mathbf{W}_o\left(\mathbf{z}^2_{i,t} \odot \mathbf{s}_z\right) + \mathbf{U}_o\left(\mathbf{h}^3_{i,t-1} \odot \mathbf{s}_h\right) + b_o\right) \\
\mathbf{g}_t &= tanh\left(\mathbf{W}_g\left(\mathbf{z}^2_{i,t} \odot \mathbf{s}_z\right) + \mathbf{U}_g\left(\mathbf{h}^3_{i,t-1} \odot \mathbf{s}_h\right) + b_g\right)
\end{aligned} \tag{3.5}
$$

where $\mathbf{I}_t$,$\mathbf{f}_t$,$\mathbf{o}_t$ and $\mathbf{g}_t$ represent input gate, forget gate, output gate, and input modulation of LSTM cell in the Layer 3, respectively. $sigmoid$ and $tanh$ denote the sigmoid

activation function and hyperbolic tangent function.

$$\begin{aligned} \mathbf{c}_{i,t}^3 &= \mathbf{f}_t \odot \mathbf{c}_{i,t-1}^3 + \mathbf{I}_t \odot \mathbf{g}_t \\ \mathbf{h}_{i,t}^3 &= \mathbf{o}_t \odot tanh\left(\mathbf{c}_{i,t}^3\right) \\ \mathbf{z}_{i,t}^3 &= \mathbf{h}_{i,t}^3 \end{aligned} \tag{3.6}$$

According to Eq. (3.6), we can obtain the input of the projection layer, which is denoted by $\mathbf{z}_{i,t}^3$.

3.1.6 Probabilistic Forecasting

The probabilistic forecasting of multiple time series by MTSNet is defined as follows. Given a set of time series $\{x_{i,t}\}_{i=1}^N$ from N sources with the source index $i = 1,2\ldots,N$, where $x_{i,t} \in \mathbb{R}$ denotes the value of target multiple MTSs in source i at time t. Assume the current time point is at t_0, the historical time series values (e.g., COVID-19 case numbers) are denoted by $\mathbf{x}_{i,1:t_0}$ and the future values are $\mathbf{x}_{i,t_0+1:t_0+m}$. The goal of multifactor probabilistic forecasting is to model the conditional probability distributions of the future values $\mathbf{x}_{i,t_0+1:t_0+m}$, given their historical data of the target time series $\mathbf{x}_{i,1:t_0}$ and exogenous variables $\mathbb{F}_{i,1:t_0+m}$. The probabilistic forecasting is represented by Eq. (3.7).

$$P\left(\mathbf{x}_{i,t_0+1:t_0+m} \mid \mathbf{x}_{i,1:t_0}, \mathbb{F}_{i,1:t_0+m}, \Theta\right) \tag{3.7}$$

where Θ denotes the parameters of the global neural network, which is learnable and shared with all time series from N sources. We set the time range $\{1,2,\ldots,t_0\}$ to be the conditioning range for training and the prediction range as $\{t_0+1, t_0+2, \ldots, t_0+m\}$, $m \in \mathbb{N}$.

Solving Eq. (3.7) requires inferring a joint distribution over a period of time conditional on its past period, external factors, and modeling parameters. Assume the time series points are i.i.d. drawn from an underlying distribution, then it can be converted to the multiplication of past state probabilities:

$$\begin{aligned} &P\left(\mathbf{x}_{i,t_0+1:t_0+m} \mid \mathbf{x}_{i,1:t_0}, \mathbb{F}_{i,1:t_0+m}, \Theta\right) = \\ &\prod_{t=t_0}^{t_0+m} P\left(x_{i,t+1} \mid \mathbf{x}_{i,1:t}, \mathbb{F}_{i,1:t}, \Theta\right) \end{aligned} \tag{3.8}$$

As $\mathbf{z}_{i,t}^3$ is generated by the neural network as shown in Eq. (3.6), then $\mathbf{z}_{i,t}^3$ can be written as the function of variables $\mathbf{x}_{i,1:t}$ and $\mathbb{F}_{i,1:t}$ in Eq. (3.9), where $h(.)$ denote the

transformation of stacked LSTM (Layer 1, Layer 2 and Layer 3 as shown in Fig. 3.1). Θ_h represents their learnable parameters.

$$\mathbf{z}_{i,t}^3 = h\left(\mathbf{x}_{i,1:t}, \mathbb{F}_{i,1:t}, \mathbf{h}_{i,t-1}^3, \Theta_h\right) \tag{3.9}$$

One approach to solve Eq. (3.8) is to treat it as the product of likelihood functions and then use the maximum likelihood estimation for solutions. Accordingly, it assumes $x_{i,t+1}$ is fixed and the parameters of the likelihoods are determined by a function $k\left(\mathbf{z}_{i,t}^3, \Theta_k\right)$. $k(.)$ denotes the projection layer (the orange arrows in Fig. 3.1) with their parameters Θ_k and the input is $\mathbf{z}_{i,t}^3$ containing the information of the past target MTS values and exogenous factors. Then Eq. (3.8) can be converted to the product of the likelihood functions $\ell\left(x_{i,t+1} \mid k\left(\mathbf{z}_{i,t}^3, \Theta_k\right)\right)$ in Eq. (3.10), as the global parameters Θ consists of Θ_h in $h(.)$ and Θ_k in $k(.)$.

$$\begin{aligned} &\prod_{t=t_0}^{t_0+m} P\left(x_{i,t+1} \mid \mathbf{x}_{i,1:t}, \mathbb{F}_{i,1:t}, \Theta\right) \\ &= \prod_{t=t_0}^{t_0+m} \ell\left(x_{i,t+1} \mid k\left(\mathbf{z}_{i,t}^3, \Theta_k\right)\right) \end{aligned} \tag{3.10}$$

Therefore, MTSNet makes multiple steps of transformations to generate their hidden representation $\mathbf{z}_{i,t}^3$ and then uses $k(.)$ to map it to the unknown parameters of the likelihood functions.

3.1.7 Probabilistic Projection

In probabilistic modeling, different distributions determine the dynamics and trends of underlying factors. Their likelihood function generates target values by sampling from the distributions. The statistical properties of the training data are sensitive to the assumed probability distribution. Numerical MTSs like COVID-19 confirmed cases follow continuous probability distributions. Since MTS and its external multifactors may follow different distributions, we test three probability distributions: Gaussian, Laplace, and Student's t-distribution. Eq. (3.11), Eq. (3.12) and Eq. (3.13) represent their likelihood functions ($\ell_L(.)$, $\ell_G(.)$, $\ell_S(.)$) and the corresponding formula of its projection layer ($k_L(.)$, $k_G(.)$, $k_S(.)$) for Laplace, Gaussian, and Student's t distribution, respectively.

$$\begin{aligned} \ell_L(x_{i,t} \mid \mu, \sigma) &= \frac{1}{2\sigma} \exp\left(-\frac{|x_{i,t} - \mu|}{\sigma}\right) \\ k_L(\mathbf{z}_{i,t}^3) &= \begin{cases} \mu\left(\mathbf{z}_{i,t}^3\right) = \mathbf{w}_\mu^T \mathbf{z}_{i,t}^3 + b_{L\mu} \\ \sigma\left(\mathbf{z}_{i,t}^3\right) = \log\left(1 + \exp\left(\mathbf{w}_\sigma^T \mathbf{z}_{i,t}^3 + b_\sigma\right)\right) \end{cases} \end{aligned} \tag{3.11}$$

$$\ell_G(x_{i,t} \mid \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x_{i,t}-\mu)^2}{2\sigma^2}\right)$$
$$k_G(\mathbf{z}_{i,t}^3) = \begin{cases} \mu\left(\mathbf{z}_{i,t}^3\right) = \mathbf{w}_\mu^T \mathbf{z}_{i,t}^3 + b_\mu \\ \sigma\left(\mathbf{z}_{i,t}^3\right) = \log\left(1 + \exp\left(\mathbf{w}_\sigma^T \mathbf{z}_{i,t}^3 + b_\sigma\right)\right) \end{cases} \tag{3.12}$$

$$\ell_S(x_{i,t} \mid \nu, \mu, \sigma) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\nu}\,\sigma}\left(1 + \frac{1}{\nu}\left(\frac{x_{i,t}-\mu}{\sigma}\right)^2\right)^{-\frac{\nu+1}{2}}$$
$$k_S(\mathbf{z}_{i,t}^3) = \begin{cases} \mu\left(\mathbf{z}_{i,t}^3\right) = \mathbf{w}_\mu^T \mathbf{z}_{i,t}^3 + b_\mu \\ \sigma\left(\mathbf{z}_{i,t}^3\right) = \log\left(1 + \exp\left(\mathbf{w}_\sigma^T \mathbf{z}_{i,t}^3 + b_\sigma\right)\right) \\ \nu\left(\mathbf{z}_{i,t}^3\right) = 2 + \log\left(1 + \exp\left(\mathbf{w}_\nu^T \mathbf{z}_{i,t}^3 + b_\nu\right)\right) \end{cases} \tag{3.13}$$

The projection layer $k_L(.)$, $k_G(.)$ and $k_S(.)$ map the hidden representation $\mathbf{z}_{i,t}^3$ to the corresponding parameters of assumed distribution. The calculation of their standard deviation is slightly different from their mean. Specifically, we apply a Softplus activation function to generate their positive values for the standard deviation. According to [12], this projection layer can facilitate a fast convergence. For the additional parameter ν in Student's t-distribution, which denotes the degree of freedom, we apply a similar approach to let $\nu \geq 2$ in Eq. (3.13).

With the above distributional assumptions and the mechanism of the projection layer, we can establish the connection between the DNNs and the probability distribution. Further, it aims to define a likelihood function and calculate the gradients of the loss. Then, the parameters of MTSNet are optimized by backpropagation.

3.1.8 Likelihood Loss

After specifying the conditional distribution of target data, we apply maximum likelihood estimation to the loss function to update the parameters of MTSNet. The loss function is shown in Eq. (3.14), where we use the negative log-likelihood. As the previously defined likelihood functions meet all the requirements for calculating the gradients in Eq. (3.14), it further can be optimized by stochastic gradient descent directly.

$$\mathscr{L} = -\sum_{i=1}^{N}\sum_{t=t_0}^{t_0+m} \log \ell\left(x_{i,t} \mid k\left(\mathbf{z}_{i,t}^3\right)\right) \tag{3.14}$$

During the training process, both the conditioning range $[1, 2, \ldots, t_0]$ and the prediction range $[t_0+1, t_0+2, \ldots, t_0+m]$ are available. However, in the prediction process, $x_{i,t}$

is only available in the conditioning range. We use $x_{i,t}$ ($t \leq t_0$) as input for training and $\tilde{x}_{i,t}$ ($t > t_0$) is drawn from the distribution $p\left(\cdot \mid \Theta_{i,t}\right)$. Then $\tilde{x}_{i,t}$ is used to generate the next distribution and conduct the sampling again until $t = t_0 + m$. After repeating this process, we can get a set of $\tilde{x}_{i,t}$ in the time range $[t_0 + 1,\, t_0 + m]$ by ancestral sampling from the estimated probability distributions as shown in Eq. (3.15).

$$\tilde{\mathbf{x}}_{i,t_0+1:t_0+m} \sim P\left(\mathbf{x}_{i,t_0+1:t_0+m} \mid \mathbf{x}_{i,1:t_0}, \mathbb{F}_{i,1:t_0+m}, \Theta\right) \tag{3.15}$$

Then, we use these samples to calculate the quantiles and other indices of the posterior distribution in the future time period.

3.2 Experiments and Evaluation

3.2.1 The COVID-19 Data with External Factors

Here, we focus on COVID-19 time-series cases and their affiliated external factors, including the proportion of Omicron, vaccine booster rates, Stringency Index [46], and reproduction number. They represent the factors of virus mutation, vaccination, government interventions, and virus infectivity, respectively. We choose this COVID-19 data from six representative countries: UK, US, Canada, France, Spain, and Italy. These countries are chosen because their COVID-19 epidemic showed distinct characteristics and they took different NPI strategies in mitigating COVID-19.

- *Historical cases:* The COVID-19 case data includes daily confirmed cases and the number of ICU patients, acquired from Johns Hopkins University's COVID-19 data repository. The UK, US, Canada, France, Spain, and Italy data from March 13, 2020 to February 15, 2022 are crawled. Our model works as a global one to fit the data from different countries, which thus illustrates the generalization capability of the model.

- *Virus mutation:* We use the proportion of Omicron variants in the UK, US, Canada, France, Spain, and Italy. This data was obtained from the GISAID initiative[1].

- *Vaccination:* We use the rate of vaccine booster, which is derived from the total number of applied vaccine boosters divided by the total population of the country. The data is cumulative and the time interval is one day[2].

[1]The data is available at covariants.org.

[2]The full dataset is available at: https://github.com/owid/covid-19-data/tree/master/public/data/vaccinations

- *Government interventions:* For quantitative information about a government's NPIs on COVID-19, we use the Stringency Index [46] of the six countries from March 13, 2020, to February 15, 2022. The Stringency Index is calculated on nine different policy indicators: school closing, workplace closing, canceling public events, restrictions on gathering size, closed public transport, stay-at-home requirements, restrictions on internal movement, restrictions on international travel, and public information campaigns. The data is available at the Oxford Covid-19 Government Response Tracker.

- *Infectivity:* Our model also involves the daily reproduction rate [64] from different countries to represent the infectivity of the virus[3].

We extract the COVID-19 data from 13/03/2020 to 15/02/2022 for training to predict the daily confirmed cases and daily ICU patient number in the coming 10 days. At this moment, we cannot find other time-series data with multiple external factors like the COVID-19 problem for the experiments and evaluating our design.

3.2.2 Baseline Models and Settings

We compare our proposed model MTSNet with SOTA RNN networks for probabilistic forecasting, including DeepAR, Deep State Space Model (DSM), and Deep Factor (DF). We do not compare MTSNet with classic methods including ARIMA and ETS and DNNs including RNN and LSTM since they have been shown incapable of competing with DeepAR, DSM, and DF and they focus on point-based rather than probabilistic forecasting.

- DeepAR [12]: It contains two layers of RNNs to extract features and then maps the hidden states of the RNN cells to the parameters of an assumed distribution of the input.

- DSM [13]: It combines a linear state space model with RNNs. By introducing Gaussian noise, the marginal likelihood function of the linear state space model can be analytically tractable, which is used in the loss function to train the neural networks.

- DF [14]: It uses RNNs to process global time series and generate the latent representation of each time series. Then the Kalman filter and other two likelihood

[3]The data is available at:https://github.com/crondonm/ TrackingR/tree/main/Estimates-Database

function estimation techniques based on the assumption of a Gaussian distribution are tested in this study.

For MTSNet, we create its variants for testing the effect of different distributions and for the ablation study of involving external factors or not in the network.

- fMTSNet-Laplace: referring to the MTSNet incorporated with Laplace distribution and external factors;
- vMTSNet-Laplace: referring to the MTSNet incorporated with Laplace distribution without variational zoneout;
- MTSNet-Laplace: referring to the MTSNet incorporated with Laplace distribution without external factors;
- fMTSNet-Gaussian: referring to the MTSNet incorporated with Gaussian distribution and external factors;
- vMTSNet-Gaussian: referring to the MTSNet incorporated with Gaussian distribution without variational zoneout;
- MTSNet-Gaussian: referring to the MTSNet incorporated with Gaussian distribution without external factors;
- fMTSNet-StudenT: referring to the MTSNet incorporated with Student's t-distribution and external factors;
- vMTSNet-StudenT: referring to the MTSNet incorporated with Student's t-distribution distribution without variational zoneout;
- MTSNet-StudenT: referring to the MTSNet incorporated with Student's t-distribution distribution without external factors.

The probabilistic baseline models are based on MXNet [65] and Gluon Time Series Toolkit [66]. All experiments are undertaken with the same CPU. We set the length of context windows t_0 and prediction windows length m to be $5:1$. All models initially optimize hyperparameters through a back-test strategy before we make the prediction. Accordingly, 100 samples are drawn from the decoder to generate future results. The above settings are also applied in the baseline models. In addition, we also conduct a comprehensive ablation study to gauge the effect of incorporating the exogenous variables for COVID-19 prediction.

3.2.3 Evaluation Metrics

Metrics including normalized deviation (ND), root mean square error (RMSE), and weighted quantile loss (WQL) are used to gauge the performance of MTSNet and its baselines. We apply them to comparing the gaps between observations $x_{i,t}$ and the predicted values $\tilde{x}_{i,t}$ at time t for country i.

ND is defined in Eq. (3.16). It measures the overall deviation of the predicted results from the true values in terms of the sum of error between observations and predicted results, divided by the sum of observations. A lower value of ND indicates better performance.

$$\text{ND} = \frac{\sum_{i,t} \left| x_{i,t} - \tilde{x}_{i,t} \right|}{\sum_{i,t} \left| x_{i,t} \right|} \tag{3.16}$$

RMSE, defined in Eq. (3.17), calculates the root of the sum of squared errors. In comparison with ND, RMSE is more sensitive to the deviation between predicted values and true values. n refers to the number of time series, i.e., $n = 6$ for six countries. T refers to the number of time steps for prediction.

$$\text{RMSE} = \sqrt{\frac{1}{nT} \sum_{i,t} \left(\tilde{x}_{i,t} - x_{i,t} \right)^2} \tag{3.17}$$

ND and RMSE only use the median value of the predicted distribution, they are thus limited in evaluating the accuracy of the probability distribution. Alternatively, we calculate the weighted quantile loss (WQL) of the results at different quantiles in the assumed distribution. WQL is defined in Eq. (3.18), where τ represents a quantile, which is set to $[0.1, 0.5, 0.9]$ in our experiments. $q_{i,t}^{(\tau)}$ denotes the value at τ-quantile of the predicted probability distribution. WQL measures the accuracy of probabilistic forecasting at a given quantile.

$$\text{WQL}[\tau] = 2 \frac{\sum_{i,t} Q_{i,t}^{(\tau)}}{\sum_{i,t} \left| x_{i,t} \right|}$$

$$Q_{i,t}^{(\tau)} = \begin{cases} (1-\tau) \left| q_{i,t}^{(\tau)} - x_{i,t} \right|, \text{if } q_{i,t}^{(\tau)} > x_{i,t} \\ \tau \left| q_{i,t}^{(\tau)} - x_{i,t} \right|, \text{otherwise} \end{cases} \tag{3.18}$$

To report the overall performance of a model over all quantiles $[0.1, 0.5, 0.9]$, we calculate the mean weighted Quantile Loss, mWQL, which is the mean of WQL at all quantiles.

3.2.4 Result Analysis

We compare the three variants of MTSNet with DeepAR, DSM, and DF. The experimental results on the next 10-day confirmation cases and ICU patients are shown in Table 3.1 and Table 3.2, respectively.

The results indicate the effectiveness of MTSNet in competing with the baselines in probabilistic forecasting. In terms of both ND and RMSE, the MTSNet with Student's t-distribution achieves the best performance for daily confirmed case prediction. The MTSNet with the Laplace distribution reaches the best RMSE results for ICU patient prediction, whereas DSM has the best ND for this task.

Specifically, as shown in Table 3.1, MTSNet variants improve the performance by at least 10.12%, 1.79%, and 10.5% over the baseline models in terms of mWQL, RMSE, and ND, respectively. In Table 3.2, the fMTSNet-Laplace outperforms the baseline models in terms of most metrics except ND. Although the MTSNets based on the Gaussian and Laplace distribution outperform all baselines well, they achieve similar results as DeepAR and DSM but better results than DF.

Among the variants of MTSNet, the model that performs well in predicting daily confirmed cases cannot guarantee consistent performance in forecasting ICU patients. This result may be owing to the different underlying distributions between cases and ICU patient number. Accordingly, a probability distribution characterizes the case time series well and does not capture the uncertainty of ICU patient number over time. As indicated in Table 3.1 and Table 3.2, the experimental results are very sensitive to the probability distribution. For example, the MTSNet with Gaussian distribution achieved the best WQL[0.9] in predicting daily confirmed cases, which, however, performs poorly in predicting the number of ICU patients. This finding shows the challenge in forecasting heterogeneous MTS and in designing a both flexible and general MTS model fitting all MTSs. The MTSNet architecture allows to incorporate different probability distributions for heterogeneous MTSs.

Fig. 1.1 further shows the predicted daily confirmed cases in the next 10 days from 06-02-2022 to 15-02-2022 by MTSNet-Laplace. We show the median value, and value at 50% and 90% confidence intervals, respectively. The results show, while it is promising to capture the case movement by deep neural probabilistic forecasting, it is indeed very challenging to have a model reliable and general for small data with high uncertainty and inconsistency across multiple countries.

Fig. 3.3 further shows the forecasted number of daily ICU patients in the next 10 days by fMTSNet-Laplace from 06-02-2022 to 15-02-2022. Due to the high dimensionality

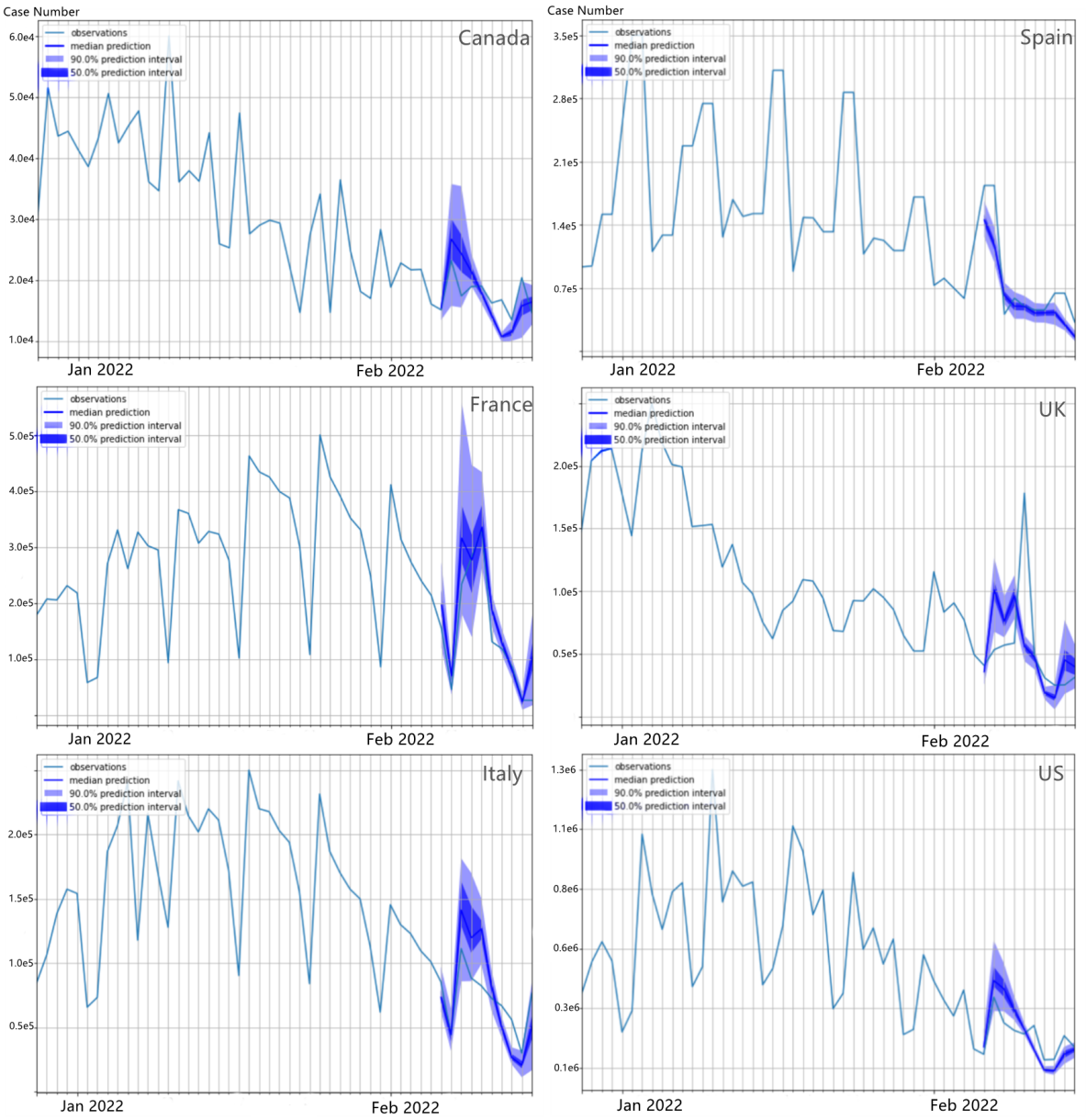


Figure 3.2: Prediction of Daily Confirmed Cases in 6 Countries by fMTSNet-Laplace with exogenous variables.

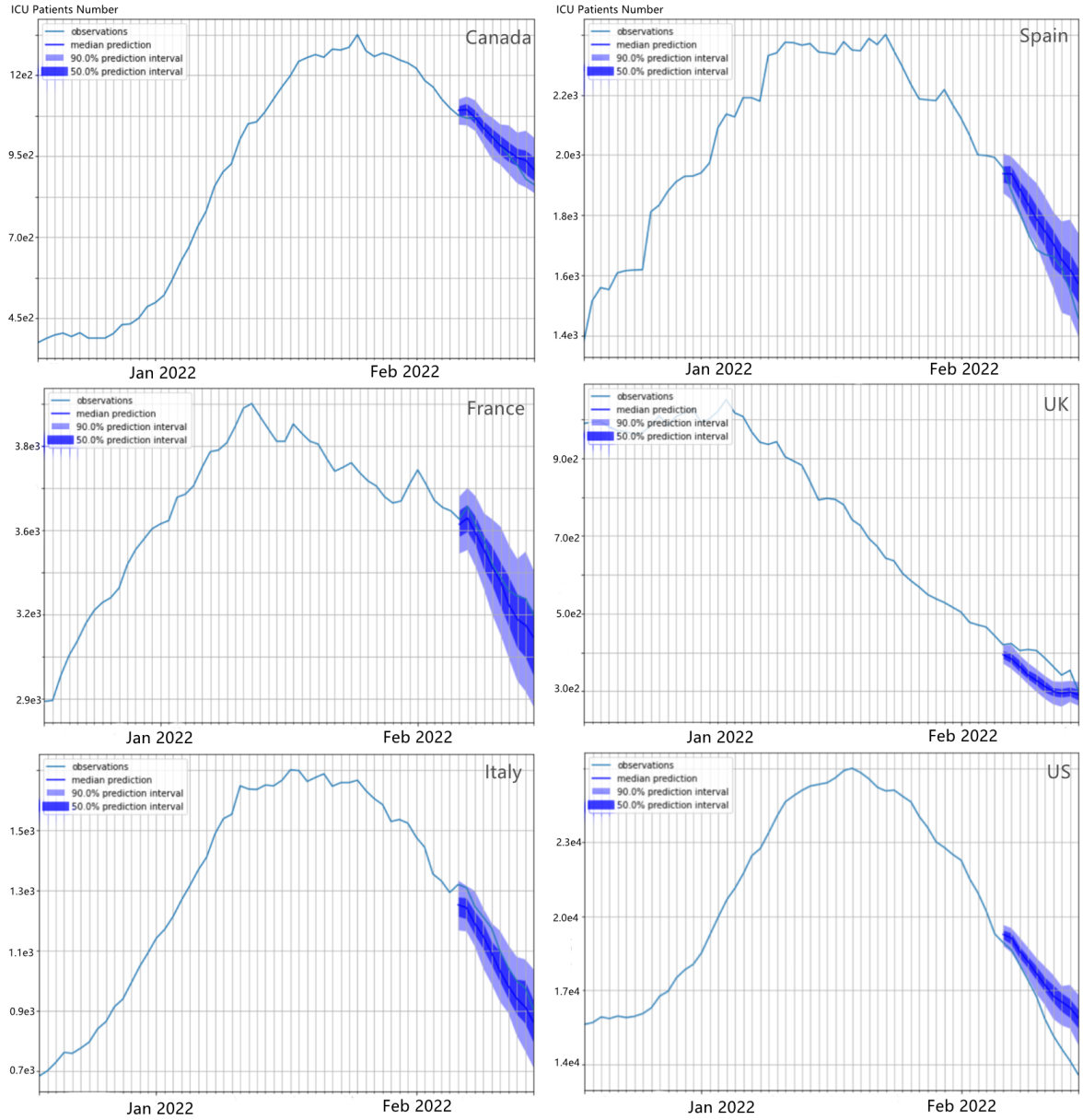


Figure 3.3: Prediction of ICU patients in 6 Countries by fMTSNet-Laplace with exogenous variables.

of the prediction results from multiple countries, it is infeasible to compare them by plotting them in a single graph. Here, we select the probabilistic forecasting results from fMTSNet-Laplace as an example to visualize the predictions in median, 50%, and 90% confidence intervals respectively. The results show that our network is promising in capturing the ICU patient occupancy trend in general. However, it is challenging to fit all countries well in one model in the small COVID-19 ICU data.

Table 3.1: Prediction results for the daily confirmed cases in the next 10 days

Metrix	fMTSNet-Laplace	fMTSNet-Gaussian	fMTSNet-StudentT	DeepAR	DF	DSM
WQL[0.1]	**0.1461**	0.1741	0.1613	0.3107	0.3850	0.6970
WQL[0.5]	0.3401	0.3676	**0.3395**	0.5403	0.4727	0.6599
WQL[0.9]	0.2692	**0.2324**	0.2817	0.2353	0.2627	0.2781
mWQL	**0.2518**	0.2580	0.2608	0.3621	0.3735	0.5450
RMSE	0.1046	0.1173	**0.1031**	0.1351	0.1450	0.1858
ND	0.3401	0.3676	**0.3395**	0.5403	0.4727	0.6599

Table 3.2: Prediction results for the daily ICU patients in the next 10 days

Metrix	fMTSNet-Laplace	fMTSNet-Gaussian	fMTSNet-StudentT	DeepAR	DF	DSM
WQL[0.1]	**0.0682**	0.0804	0.0709	0.0993	0.2529	0.1269
WQL[0.5]	0.1388	0.1950	0.1794	0.1311	0.1720	**0.1157**
WQL[0.9]	**0.0542**	0.1196	0.1005	0.0721	0.0817	0.0745
mWQL	**0.0871**	0.1317	0.1169	0.1008	0.1689	0.1057
RMSE	**0.0934**	0.1328	0.1210	0.0942	0.1318	0.1346
ND	0.1388	0.1950	0.1794	0.1311	0.1720	**0.1157**

Table 3.3: Ablation studies of predicting daily confirmed cases for external factors

Metrix	fMTSNet-Laplace	MTSNet-Laplace	fMTSNet-Gaussian	MTSNet-Gaussian	fMTSNet-StudentT	MTSNet-StudentT
WQL[0.1]	**0.1461**	0.2731	**0.1741**	0.2899	**0.1613**	0.3847
WQL[0.5]	**0.3401**	0.4491	**0.3676**	0.5644	**0.3395**	0.4754
WQL[0.9]	**0.2692**	0.3835	**0.2324**	0.3860	0.2817	**0.2733**
mWQL	**0.2518**	0.3686	**0.2580**	0.4134	**0.2608**	0.3778
RMSE	**0.1046**	0.1394	**0.1173**	0.1747	**0.1031**	0.1409
ND	**0.3401**	0.4491	**0.3676**	0.5644	**0.3395**	0.4754

Table 3.4: Ablation studies of predicting daily ICU patients for external factors

Metrix	fMTSNet-Laplace	MTSNet-Laplace	fMTSNet-Gaussian	MTSNet-Gaussian	fMTSNet-StudentT	MTSNet-StudentT
WQL[0.1]	**0.0682**	0.1222	**0.0804**	0.1296	**0.0709**	0.0719
WQL[0.5]	**0.1388**	0.1985	**0.1950**	0.2637	**0.1794**	0.2209
WQL[0.9]	**0.0542**	0.1271	**0.1196**	0.1563	**0.1005**	0.1123
mWQL	**0.0871**	0.1493	**0.1317**	0.1832	**0.1169**	0.1350
RMSE	**0.0934**	0.1305	**0.1328**	0.1918	**0.1210**	0.1655
ND	**0.1388**	0.1985	**0.1950**	0.2637	**0.1794**	0.2209

Table 3.5: Ablation studies of predicting daily confirmed cases for variational zoneout

Metrix	fMTSNet-Laplace	vMTSNet-Laplace	fMTSNet-Gaussian	vMTSNet-Gaussian	fMTSNet-StudentT	vMTSNet-StudentT
WQL[0.1]	**0.1461**	0.1597	0.1741	**0.1591**	**0.1613**	0.1815
WQL[0.5]	0.3401	**0.3305**	**0.3676**	0.3947	**0.3395**	0.3545
WQL[0.9]	**0.2692**	0.2728	**0.2324**	0.3022	0.2817	**0.2720**
mWQL	**0.2518**	0.2543	**0.2580**	0.2853	**0.2608**	0.2693
RMSE	**0.1046**	0.1063	**0.1173**	0.1219	**0.1031**	0.1119
ND	**0.3401**	0.3305	**0.3676**	0.3947	**0.3395**	0.3545

Table 3.6: Ablation studies of predicting daily ICU patients for variational zoneout

Metrix	fMTSNet-Laplace	vMTSNet-Laplace	fMTSNet-Gaussian	vMTSNet-Gaussian	fMTSNet-StudentT	vMTSNet-StudentT
WQL[0.1]	**0.0682**	0.0907	**0.0804**	0.1210	0.0709	**0.0650**
WQL[0.5]	**0.1388**	0.1548	**0.1950**	0.2212	**0.1794**	0.2024
WQL[0.9]	**0.0542**	0.0768	0.1196	**0.0818**	**0.1005**	0.1330
mWQL	**0.0871**	0.1074	**0.1317**	0.1413	**0.1169**	0.1335
RMSE	**0.0934**	0.1072	**0.1328**	0.1738	**0.1210**	0.1500
ND	**0.1388**	0.1548	**0.1950**	0.2212	**0.1794**	0.2024

3.2.5 Effect of Exogenous Variables

Here, we evaluate the effect of concatenating external variables on forecasting the COVID-19 confirmed cases and ICU patient number in the next 10 days. We conduct an ablation study by comparing each MTSNet variant using external variables with the variant without these variables, i.e., fMTSNet-Laplace vs. MTSNet-Laplace, fMTSNet-Gaussian vs. MTSNet-Gaussian, and fMTSNet-StudenT vs. MTSNet-StudenT. The experimental results are presented in Table 3.3 and Table 3.4. Overall, they show the effect of incorporating exogenous variables for COVID-19 probabilistic forecasting.

Specifically, Table 3.3 for predicting daily confirmed cases shows that the networks with exogenous variables achieve superior results with an improvement of at least 11.68%, 3.48%, and 10.90% in terms of mWQL, RMSE, and ND, respectively. Furthermore, all networks using exogenous variables outperform their corresponding ones without these variables regarding WQL at [0.1],[0.5],[0.9]-quantile for this task.

In Table 3.4 for predicting the number of ICU patients, the MTSNet variants with exogenous variables improve the performance by at least 1.809%, 4.45%, and 4.14% in terms of mWQL, RMSE, and ND, respectively. A notable point is that the overall improvement is not as significant as for case forecasting. This may be because the case time series may be more strongly correlated with the external variables than the ICU patient number.

3.2.6 Effect of Variational Zoneout

Tables 3.5 and 3.6 report the results of the ablation study by comparing each MTSNet variant using the variational zoneout (fMTSNet-Laplace, fMTSNet-Gaussian, fMTSNet-StudenT) with the variant without the variational zoneout (vMTSNet-Laplace, vMTSNet-Gaussian, vMTSNet-StudenT). All of the MTSNet variants incorporate external factors for comparison. Although several indicators such as WQL[0,1] and WQL[0.5] have not improved in some model variants (fMTSNet-Laplace, fMTSNet-Gaussian), most indicators for the overall performance demonstrate the contribution of incorporating the

variational zoneout in the two tasks. We note that all networks using variational zoneout outperform the variants without using it regarding mWQL, RMSW, and ND.

3.3 Summary

Modeling coupled multivariate time series is crucial in solving the problem of accurately forecasting COVID-19 transmission using multiple external factors. Due to the high-level uncertainty of COVID-19 transmission caused by various external factors. It is necessary to develop a probabilistic forecasting model capable of handling multiple exogenous variables. However, the majority of the existing studies focus on point-based estimation instead of investigating probabilistic prediction using multiple external factors including virus mutations, vaccination, government interventions, and infectivity. In this chapter, we propose a global model for probabilistic forecasting of COVID-19 daily confirmed cases and the number of ICU patients from multiple countries using exogenous variables. And this approach is flexible and compatible with different assumed probability distributions. We utilize a stacked RNN structure to learn a compound tensor and generate the parameters of the assumed probability distribution. And the Residual Connection and Variational Dropout are implemented to improve the robustness. The experimental results demonstrate the superiority of the model and the effectiveness of incorporating the exogenous variables.

The Deep Probabilistic Cross-MTS Network presents notable limitations that merit careful consideration. Firstly, the inherent complexity of its model architecture entails challenges in terms of interpretability and scalability, necessitating substantial computational resources and training efforts. Secondly, the assumptions regarding data distribution, often relying on certain probabilistic distribution assumptions, may not adequately capture the intricate nature of real-world data, potentially compromising the model's performance. Furthermore, while MTSNet has demonstrated promise in specific domains, its generalizability to diverse contexts remains an open question, urging further investigation. Addressing these limitations is imperative for refining the robustness, scalability, and applicability of MTSNet in the analysis of coupled multivariate time series data.

CHAPTER 4

CONTRIBUTION - EXPLICITLY MODELING INTRA- AND INTER-MTS COUPLING WITH VOLATILE COVARIATES

4.1 Deep spectral copula mechanisms

4.1.1 Introduction

As the previous chapters introduce a research question concerning modeling coupled and volatile multivariate time series, this chapter proposes deep spectral copula mechanisms (DSCM) to tackle this problem with a focus on its challenges (5)-(7).

Firstly, DSCM incorporates a singular spectral analysis (SSA) module to reduce the volatility of multiple covariates. Secondly, it applies an intra-MTS coupling module to explicitly model the temporal couplings within a single set of multivariate time series; Thirdly, DSCM transforms target variables into joint probability distributions via Gaussian copula transformation to establish inter-MTS couplings across multiple multivariate time series. Substantial experiments on COVID-19 time-series data from multiple countries indicate the superiority of DSCM over state-of-the-art approaches.

4.1.2 The DSCM Architecture

DSCM aims to tackle the challenges of modeling Coupled and volatile multivariate time series (CVMTS). As depicted in Figure 4.4, to handle volatile covariates, DSCM implements the SSA Module to extract the trend features of the volatile covariates and

input them into the neural network to obtain hidden representations. To capture the temporal intra-couplings within multiple multivariate time series, DSCM utilizes a deep neural network to learn the weights of hidden representations at each time step. To model the inter-couplings between multiple multivariate time series, DSCM incorporates Gaussian copula probabilistic transformation into the algorithms. The copula function connects multiple target time series at each time step to form a joint multivariate probabilistic distribution.

Coupled and volatile multivariate time series (CVMTS) consisting of target variables time-series $\mathbf{x}$ and multiple volatile covariates time-series $\mathbf{F}$ have both inter- and intra-MTS couplings as shown in Figure 1.1. For the consider COVID-19 data, we assume $\mathbf{x}_t$ represents a target vector composed of the values of the multivariate time series at time step t, e.g., confirmed cases from N countries. $\mathbf{F}_t$ represents the volatile multiple covariates corresponding to the target multivariate time series, e.g., test number, stringency index, the share of virus variants, vaccination rate, non-pharmaceutical interventions (NPIs), and virus infectivity from N countries.

The deep probabilistic forecasting and projection mechanisms in DSCM follow the DeepVAR framework [15]. The difference lies in that we incorporate the volatile covariates of multivariate time series and the intra-coupling module into the DSCM. DSCM aims to tackle the challenges of modeling CVMTS. It characterizes the CVMTS forecasting problem in Eq. (4.1),

$$p\left(\mathbf{x}_{t_0+1:t_0+m} \mid \mathbf{x}_{1:t_0}, \mathbf{F}_{1:t_0+m}, \Theta\right), \tag{4.1}$$

where Θ denotes the parameters of DSCM, the historical time series steps are $\{1, 2, \ldots, t_0\}$, and the prediction time length is $m \in \mathbb{N}$. DSCM models the conditional joint multivariate probability distributions of the future multivariate time series $\mathbf{x}_{t_0+1:t_0+m}$, given the historical multivariate time series $\mathbf{x}_{1:t_0}$ and their volatile covariates $\mathbf{F}_{1:t_0+m}$.

4.1.3 The SSA Module

The SSA module serves the purpose of trend extraction and volatility reduction. It has been experimentally proven that eliminating the volatility caused by random fluctuations in multiple covariates can improve the overall robustness of the framework. In addition, the trend characteristics of covariates also facilitate long-term probabilistic forecasting.

The research work [67–70] reveals that SSA can improve smoothing, trend feature extraction, and volatility reduction for time series through decomposition and reconstruction. Accordingly, we apply a similar approach to the fluctuated covariates. The difference

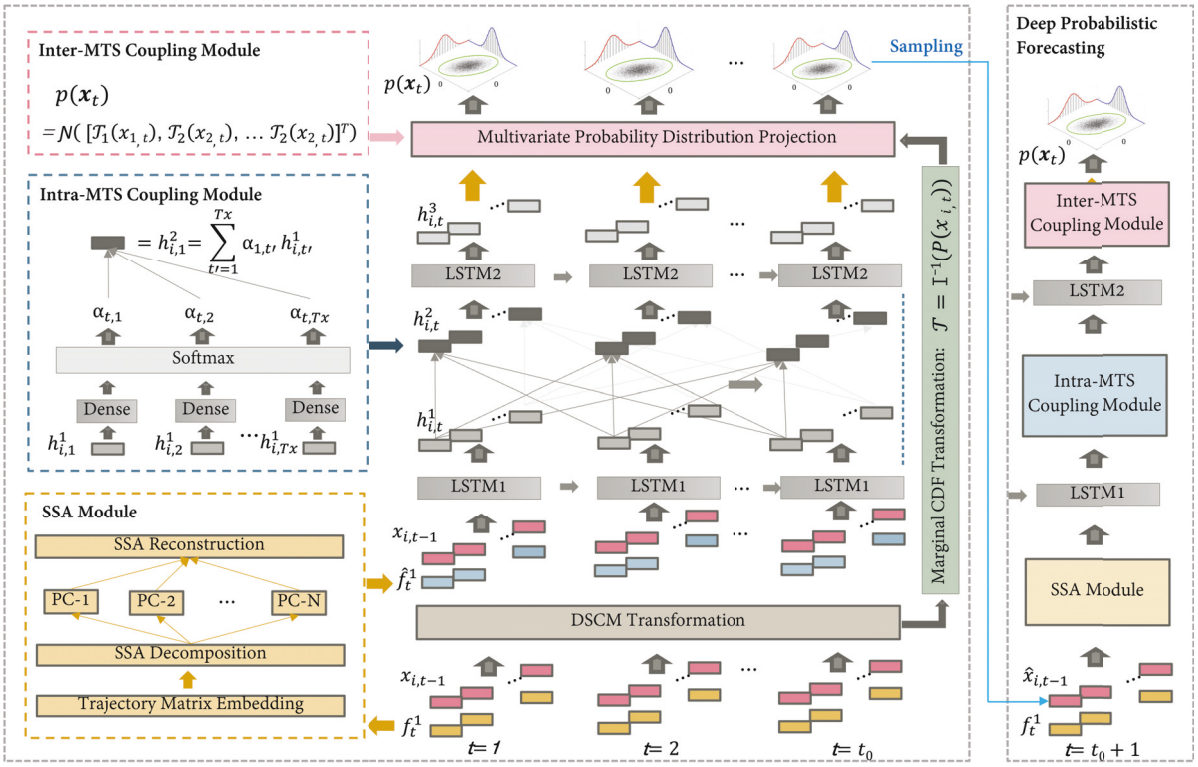


Figure 4.1: The framework of Deep Spectral Copula Mechanisms (DSCM). (1) It incorporates a singular spectral analysis (SSA) module to reduce the fluctuation of covariates. (2) It applies an intra-MTS coupling module to explicitly learn the temporal couplings inside a single set of multivariate time series. (3) To establish inter-MTS couplings across the correlated multivariate time series, the target variables are converted into joint probability distributions using the Gaussian copula transformation.

is that we use multiple reconstructed covariates vectors to construct a two-dimensional tensor as input. The SSA Module includes trajectory matrix embedding, decomposition, and multiple covariates reconstruction.

4.1.3.1 Trajectory Matrix Embedding and Decomposition

Let $\mathbf{f^i} = \left(f_1^i, f_2^i, \ldots, f_T^i\right)$ denote a covariates time series (e.g., test number of country i), where $f_t^i \in \mathbb{R}$ is a single value at time t. In order to perform the singular value decomposition (SVD) on a time series, we first need to construct a Hankel matrix, $\mathbf{H^i} \in \mathbb{R}^{K \times L}$, which contains L lagged vectors with $K = T - L + 1$ window size as shown in

Eq. (4.2).

$$\begin{aligned}\mathbf{H}^i &= \left[\boldsymbol{f}_1^i, \boldsymbol{f}_2^i, \ldots, \boldsymbol{f}_K^i\right] \\ &= \begin{bmatrix} f_1^i & f_2^i & f_3^i & \cdots & f_K^i \\ f_2^i & f_3^i & f_4^i & \cdots & f_{K+1}^i \\ f_3^i & f_4^i & f_5^i & \cdots & f_{K+2}^i \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ f_L^i & f_{L+1}^i & f_{L+2}^i & \cdots & f_T^i \end{bmatrix}.\end{aligned} \tag{4.2}$$

In the trajectory matrices with a stacked Hankel structure, the elements along the antidiagonal are equal to each other. And the multiple covariates are represented by those Hankel matrices, which will be decomposed using SVD and formulated as,

$$\mathbf{H}^i = \mathbf{U}\Sigma\mathbf{V}^{\mathbf{T}}, \tag{4.3}$$

Where $\mathbf{U} \in \mathbb{R}^{K\times K}$ and $\mathbf{V} \in \mathbb{R}^{L\times L}$ are the matrices composed of the singular vectors. $\mathbf{\Sigma} \in \mathbb{R}^{K\times L}$ is a rectangular matrix in which the diagonal elements are all the singular values of the trajectory matrix $\mathbf{H^i}$. Then it can be written in the form of the sum of linearly independent matrices $\mathbf{G}^j$ $(j = 1,\ldots,r)$,

$$\mathbf{H}^i = \sum_{j=1}^{r} \sigma_j \mathbf{u}_j \mathbf{v}_j^T = \mathbf{G}^1 + \mathbf{G}^2 + \ldots + \mathbf{G}^r, \tag{4.4}$$

which can be represented by the singular values σ_i, vector $\mathbf{u}_i$ and $\mathbf{v}_i$ (eigentriple). r is the total number of decomposed components.

4.1.3.2 Multiple Covariates Reconstruction

After the SVD decomposition, we select a group of decomposed components to reconstruct the multiple time series. Traditionally, the grouping process divides the selected basic matrices $\mathbf{G}^j$ into a number of groups and then adds the submatrices together in each group. Here, we assume each non-zero singular value is grouped individually. We then choose v components to reconstruct the covariates.

To reconstruct the time series using selected components, we calculate each anti-diagonal average $\mathbf{g}^j$ over $\mathbf{G}^i$ as shown in Eq. (4.5),

$$g_k^j = \begin{cases} \frac{1}{k}\sum_{m=1}^{k} G_{m,k-m+1}^j, & 1 \le k < L^* = \min\{L,K\} \\ \frac{1}{L^*}\sum_{m=1}^{L^*} G_{m,k-m+1}^j, & L^* \le k \le K^* = \max\{L,K\} \\ \frac{1}{T-k+1}\sum_{k-K^*+1}^{T-K^*+1} G_{m,k-m+1}^j, & K^* < k \le T \end{cases}, \tag{4.5}$$

where $G^i_{m,k-m+1}$ denotes the elements of row m and column $k-m+1$ of the matrix $\mathbf{G}^i$. The aim of Eq. (4.5) is to calculate the anti-diagonal average for each selected component.

Then, g^j_k is used to construct a vector $\mathbf{g}^j$

$$\begin{aligned} \mathbf{g}^j &= \left[g^j_1, g^j_2, \ldots, g^j_T\right]^T, \\ \hat{\mathbf{f}}^i &= \sum_{j=1}^{v} \mathbf{g}^j, \end{aligned} \tag{4.6}$$

which has the same length as the original covariates. Further, we sum up all the components to obtain a single reconstructed vector. Finally, for the multiple volatile covariates time-series $\mathbf{F}$, its reconstructed multiple covariates tensor $\hat{\mathbf{F}}$ can be obtained through a concatenation operation over these vectors,

$$\hat{\mathbf{F}} = \left[\hat{\mathbf{f}}^1, \hat{\mathbf{f}}^2, \ldots, \hat{\mathbf{f}}^i\right]. \tag{4.7}$$

Figure 4.2 uses the test number data of Italy to illustrate the effects of volatility reduction and trend feature extraction in this module.

4.1.4 The Inter-MTS Coupling Module

The inter-MTS coupling module in DSCM implements the Gaussian copula function. It transfers the multiple marginal probability distributions of the target random variables at each time step to a joint probabilistic distribution. The marginal probability distribution is assumed to be a multivariate Gaussian distribution as described in [15].

Given the observations $x_{i,t}$ from target time series i at the time point t, we assume that the set $\{x_{i,t}\}_{i=1}^N$ follows a joint probability distribution. According to Sklar‚Äôs theorem [71], the joint cumulative distribution $Q(.)$ can be represented by a copula function $C(.)$ and their marginal distribution $P(.)$.

$$Q\left(x_{1,t}, \ldots, x_{N,t}\right) = C\left(P_1\left(x_{1,t}\right), \ldots, P_N\left(x_{N,t}\right)\right). \tag{4.8}$$

To make the joint probability distribution tractable, we assume the copula function is a multivariate Gaussian. As illustrated in Eq. (4.9), $\Phi_{\boldsymbol{\mu},\Sigma}$ denotes a multivariate Gaussian distribution with mean $\boldsymbol{\mu}$, variance $\boldsymbol{\Sigma}$, and $\mathscr{I}$ is standard univariate normal cumulative distribution function (CDF). $\mathscr{I}^{-1}$ indicates the standard inverse transformation, which aims to obtain the intermediate variables following a univariate Gaussian distribution.

$$\begin{aligned} &C\left(P_1\left(x_{1,t}\right), \ldots, P_N\left(x_{N,t}\right)\right) \\ &= \Phi_{\boldsymbol{\mu},\boldsymbol{\Sigma}}\left(\mathscr{I}^{-1}\left(P_1\left(x_{1,t}\right)\right), \ldots, \mathscr{I}^{-1}\left(P_N\left(x_{N,t}\right)\right)\right). \end{aligned} \tag{4.9}$$

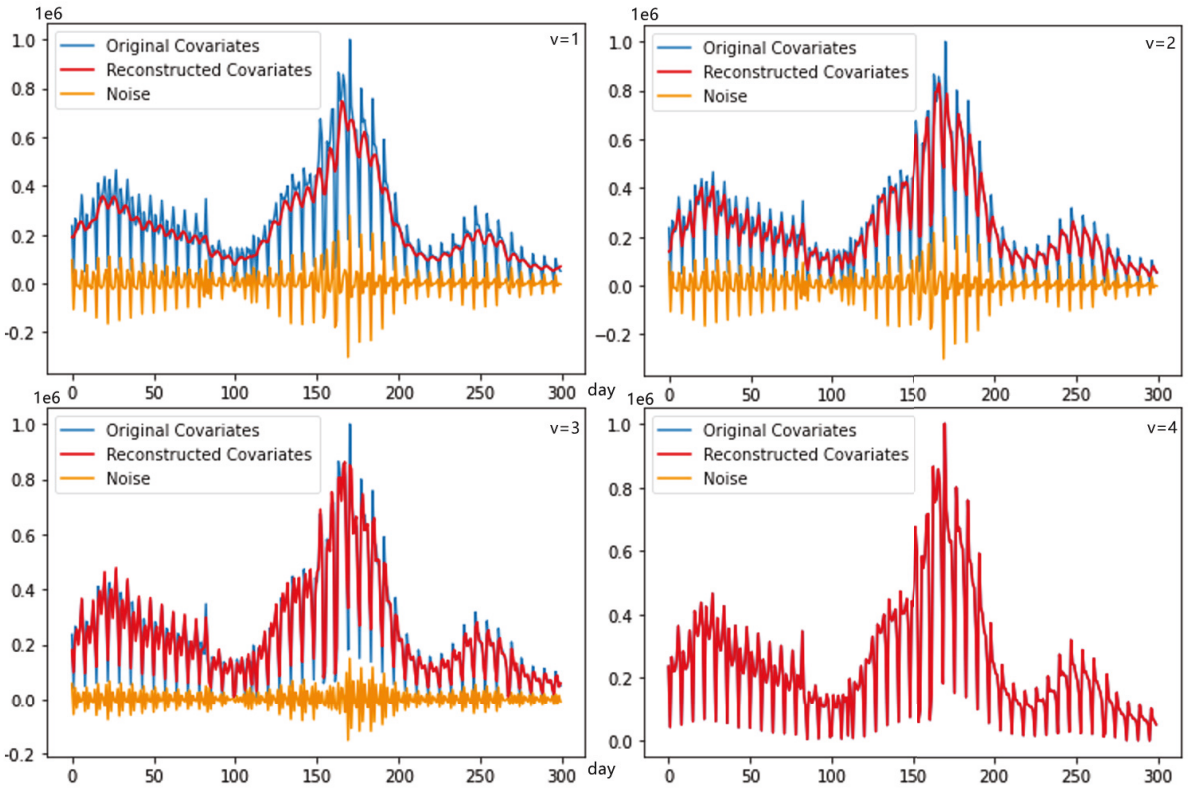


Figure 4.2: The SSA Module in DSCM. We visualize the Italian covariate (daily test number) for SSA decomposition and reconstruction as an example. The four graphs represent the different numbers v of decomposed components selected for reconstruction.

We then apply the same method described in [72] to estimate the empirical CDF of the marginal distributions P_i. This is a nonparametric, frequency-based approach to obtain the empirical distribution function P_i. Let $\mathcal{T}_i = \mathcal{I}^{-1} P_i(\cdot)$ represent this transformation, we can have the joint cumulative distribution,

$$\begin{aligned} &Q\left(x_{1,t},\ldots,x_{N,t}\right) \\ &= \Phi_{\boldsymbol{\mu},\boldsymbol{\Sigma}}\left(\mathcal{T}_1\left(x_{1,t}\right),\mathcal{T}_2\left(x_{2,t}\right),\ldots,\mathcal{T}_N\left(x_{N,t}\right)\right). \end{aligned} \tag{4.10}$$

Accordingly, let the vector $\mathbf{x}_t$ represent target variables $\{x_{i,t}\}_{i=1}^{N}$, we can obtain joint emission distribution in Eq. (4.11), where the variables from the originally observed time series follow a temporal multivariate Gaussian probability distribution $\mathcal{N}$ at time t. Moreover, the corresponding likelihood function can be derived [15],

$$p\left(\mathbf{x}_t\right) = \mathcal{N}\left(\left[\mathcal{T}_1\left(x_{1,t}\right),\mathcal{T}_2\left(x_{2,t}\right),\ldots,\mathcal{T}_N\left(x_{N,t}\right)\right]^T\right). \tag{4.11}$$

4.1.5 The Intra-MTS Coupling Module

The intra-MTS coupling module of DSCM captures the intra-couplings cross time length T_x within a single set of multivariate time series.

Specifically, $\mathbf{h}_{i,t}^1$ captures the hidden states generated by the first LSTM layer in Figure 4.1,

$$\mathbf{h}_{i,t}^1 = \mathrm{LSTM}_1\left(\mathbf{x}_{i,t}, \hat{\mathbf{f}}_t^i, \mathbf{h}_{i,t-1}^1\right), \tag{4.12}$$

which contains the nonlinearly transformed information of both the target variables $\mathbf{x}_{i,t}$ and the reconstructed covariates $\hat{\mathbf{f}}_t^i$ at time t. Then, $\mathbf{h}_{i,t}^1$ is fed into the intra-MTS coupling module, which consists of a dense layer and a softmax layer to compute the vector $\boldsymbol{\alpha}_t$,

$$\boldsymbol{\alpha}_t = \mathrm{Softmax}\left(\mathrm{Dense}\left(\mathbf{h}_i^1\right)\right). \tag{4.13}$$

As shown in Figure 4.1, each element $\alpha_{t,t'}$ in vector $\boldsymbol{\alpha}_t$ represents the weights of the corresponding hidden states indexed by t'. The sum of $\alpha_{t,t'}$ is equal to 1. It is noted that each hidden state $\mathbf{h}_{i,t'}$ has its time-step related weight $\alpha_{t,t'}$. We then use the following method to combine the weights with the hidden states to obtain $\mathbf{h}_{i,t}^2$,

$$\mathbf{h}_{i,t}^2 = \sum_{t'=1}^{T_x} \alpha_{t,t'} \mathbf{h}_{i,t'}^1. \tag{4.14}$$

Further, Eq. (4.15) uses the current hidden states $\mathbf{h}_{i,t}^2$ and the previous hidden states $\mathbf{h}_{i,t-1}^3$ to obtain the hidden states $\mathbf{h}_{i,t}^3$. $\mathbf{h}_{i,t}^3$ will subsequently map to the parameters of the previously established joint probability distribution via projection layers,

$$\mathbf{h}_{i,t}^3 = \mathrm{LSTM}_2\left(\mathbf{h}_{i,t}^2, \mathbf{h}_{i,t-1}^3\right), \tag{4.15}$$

where $\mathrm{LSTM}_2(\cdot)$ represents the second LSTM layer of the DSCM in Figure 4.1.

4.1.6 Deep Probabilistic Forecasting

As DSCM essentially follows the Seq2Seq architecture and we further assume the target variables are i.i.d. drawn from an underlying joint distribution. Considering the reconstructed covariates $\hat{\mathbf{F}}$ from SSA of the volatile covariates $\mathbf{F}$, Eq. (4.1) can be written as,

$$p\left(\mathbf{x}_{t_0+1:t_0+m} \mid \mathbf{x}_{1:t_0}, \hat{\mathbf{F}}_{1:t_0+m}, \Theta\right) = \prod_{t=t_0}^{t_0+m} p\left(\mathbf{x}_t \mid \mathbf{x}_{1:t}, \hat{\mathbf{F}}_{1:t}, \Theta\right). \tag{4.16}$$

As stated in the previous studies [12, 13, 15], an approach to tackle Eq. (4.16) is to use the maximum likelihood estimation for solutions. Accordingly, it assumes the observations are fixed and the parameters of the likelihoods are determined by functions $\mu\left(\mathbf{h}_{i,t}^3\right)$ and $\Sigma\left(\mathbf{h}_{i,t}^3\right)$, which denote the low-rank projection operations described in [15]. Since the hidden states $\mathbf{h}_{i,t}^3$ generated by the neural networks containing the information of $\mathbf{x}_{1:t}$ and $\mathbf{F}_{1:t}$ as shown in Eq. (4.12-4.15), the global parameters Θ in Eq. (4.17) can be learned and updated by the gradient descent algorithm when the assumed likelihood loss function is tractable,

$$p\left(\mathbf{x}_{t_0+1:t_0+m} \mid \mathbf{x}_{1:t_0}, \hat{\mathbf{F}}_{1:t_0+m}, \Theta\right) = \prod_{t=t_0}^{t_0+m} p\left(\mathbf{x}_t \mid \mu\left(\mathbf{h}_{i,t}^3\right), \Sigma\left(\mathbf{h}_{i,t}^3\right), \Theta\right). \tag{4.17}$$

Consequently, DSCM obtains hidden representation $\mathbf{h}_{i,t}^3$ after multiple nonlinear transformations using the target variable and the reconstructed covariates by a neural network and an intra-MTS coupling module. Then, we use low-rank projection functions $\mu\left(\mathbf{h}_{i,t}^3\right)$ and $\Sigma\left(\mathbf{h}_{i,t}^3\right)$ to map the hidden representation $\mathbf{h}_{i,t}^3$ to the unknown parameters of the likelihood functions,

$$\begin{aligned} \Sigma\left(\mathbf{h}_{i,t}^3\right) &= \mathbf{D}_t + \mathbf{V}_t \mathbf{V}_t^T, \\ d_{i,t}\left(\mathbf{h}_{i,t}^3\right) &= \log\left(1 + \exp\left(W_d(\mathbf{h}_{i,t}^3\right)\right), \\ \mathbf{v}_{i,t}\left(\mathbf{h}_{i,t}^3\right) &= W_v \mathbf{h}_{i,t}^3, \\ \mu\left(\mathbf{h}_{i,t}^3\right) &= W_\mu^T \mathbf{h}_{i,t}^3. \end{aligned} \tag{4.18}$$

The low-rank projection functions are the same as that of DeepVAR. It uses the sum of a diagonal matrix $\mathbf{D}_t$ and a low-rank matrix $\mathbf{V}_t$ to replace the covariance matrix of the assumed joint distribution. Specifically, $d_{i,t}\left(\mathbf{h}_{i,t}^3\right)$ and $\mathbf{v}_{i,t}\left(\mathbf{h}_{i,t}^3\right)$ calculate each diagonal element of $\mathbf{D}_t$ and the elements of low-rank matrix $\mathbf{V}_t$, respectively. They will be used to calculate the $\Sigma\left(\mathbf{h}_{i,t}^3\right)$ subsequently.

4.1.7 Likelihood Loss

In Section 4.1.4, we obtain the joint probability distribution of the target variables through the marginal distribution transformations and Gaussian Copula assumptions. The likelihood function of this joint probability distribution is approachable [15] and meets the requirements for gradient descent optimization. Hence, it can be directly used as an objective function of DSCM. As we need to implement the maximum likelihood

estimation to update the parameters of DSCM, the objective function is in the form of a negative log-likelihood,

$$\mathscr{L} = -\sum_{t=t_0}^{t_0+m} \log p\left(\mathbf{x}_t \mid \mu\left(\mathbf{h}_{i,t}^3\right), \Sigma\left(\mathbf{h}_{i,t}^3\right)\right). \tag{4.19}$$

The prediction process of DSCM follows the DeepAR and DeepVAR architecture[12, 15]. Specifically, we use $\mathbf{x}_t$ ($t \leq t_0$) as input, and the predicted value $\mathbf{x}_t$ ($t > t_0$) at the next time step can be obtained via ancestral sampling from the generated distribution. After repeating this process, we can get a set of predicted values at the prediction time length. Then, we use these samples for evaluation.

4.2 Experiments

To verify the effect of the proposed DSCM method on CVMTS data, we perform experiments on COVID-19 data and compare it with the state-of-the-art multivariate time series methods.

4.2.1 Setup

4.2.1.1 Coupled and Volatile Multivariate Time Series Data

The considered COVID-19 case data consists of multiple target time series and covariates time series from 13 European countries. Specifically, the target time series (e.g., confirmed cases, deaths, and hospital patients) are acquired from the COVID-19 data repository of Johns Hopkins University between June 10th, 2020 and June 10th, 2022. The covariates (i.e., daily test number, reproduction number, vaccine rates, and the Stringency Index) time series are obtained from the GISAID initiative[1] and Github [2] for the same period.

We calculate the Pearson correlation coefficient [73] to analyze the inter-MTS couplings of COVID-19 cases data from 13 European countries. As shown in Figure 4.4, France, Italy, and Portugal show strong couplings with Pearson coefficients 0.87, 0.72, and 0.68, respectively. In addition, the most correlated covariates are the daily test number with a coefficient at 0.55. In this study, we leave the test on other time series data for future work as we are still struggling in constructing such multiple correlated multivariate time series with volatile external covariates like COVID-19.

[1]The data is available at covariants.org.

[2]The full dataset is available at: https://github.com/owid/covid-19-data/tree/master/public/data/vaccinations

4.2.1.2 Deep Probabilistic Models as Baselines

The considered deep probabilistic models in the experiment include DeepAR, Deep State Space Model (DSM), Deep Factor (DF), DeepVAR, and GPVAR, which are from MXNet [65] and Gluon Time Series Toolkit [66]. The experiments run via the same CPU. Let m be the prediction length (window size), and all models initially receive time series data with t_0 length to tune hyperparameters using a back-test method. We test two tasks with window size m set to 10 and 20, respectively. After the optimal hyperparameters have been identified, all models will be assessed on the prediction results from the time step t_0+1 to t_0+m. We then calculate metrics for each of these training runs using 100 samples from the validation set.

4.2.1.3 Evaluation Metrics

To evaluate the performance of DSCM and the other deep probabilistic models, the adopted metrics include the continuous ranked probability score (CRPS), weighted quantile loss (WQL), and root mean square error (RMSE). To test the comprehensive performance of DSCM at various forecasting lengths, we calculate these metrics for 10-day prediction and 20-day prediction, respectively.

CRPS is a primary performance metric for probabilistic predictions. It is a measurement of the discrepancy between the empirical cumulative distribution function (CDF) of the observation and the predicted CDF. The continuous ranked probability score is defined as:

$$\mathrm{CRPS}(P,\mathbf{x}) = \int_{-\infty}^{\infty} (P(y) - \mathbf{1}\{y \geq x_{i,t}\})^2 \, \mathrm{d}y, \tag{4.20}$$

where P denotes the cumulative distribution function. $\mathbf{1}\{y \geq x_{i,t}\}$ and $x_{i,t}$ represent the indicator function and observations.

The Energy Score [74] is a metric for measuring multivariate probability distribution. A smaller Energy Score indicates a more accurate predicted distribution. Unlike CRPS, the Energy Score is more sensitive to the correlation with the outgoing variables. The Energy Score is defined as:

$$\begin{aligned} \mathrm{Energy}(\{\tilde{x}_t, x_t\}) = & \mathop{\mathbb{E}}_{\tilde{x}_t} \|\tilde{x}_t - x_t\|_{Fro}^{\beta} \\ & - \frac{1}{2} \mathop{\mathbb{E}}_{\tilde{x}_t, \tilde{x}_t'} \left\|\tilde{x}_t - \tilde{x}_t'\right\|_{Fro}^{\beta}, \end{aligned} \tag{4.21}$$

where x_t denotes the observed value in the form of a matrix, $\tilde{x}_t$ and $\tilde{x}_t'$ are two independent variables in matrix forms sampled from a predicted probability distribution, β represents a parameter set to 1, and $\|.|_{For}\|$ denotes the Frobenius matrix norm.

Table 4.1: Prediction results for the case number in the next 10 days. Bold values represent relatively better performance.

Metrics	DSCM	DeepAR	DeepVAR	DSM	GPVAR	DF
CRPS	**0.2609**	0.3785	0.3147	0.4210	0.3019	0.4743
WQL[0.1]	0.1221	0.1617	0.1276	0.3395	**0.1155**	0.2374
WQL[0.5]	0.3643	0.5093	0.3689	0.4645	**0.3323**	0.4818
WQL[0.9]	**0.2417**	0.3573	0.3728	0.4684	0.4425	0.7027
RMSE	**0.0944**	0.1138	0.1051	0.1097	0.0948	0.1286

Table 4.2: Prediction results for the case number in the next 20 days. Bold values represent relatively better performance.

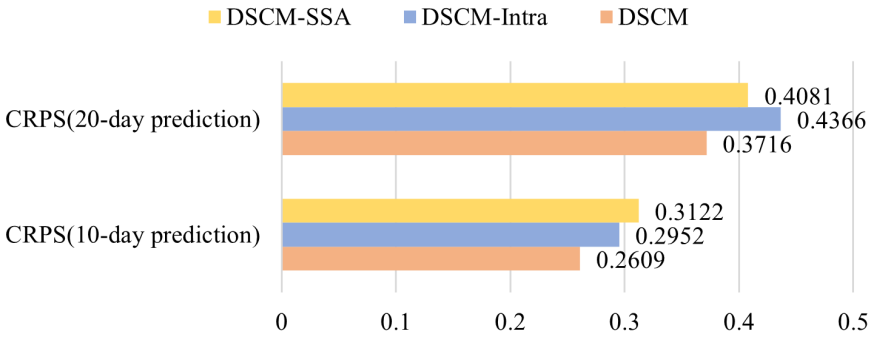
Metrics	DSCM	DeepAR	DeepVAR	DSM	GPVAR	DF
CRPS	**0.3716**	0.4285	0.4254	0.6274	0.5466	0.5003
WQL[0.1]	0.2126	**0.1276**	0.2182	0.4670	0.2174	0.4104
WQL[0.5]	0.5177	**0.3689**	0.5402	0.6856	0.6540	0.5866
WQL[0.9]	**0.2422**	0.5124	0.3679	0.7324	0.6713	0.5019
RMSE	**0.1384**	0.1587	0.1520	0.1719	0.1566	0.1433

Weighted quantile loss measures the accuracy of probabilistic forecasting at various quantiles, which can be applied to the predicted probability distribution. RMSE measures the accuracy of the probability distribution using the median value of the predicted samples. Both WQL and RMSE are defined in Chapter 3.

4.2.2 Comparison Results

We compare DSCM with the SOTA deep probabilistic models, including DeepAR, Deep State Space Model (DSM), Deep Factor (DF), DeepVAR, and GPVAR, to verify their capable of generating probability distributions. Point estimation models such as ARIMA, ETS, RNNs and other DNN-based models are not considered in the experiments because they cannot conduct probabilistic forecasting and evaluate the probabilistic performance. We introduce DeepAR, DSM, and DF in Chapter 3. The details of the DeepVAR, and GPVAR are shown below,

- DeepVAR [15] uses a Gaussian Copula function to represent the joint probability distribution of the multivariate time series. The parameters are solved by a gradient descent algorithm.



	CRPS(10-day prediction)	CRPS(20-day prediction)
DSCM-SSA	0.3122	0.4081
DSCM-Intra	0.2952	0.4366
DSCM	0.2609	0.3716

Figure 4.3: Comparison of the CRPS score of 10-day vs 20-day prediction. The figure shows that the overall CRPS is more accurate for a 10-day forecast than a 20-day forecast. In addition, the DSCM model performs the best in both 10-day and 20-day predictions, with CRPS values of 0.2609 and 0.3716, respectively. This indicates that short-term prediction can be better than long-term prediction

- GPVAR [15] uses a low-rank approximation to eliminate issues caused by a large number of parameters in a full covariate matrix. The Gaussian process is implemented for parameter optimization.

Table 4.1 and 4.2 report the comparison results. The results demonstrate that DSCM outperforms the SOTA models in the probabilistic forecasting tasks at different time lengths. The overall prediction performance is gauged by CRPS and RMSE, which represents the accuracy of the overall predicted probability distribution and the error between the predicted mean and test data, respectively. Specifically, for the 10-day prediction as shown in Table 4.1, DSCM improves the performance over the SOTA models by at least 10.12% and 10.5% over the SOTA models in terms of CRPS and RMSE respectively. For the 20-day prediction in Table 4.2, DSCM outperforms the other deep probabilistic models by at least 10.12% and 10.5% in terms of CRPS and RMSE, respectively. We note that all models perform better on the 10-day prediction than the 20-day prediction. This demonstrates that it is more challenging in long-term forecasting than short-term forecasting for such non-stationary and volatile data. Some deep probabilistic models, such as GPVAR and DeepAR, achieve the best scores in terms of WQL[0.1] and WQL[0.5]. However, these metrics do not provide sufficient evidence for the overall performance. Overall, DSCM achieves the best WQL[0.1] and WQL[0.5]

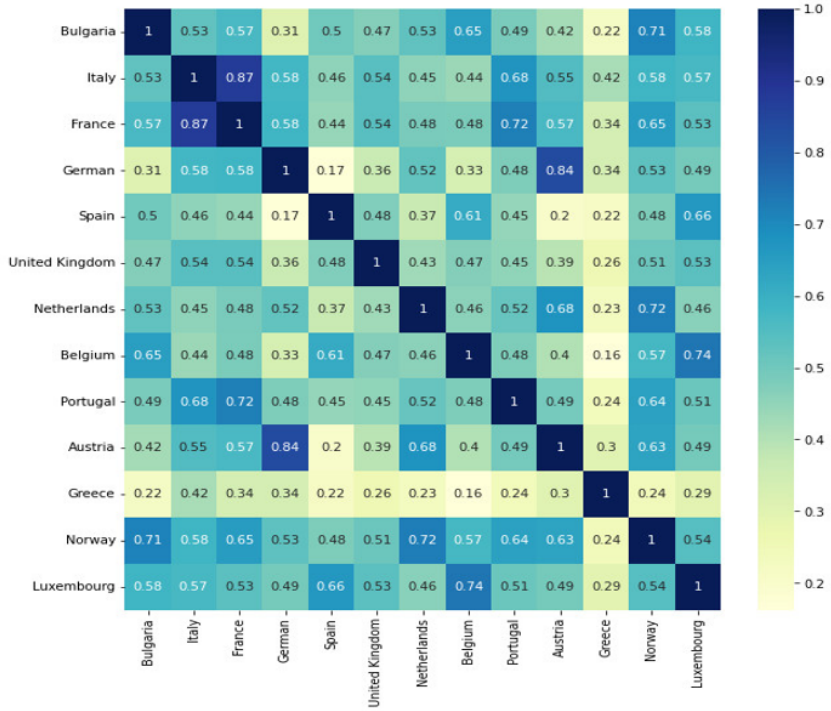


Figure 4.4: Pearson correlation coefficient of COVID-19 case data from 13 European countries. France, Italy, and Portugal show strong couplings with Pearson coefficients 0.87, 0.72, and 0.68 between them, respectively.

Table 4.3: Ablation studies for 10-day prediction of case number. Bold values represent relatively better performance.

Metrics	DSCM	DSCM-Intra	DSCM-SSA
CRPS	**0.2609**	0.2952	0.3122
WQL[0.1]	**0.1221**	0.1225	0.1351
WQL[0.5]	0.3643	0.3760	**0.3640**
WQL[0.9]	**0.2417**	0.3304	0.2596
RMSE	**0.0944**	0.1051	0.1018

and outperforms the other probabilistic models in other metrics, such as CRPS and RMSE. These results demonstrate that the proposed model is more robust to the CVMTS probabilistic prediction task.

4.2.3 Ablation Study

For the ablation study, three DSCM variants are adopted to evaluate the effectiveness of the respective modules, including the SSA and intra-MTS coupling modules.

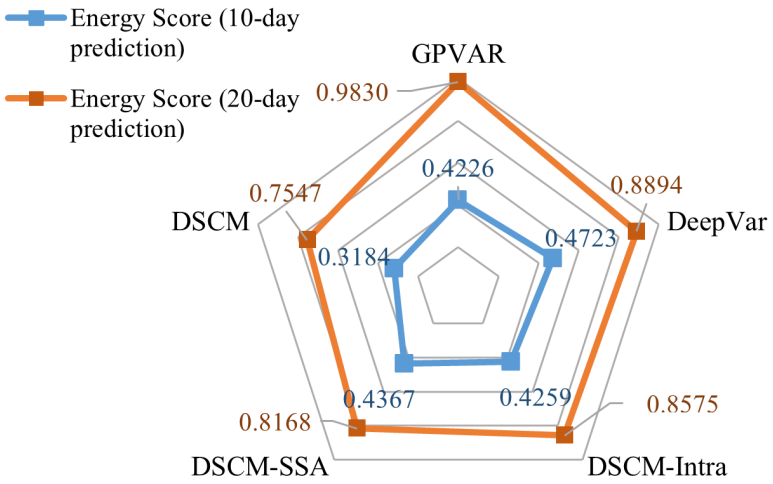


Figure 4.5: Comparison of the Energy score of 10-day vs 20-day prediction. The Figure displays Energy scores for five models, namely DSCM, DSCM-Intra, DSCM-SSA, GPVAR, and DeepVAR. The Energy score is a metric that can only assess multivariate probability distributions, not univariate ones. Therefore, we computed the Energy score for methods that can generate such probability distributions.

Table 4.4: Ablation studies for 20-day prediction of the case number. Bold values represent relatively better performance.

Metrics	DSCM	DSCM-Inta	DSCM-SSA
CRPS	**0.3716**	0.4366	0.4081
WQL[0.1]	0.2126	0.2351	**0.2003**
WQL[0.5]	0.5177	0.5453	**0.4913**
WQL[0.9]	**0.2422**	0.2629	0.2537
RMSE	**0.1384**	0.1471	0.1378

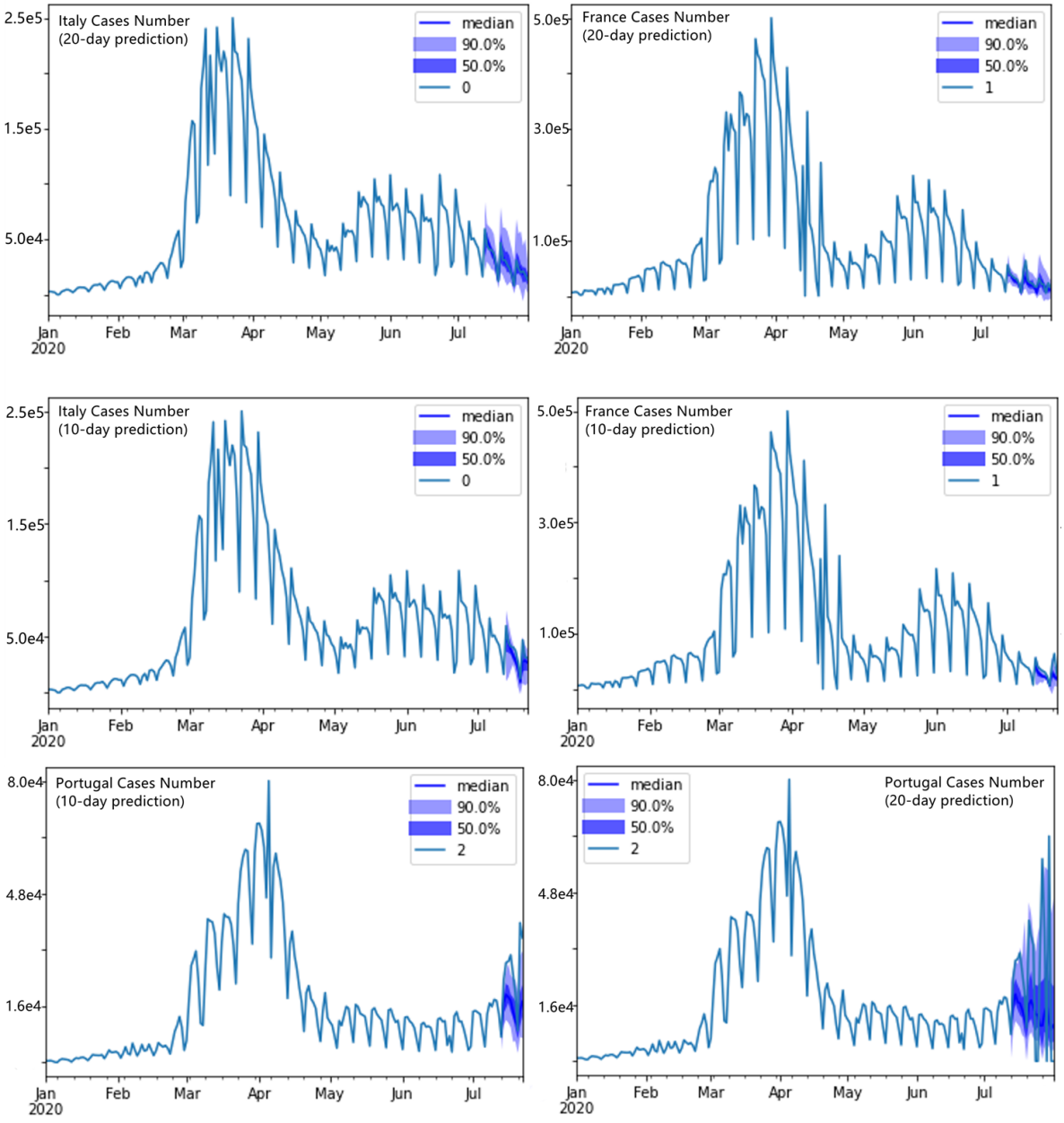


Figure 4.6: Visualization of representative prediction results showing the performance of DSCM for 10-day and 20-day prediction. As the forecasting results are multidimensional and there are 100 samples per time point, it is not feasible to plot all the model results together for comparison.

DSCM-SSA refers to DSCM with the singular spectral analysis (SSA) module only. DSCM-Intra incorporates an intra-MTS coupling module, explicitly capturing temporal couplings within a single set of multivariate time series. Finally, DSCM incorporates both the intra-MTS coupling and SSA modules, offering a comprehensive approach for exploring inter- and intra-MTS couplings and handling volatile covariates in CVMTS forecasting.

Tables 4.3 and 4.4 report the performance of the ablation experiments for the 10-day and 20-day prediction tasks, respectively. It shows that the DSCM with both intra-MTS coupling and SSA modules performs better than other DSCM variants on an overall basis. The results in terms of WQL[0.1] and WQL[0.5] of DSCM are higher than that of DSCM-SSA for the task of 20-day prediction. However, WQL only represents the accuracy of a certain quantile of the predicted samples, which is limited to gauging the whole forecasting results. The underperformance of the proposed DSCM with both SSA and Intra modules compared to DSCM-SSA and DSCM-Intra in terms of WQL[0.1] and WQL[0.5] could be attributed to the increased complexity introduced by the combination of these modules. The additional complexity may lead to overfitting and potential conflicts or redundancy in the information extracted, resulting in less accurate predictions for extreme quantiles. Fine-tuning the model's design and parameters and optimizing the interaction between the modules could help address this underperformance.

The CRPS score reflects the overall accuracy of the probabilistic prediction, which illustrates that DSCM outperforms DSCM-SSA. Specifically, the CRPS score of DSCM achieves an improvement of 3.4% and 5.1% over DSCM-Intra and DSCM-SSA for the 10-day prediction, respectively. Moreover, the CRPS score of DSCM improves over the variants DSCM-Intra and DSCM-SSA in predicting 20-day case numbers by 6.5% and 3.6%, respectively. In Figure 4.3, we note that all model variants perform better on the 10-day prediction task than on the 20-day task. In addition, for the 20-day prediction task, the CRPS score of DSCM-SSA is 2.8% better than the DSCM-Intra model. This may indicate that the trend features of covariates extracted by the SSA module might contribute more to long-term forecasting than the intra-MTS coupling module. Figure 4.5 compares the Energy scores of five models including DSCM, DSCM-Intra, DSCM-SSA, GPVAR, and DeepVAR. Since this metric is merely suitable for measuring multivariate probability distributions and not univariate probability distributions, we calculate the Energy score of the approaches that can generate such probability distributions. Figure 4.5 reports that the Energy scores for the 10-day prediction are generally better than the 20-day prediction. Moreover, DSCM achieves the best values of 0.3184 and 0.3547 in

the task of 10-day prediction and 10-day prediction, respectively. The Energy score also demonstrates the superiority of our approach in multivariate probabilistic forecasting.

In Figure 4.3, we note that all model variants perform better on the 10-day prediction task than on the 20-day task. In addition, for the 20-day prediction task, the CRPS score of DSCM-SSA is 2.8% better than the DSCM-Intra model. This may indicate that the trend features of covariates extracted by the SSA module might contribute more to long-term forecasting than the intra-MTS coupling module.

4.3 Summary

Coupled and Volatile Multivariate Time Series (CVMTS) widely exist in many industries. Cross-country COVID-19 modeling is a typical application of CVMTS, where multiple COVID-19 time series contain high levels of volatility caused by various underlying factors. To explore the inter- and intra-MTS couplings and handle volatile covariates for CVMTS forecasting, we propose Deep Spectral Copula Mechanisms (DSCM).

DSCM includes a singular spectral analysis module to extract information from volatile variables. Additionally, an intra-MTS coupling module explicitly captures the temporal couplings inside a single set of multivariate time series, and Gaussian copula transformation is applied to convert target variables into joint probability distributions for exploring inter-MTS couplings between multivariate time series. Numerous experiments on COVID-19 time-series data from 13 countries show that DSCM achieves superior performance in predicting the probability distribution of CVMTS. Furthermore, ablation experiments demonstrate the effectiveness of different designed modules in DSCM.

DSCM proposed for handling Coupled and volatile multivariate time series, particularly in the context of cross-country COVID-19 modeling, demonstrates promising capabilities. However, it is important to consider certain limitations associated with the approach. One notable limitation is its reliance on supervised learning algorithms, which may have inherent constraints in capturing complex relationships and dynamics within the data. These algorithms heavily depend on the quality, representativeness, and quantity of the training data, which can potentially lead to issues of overfitting or underfitting. As a result, the model's ability to generalize to unseen data or different contexts may be compromised.

Furthermore, the generalization ability of DSCM might be a concern when applying the model to diverse scenarios beyond COVID-19 time series. While the model has demon-

strated superior performance in predicting the probability distribution of CVMTS using COVID-19 data from multiple countries, the underlying assumptions and characteristics specific to COVID-19 might not hold true for other types of multivariate time series. Therefore, it is necessary to thoroughly validate and assess the model's performance in various domains to ensure its broader applicability.

In order to mitigate these limitations, future iterations of DSCM should consider integrating domain knowledge and expertise into the modeling process. By leveraging a hybrid approach that combines data-driven techniques with domain-driven insights, the model can enhance its robustness and generalizability. This integration would enable a more comprehensive understanding of the complex interdependencies within CVMTS and improve the accuracy and practicality of the model's predictions.

CHAPTER 5

CONCLUSION AND PLAN

5.1 Conclusion

Deep probabilistic modeling for coupled multivariate time series has been significant in various fields such as finance, economics, weather forecasting, and COVID-19 transmission prediction. The challenges come from modeling the interactions within and between coupled MTSs, and the uncertainty, heterogeneity, and dynamics. COVID-19 cases and external factors present an imperative use case for this type of research issue. Deep probabilistic modeling is thus essential to address the complexity of the problem and to model the uncertainty of the predictions in reality.

The proposed methods, MTSNet and DSCM, offer innovative solutions to these problems by incorporating advanced techniques. Particularly, MTSNet shows promising potential in the case study of forecasting COVID-19 confirmed cases and ICU patient numbers. MTSNet serves as a global model to jointly predict them in six countries by involving temporal external factors including virus mutations, vaccination, government interventions, and infectivity. MTSNet stacks LSTM vertically and horizontally, improving prediction accuracy and robustness by residual connection and variational zoneout. Experimental results demonstrate the accuracy and effectiveness of our method in the COVID-19 multivariate time series modeling task. On the other hand, DSCM incorporates the Singular Spectral Analysis (SSA) module to reduce the volatilities of multiple covariates and utilizes an intra-MTS Coupling module to explicitly model the temporal correlation within a single set of multivariate time series. To establish

inter-MTS correlations across multiple multivariate time series, it transforms the target variables into joint probability distributions via Gaussian copula transformation. The massive experiments on COVID-19 time-series data from multiple countries indicate the superiority of DSCM over the deep probabilistic SOTA approaches (DeepAR, DeepVAR, Deep State Space Model, Deep Factor Model, and GPVAR) in modeling CVMTS.

Overall, in the broader context of deep probabilistic modeling for coupled multivariate time series in COVID-19 prediction, Chapter 3 introduces MTSNet while Chapter 4 introduces DSCM, both of which offer encouraging resolutions to the challenges encountered in this domain. These contributions collectively showcase the efficacy of the proposed algorithms in effectively addressing the intricacies associated with coupled multivariate time series. Moreover, the implications of the findings presented in this thesis are anticipated to have substantial significance across diverse domains that incorporate the analysis of coupled multivariate time series.

5.2 Future Work

There are several possible directions for future work. First, all experiments in this thesis focus on the COVID-19 pandemic, but the proposed methods are applicable to a wide range of coupled multivariate time series. It is worth investigating their performance on other datasets from different domains, such as finance, energy, and transportation.

Second, both MTSNet and DSCM follow Seq2Seq, which may cause accumulated errors for long-term multivariate time series prediction. Specifically, the next time step will use the value sampled from the previously generated multivariate probabilistic distribution. Repeating this process will increase the accumulated errors. It is worth exploring alternative architectures or optimization techniques to reduce the error accumulation for long-term predictions. Some possible directions include introducing a feedback mechanism to adjust the predicted values, incorporating external information or expert knowledge to improve the model's performance, or designing a hierarchical structure that decomposes the CVMTS into sub-series for better prediction accuracy. These potential directions could enhance the robustness and accuracy of the proposed methods, and extend the scope of the research to address long-term multivariate time series prediction issues in a wider range of applications.

The last future research direction is from the perspective of epidemic modeling. although COVID-19 is no longer receiving much attention as before, infectious disease modeling is still important because we may meet new or unknown outbreaks in the

future. So we may focus on modeling epidemic transmission at an earlier stage with limited data and labels. Some possible directions include exploring an unsupervised learning algorithm for a small dataset of an unknown disease based on the previous outbreak and incorporating the domain knowledge of epidemiology into deep learning algorithms instead of pure data-driven approaches.

APPENDIX

Table A.1: Notation of Variables and Parameters

Symbol:	
$x_{i,t}$	original target time series value from source i at time t
$x'_{i,t}$	normalized target time series value from source i at time t
$\mathbf{x}_i$	target time series vector from source i in the time range $[1,2,...,t]$
$\mathbf{x}'_i$	normalized target time series vector from source i in the time range $[1,2,...,t]$
$\mathbf{x}_t$	target time series vector from all sources at time t
$\mathbf{x}'_t$	normalized target time series vector from all sources at time t
$\mathbb{F}_{i,j,t}$	original value of factor j at time t from source i
$\mathbb{F}'_{i,j,t}$	normalized value of factor j at time t from source i
$\mathbb{F}_{i,t}$	original vector of all factors at time t from source i
$\mathbb{F}'_{i,t}$	normalized vector of all factors at time t from source i
$\mathbb{F}_t$	original second-order tensor of all factors and sources at time t
$\mathbb{F}'_t$	normalized original second-order tensor of all factors and sources at time t
$\mathbf{F}_t$	original vector of volatile covariates from all sources at time t
$\mathbf{F}$	original tensor of volatile covariates
$\hat{\mathbf{F}}$	reconstructed tensor of volatile covariates
f_t^i	value volatile covariates from source i at time t
$\mathbf{U} \in \mathbb{R}^{K\times K}$,$\mathbf{V} \in \mathbb{R}^{L\times L}$	the matrices composed of the singular vectors
$\mathbf{H^i}$	trajectory matrices with a stacked Hankel structure
σ_i	singular values
$\mathbf{u}_i$, $\mathbf{v}_i$	eigentriple vectors
r	total number of decomposed components
$\mathbf{\Sigma} \in \mathbb{R}^{K\times L}$	rectangular matrix (the diagonal elements are all the singular values of trajectory matrices)
$\mathbf{G}^j$ $(j = 1,\ldots,r)$	linearly independent matrices
$\mathbf{h}_{i,t}^l$	hidden states at time t in the layer$l$
$\mathbf{c}_{i,t}^l$	memory states at time t in the layer$l$
$\mathbf{z}_{i,t}^l$	hidden representations at time t in the layer$l$
$\mathbf{I}_t$	input gate of LSTM cell
$\mathbf{f}_t$	forget gate of LSTM cell
$\mathbf{o}_t$	output gate of LSTM cell
$\mathbf{g}_t$	input modulation of LSTM cell
$\mathbf{W}_I, \mathbf{W}_f, \mathbf{W}_o, \mathbf{W}_g$	learnable parameter matrix
$\mathbf{U}_I, \mathbf{U}_f, \mathbf{U}_o, \mathbf{U}_g$	learnable parameter matrix
$\mathbf{s}_x$	random initialized masks applied on the input of LSTM
$\mathbf{s}_h$	random initialized masks applied on the hidden states of LSTM
Θ	global parameters of networks
Θ_h	learnable parameters of stacked LSTM layers
Θ_k	learnable parameters of projection layers

Table A.2: Notation of Functions

Symbol:	
$\ell_L(.)$	likelihood function of Laplace distribution
$\ell_G(.)$	likelihood function of Gaussian distribution
$\ell_S(.)$	likelihood function of Student's t distribution
$sigmoid(.)$	sigmoid activation function
$tanh(.)$	hyperbolic tangent function
$Q(.)$	multivariate probability distribution
$C(.)$	copula function
$P(.)$	univariate marginal distribution
$\Phi_{\boldsymbol{\mu},\Sigma}(.)$	multivariate Gaussian distribution (CDF) with mean $\boldsymbol{\mu}$, variance $\boldsymbol{\Sigma}$
$\mathscr{I}(.)$	standard univariate normal cumulative distribution function (CDF)
$\mathscr{I}^{-1}(.)$	standard inverse transformation
$\mathscr{N}(.)$	multivariate Gaussian distribution(PDF)
Softmax(.)	softmax layer operation

Table A.3: Abbreviations

MTSs	multivariate time series
CVMTS	coupled and volatile multivariate time series
MTSNet	deep probabilistic cross-MTS networks
DSCM	deep spectral copula mechanisms
LSTM	long short-term memory networks
RNN	recurrent neural networks
CNN	convolutional neural network
SSA	singular spectral analysis
DSM	deep state space model
DF	deep factor model
CDF	nonpharmaceutical Interventions
PDF	nonpharmaceutical Interventions
CRPS	continuous ranked probability score
ND	normalized deviation
WQL	weighted quantile loss

BIBLIOGRAPHY

[1] James Douglas Hamilton.
Time series analysis.
Princeton university press, 2020.

[2] Bryan Lim and Stefan Zohren.
Time-series forecasting with deep learning: a survey.
Philosophical Transactions of the Royal Society A, 379(2194):20200209, 2021.

[3] Andrea Silvestrini and David Veredas.
Temporal aggregation of univariate and multivariate time series models: a survey.
Journal of Economic Surveys, 22(3):458–497, 2008.

[4] Siyuan Ren, Bin Guo, Ke Li, Qianru Wang, Zhiwen Yu, and Longbing Cao.
Coupledmuts: Coupled multivariate utility time-series representation and prediction.
IEEE Internet of Things Journal, 9(22):22972–22982, 2022.

[5] Demetris Koutsoyiannis.
Coupling stochastic models of different timescales.
Water Resources Research, 37(2):379–391, 2001.

[6] Iman Rahimi, Fang Chen, and Amir H Gandomi.
A review on COVID-19 forecasting models.
Neural Computing and Applications, pages 1–11, 2021.

[7] Khalid Raza.
Artificial intelligence against COVID-19: A meta-analysis of current research.
Big data analytics and artificial intelligence against COVID-19: Innovation vision and approach, pages 165–176, 2020.

[8] Jakub Nowotarski and Rafał Weron.

Recent advances in electricity price forecasting: A review of probabilistic forecasting.
Renewable and Sustainable Energy Reviews, 81:1548–1568, 2018.

[9] Longbing Cao and Qing Liu.
How control and relaxation interventions and virus mutations influence the resurgence of COVID-19.
medRxiv, 2021.

[10] Longbing Cao.
Coupling learning of complex interactions.
Information Processing & Management, 51(2):167–186, 2015.

[11] Hemant Bherwani, Saima Anjum, Suman Kumar, Sneha Gautam, Ankit Gupta, Himanshu Kumbhare, Avneesh Anshul, and Rakesh Kumar.
Understanding COVID-19 transmission through bayesian probabilistic modeling and GIS-based voronoi approach: a policy perspective.
Environment, Development and Sustainability, 23(4):5846–5864, 2021.

[12] David Salinas, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski.
Deepar: Probabilistic forecasting with autoregressive recurrent networks.
International Journal of Forecasting, 36(3):1181–1191, 2020.

[13] Syama Sundar Rangapuram, Matthias W Seeger, Jan Gasthaus, Lorenzo Stella, Yuyang Wang, and Tim Januschowski.
Deep state space models for time series forecasting.
Advances in neural information processing systems, 31, 2018.

[14] Yuyang Wang, Alex Smola, Danielle Maddix, Jan Gasthaus, Dean Foster, and Tim Januschowski.
Deep factors for forecasting.
In *International conference on machine learning*, pages 6607–6617. PMLR, 2019.

[15] David Salinas, Michael Bohlke-Schneider, Laurent Callot, Roberto Medico, and Jan Gasthaus.
High-dimensional multivariate forecasting with low-rank gaussian copula processes.
Advances in neural information processing systems, 32, 2019.

[16] Daniel Svozil, Vladimir Kvasnicka, and Jiri Pospichal.

Introduction to multi-layer feed-forward neural networks.
Chemometrics and intelligent laboratory systems, 39(1):43–62, 1997.

[17] Larry Medsker and Lakhmi C Jain.
Recurrent neural networks: design and applications.
CRC press, 1999.

[18] Sepp Hochreiter and Jürgen Schmidhuber.
Long short-term memory.
Neural computation, 9(8):1735–1780, 1997.

[19] Madini O Alassafi, Mutasem Jarrah, and Reem Alotaibi.
Time series predicting of COVID-19 based on deep learning.
Neurocomputing, 468:335–344, 2022.

[20] Renato R Maaliw, Zoren P Mabunga, and Frederick T Villa.
Time-series forecasting of COVID-19 cases using stacked long short-term memory networks.
In *International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies*, pages 435–441. IEEE, 2021.

[21] Trisha Sinha, Titash Chowdhury, Rabindra Nath Shaw, and Ankush Ghosh.
Analysis and prediction of COVID-19 confirmed cases using deep learning models: A comparative study.
In *Advanced Computing and Intelligent Technologies*, pages 207–218. Springer, 2022.

[22] Rohitash Chandra, Ayush Jain, and Divyanshu Singh Chauhan.
Deep learning via LSTM models for COVID-19 infection forecasting in india.
PloS one, 17(1):e0262708, 2022.

[23] Abhishek Parashar and Yukti Mohan.
Forecasting Covid-19 cases in india using multivariate hybrid CNN-LSTM model.
In *International Conference on Innovative Computing and Communications*, pages 149–156. Springer, 2022.

[24] Jayanthi Devaraj, Rajvikram Madurai Elavarasan, Rishi Pugazhendhi, GM Shafiullah, Sumathi Ganesan, Ajay Kaarthic Jeysree, Irfan Ahmad Khan, and Eklas Hossain.

Forecasting of COVID-19 cases using deep learning models: Is it reliable and practically significant?
Results in Physics, 21:103817, 2021.

[25] Hossein Abbasimehr and Reza Paki.
Prediction of COVID-19 confirmed cases combining deep learning methods and bayesian optimization.
Chaos, Solitons & Fractals, 142:110511, 2021.

[26] Farah Shahid, Aneela Zameer, and Muhammad Muneeb.
Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM.
Chaos, Solitons & Fractals, 140:110212, 2020.

[27] Vinay Kumar Reddy Chimmula and Lei Zhang.
Time series forecasting of COVID-19 transmission in canada using LSTM networks.
Chaos, Solitons & Fractals, 135:109864, 2020.

[28] Peipei Wang, Xinqi Zheng, Gang Ai, Dongya Liu, and Bangren Zhu.
Time series prediction for the epidemic trends of COVID-19 using the improved LSTM deep learning method: Case studies in russia, peru and iran.
Chaos, Solitons & Fractals, 140:110214, 2020.

[29] Yogesh Gautam.
Transfer learning for COVID-19 cases and deaths forecast using LSTM network.
ISA transactions, 2021.

[30] İsmail Kırbaş, Adnan Sözen, Azim Doğuş Tuncer, and Fikret Şinasi Kazancıoğlu.
Comparative analysis and forecasting of COVID-19 cases in various european countries with ARIMA, NARNN and LSTM approaches.
Chaos, Solitons & Fractals, 138:110015, 2020.

[31] Mohamed Hawas.
Generated time-series prediction data of COVID-19's daily infections in brazil by using recurrent neural networks.
Data in brief, 32:106175, 2020.

[32] Saleh I Alzahrani, Ibrahim A Aljamaan, and Ebrahim A Al-Fakih.
Forecasting the spread of the COVID-19 pandemic in saudi arabia using ARIMA prediction model under current public health interventions.
Journal of infection and public health, 13(7):914–919, 2020.

[33] Zeynep Ceylan.
Estimation of COVID-19 prevalence in italy, spain, and france.
Science of The Total Environment, 729:138817, 2020.

[34] Roseline O Ogundokun, Adewale F Lukman, Golam BM Kibria, Joseph B Awotunde, and Benedita B Aladeitan.
Predictive modelling of COVID-19 confirmed cases in nigeria.
Infectious Disease Modelling, 5:543–548, 2020.

[35] Matheus Henrique Dal Molin Ribeiro, Ramon Gomes da Silva, Viviana Cocco Mariani, and Leandro dos Santos Coelho.
Short-term forecasting COVID-19 cumulative confirmed cases: Perspectives for brazil.
Chaos, Solitons & Fractals, 135:109853, 2020.

[36] Anuradha Tomar and Neeraj Gupta.
Prediction for the spread of COVID-19 in india and effectiveness of preventive measures.
Science of The Total Environment, 728:138762, 2020.

[37] Miodrag Zivkovic, Nebojsa Bacanin, K Venkatachalam, Anand Nayyar, Aleksandar Djordjevic, Ivana Strumberger, and Fadi Al-Turjman.
COVID-19 cases prediction by using hybrid machine learning and beetle antennae search approach.
Sustainable Cities and Society, 66:102669, 2021.

[38] Chandrakanta Mahanty, Raghvendra Kumar, Brojo Kishore Mishra, D Jude Hemanth, Deepak Gupta, and Ashish Khanna.
Prediction of COVID-19 active cases using exponential and non-linear growth models.
Expert Systems, 39(3):e12648, 2022.

[39] Eunju Hwang.
Prediction intervals of the COVID-19 cases by HAR models with growth rates and vaccination rates in top eight affected countries: Bootstrap improvement.
Chaos, Solitons & Fractals, page 111789, 2022.

[40] Poshan Niraula, Jorge Mateu, and Somnath Chaudhuri.

A bayesian machine learning approach for spatio-temporal prediction of COVID-19 cases.
Stochastic Environmental Research and Risk Assessment, pages 1–19, 2022.

[41] James W Taylor and Kathryn S Taylor.
Combining probabilistic forecasts of COVID-19 mortality in the united states.
European Journal of Operational Research, 2021.

[42] Kathryn S Taylor and James W Taylor.
Interval forecasts of weekly incident and cumulative COVID-19 mortality in the united states: A comparison of combining methods.
PloS one, 17(3):e0266096, 2022.

[43] Padmaksha Roy, Shailik Sarkar, Subhodip Biswas, Fanglan Chen, Zhiqian Chen, Naren Ramakrishnan, and Chang-Tien Lu.
Deep diffusion-based forecasting of COVID-19 by incorporating network-level mobility information.
In *International Conference on Advances in Social Networks Analysis and Mining, Virtual Event*, pages 168–175, 2021.

[44] Longbing Cao and Qing Liu.
COVID-19 modeling: A review.
CoRR, 2021.

[45] Jianguo Chen, Kenli Li, Zhaolei Zhang, Keqin Li, and Philip S Yu.
A survey on applications of artificial intelligence in fighting against COVID-19.
ACM Computing Surveys (CSUR), 54(8):1–32, 2021.

[46] Thomas Hale, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster, Emily Cameron-Blake, Laura Hallas, Saptarshi Majumdar, et al.
A global panel database of pandemic policies (oxford covid-19 government response tracker).
Nature Human Behaviour, 5(4):529–538, 2021.

[47] Michael A Hamilton, Danielle Hamilton, Oluwatamilore Soneye, Olorunshola Ayeyemi, and Raed Jaradat.
An analysis of the impact of policies and political affiliation on racial disparities in covid-19 infections and deaths in the usa.

International Journal of Data Science and Analytics, 13(1):63–76, 2022.

[48] Rajiv Chowdhury, Kevin Heng, Md Shajedur Rahman Shawon, Gabriel Goh, Daisy Okonofua, Carolina Ochoa-Rosales, Valentina Gonzalez-Jaramillo, Abbas Bhuiya, Daniel Reidpath, Shamini Prathapan, et al.
Dynamic interventions to control COVID-19 pandemic: a multivariate prediction modelling study comparing 16 worldwide countries.
European journal of epidemiology, 35(5):389–399, 2020.

[49] Bowen Du, Zirong Zhao, Jiejie Zhao, Le Yu, Leilei Sun, and Weifeng Lv.
Modelling the epidemic dynamics of COVID-19 with consideration of human mobility.
International Journal of Data Science and Analytics, 12(4):369–382, 2021.

[50] Cornelia Ilin, Sébastien Annan-Phan, Xiao Hui Tai, Shikhar Mehra, Solomon Hsiang, and Joshua E Blumenstock.
Public mobility data enables COVID-19 forecasting and management at local and global scales.
Scientific reports, 11(1):1–11, 2021.

[51] Gary D Hachtel, John D Stack, and Jordan A Hachtel.
Forecasting and modeling of the COVID-19 pandemic in the usa with a timed intervention model.
Scientific Reports, 12(1):1–14, 2022.

[52] Eric Zivot and Jiahui Wang.
Vector autoregressive models for multivariate time series.
Modeling financial time series with S-PLUS, pages 385–429, 2006.

[53] Agus Suharsono, Auliya Aziza, and Wara Pramesti.
Comparison of vector autoregressive (var) and vector error correction models (vecm) for index of asean stock price.
In *AIP Conference Proceedings*, volume 1913, page 020032. AIP Publishing LLC, 2017.

[54] Hairong Song and Zhiyong Zhang.
Analyzing multiple multivariate time series data using multilevel dynamic factor models.
Multivariate Behavioral Research, 49(1):67–77, 2014.

[55] Longbing Cao and Wenfeng Hou.
How have global scientists responded to tackling COVID-19?
medRxiv, 2022.

[56] Longbing Cao.
Ai in combating the COVID-19 pandemic.
IEEE Intelligent Systems, 37(2):3–13, 2022.

[57] Ilya Sutskever, Oriol Vinyals, and Quoc V Le.
Sequence to sequence learning with neural networks.
Advances in neural information processing systems, 27, 2014.

[58] Rob J Hyndman and Yeasmin Khandakar.
Automatic time series forecasting: the forecast package for R.
Journal of statistical software, 27:1–22, 2008.

[59] Sergey Ioffe and Christian Szegedy.
Batch normalization: Accelerating deep network training by reducing internal covariate shift.
In *International conference on machine learning*, pages 448–456. PMLR, 2015.

[60] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, et al.
Google's neural machine translation system: Bridging the gap between human and machine translation.
CoRR, 2016.

[61] Jun Qi, Xu Liu, and Javier Tejedor.
Variational inference-based dropout in recurrent neural networks for slot filling in spoken language understanding.
CoRR, 2020.

[62] Yarin Gal and Zoubin Ghahramani.
Dropout as a bayesian approximation: Representing model uncertainty in deep learning.
In *international conference on machine learning*, pages 1050–1059. PMLR, 2016.

[63] Yarin Gal and Zoubin Ghahramani.
A theoretically grounded application of dropout in recurrent neural networks.
Advances in neural information processing systems, 29, 2016.

[64] Francisco Arroyo-Marioli, Francisco Bullano, Simas Kucinskas, and Carlos Rondón-Moreno.
Tracking R of COVID-19: A new real-time estimation using the kalman filter.
PloS one, 16(1):e0244474, 2021.

[65] Tianqi Chen, Mu Li, Yutian Li, Min Lin, Naiyan Wang, Minjie Wang, Tianjun Xiao, Bing Xu, Chiyuan Zhang, and Zheng Zhang.
Mxnet: A flexible and efficient machine learning library for heterogeneous distributed systems.
CoRR, 2015.

[66] Alexander Alexandrov, Konstantinos Benidis, Michael Bohlke-Schneider, Valentin Flunkert, Jan Gasthaus, Tim Januschowski, Danielle C Maddix, Syama Sundar Rangapuram, David Salinas, Jasper Schulz, et al.
GluonTS: Probabilistic and neural time series modeling in python.
J. Mach. Learn. Res., 21(116):1–6, 2020.

[67] Andrés Marino Álvarez-Meza, Carlos Daniel Acosta-Medina, and Germán Castellanos-Domínguez.
Automatic singular spectrum analysis for time-series decomposition.
In *European Symposium on Artificial Neural Networks*, pages 131–136, 2013.

[68] Hossein Hassani.
A brief introduction to singular spectrum analysis.
Optimal decisions in statistics and data analysis, 2010.

[69] David Claessen and Andreas Groth.
A beginner,Äôs guide to SSA.
CERES, Ecole Normale Supérieure: Paris, France, 2002.

[70] James B Elsner and Anastasios A Tsonis.
Singular spectrum analysis: a new tool in time series analysis.
Springer Science & Business Media, 1996.

[71] M Sklar.
Fonctions de repartition an dimensions et leurs marges.
Publ. inst. statist. univ. Paris, 8:229–231, 1959.

[72] Aad W Van der Vaart.

Asymptotic statistics, volume 3.
Cambridge university press, 2000.

[73] Jacob Benesty, Jingdong Chen, Yiteng Huang, and Israel Cohen.
Pearson correlation coefficient.
In *Noise reduction in speech processing*, pages 1–4. Springer, 2009.

[74] Tilmann Gneiting and Adrian E Raftery.
Strictly proper scoring rules, prediction, and estimation.
Journal of the American statistical Association, 102(477):359–378, 2007.