Spatio-Temporal DeepKriging for Probabilistic Interpolation and Forecasting

Pratik Nag

pratik.nag@kaust.edu.sa

Orcid ID: 0000-0001-5065-5273

Feb 28, 2024

جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology

EnvStat
@ KAUST

Introduction

- Statistical modeling of evolving spatial and temporal phenomena is crucial for environmental monitoring and climate change detection. In recent years advancement in data collection technologies enabled high-resolution spatio-temporal data collection.
- However, exact likelihood-based computations necessary for traditional statistical analysis requires $\mathcal{O}(n^2)$ time and $\mathcal{O}(n^3)$ memory complexity for a covariance matrix of size $n \times n$.
- Hence it is infeasible to do exact likelihood-based analysis on large-scale spatio-temporal processes.

Prior Studies in Large-Scale Spatio-Temporal Modeling

- In the past two decades, many statistical and machine learning methods have emerged to handle large spatio-temporal datasets.
- In statistical modeling, approaches such as Veccia's approximation, Block Composite Likelihoods, and Hierarchical Bayesian space-time models have gained prominence.
- Within machine learning literature, techniques like Echo-state networks and Graphical neural networks have been developed for tasks such as spatio-temporal interpolation.
- This study expands upon the DeepKriging framework introduced by Chen et al. (2022) to encompass spatio-temporal scenarios. Furthermore, we introduce a two-stage model based on deep neural networks (DNNs) for probabilistic interpolation and forecasting of spatio-temporal processes.

Background

- Consider the real valued spatio-temporal random field $\{Y(s,t), s \in D, t \in \mathcal{T}\}, D \subseteq \mathbb{R}^p, \mathcal{T} \subseteq \mathbb{R}$. Assuming the data is observed at N locations and K time points, the realizations can be given as $\mathbf{Z}_{N,K} = \{Z(s_1,t_1), Z(s_2,t_1), \ldots, Z(s_N,t_K)\}$ such that

$$Z(s,t) = Y(s,t) + \epsilon.$$

- Given observations $\mathbf{Z}_{N,K}$, two common goals of spatio-temporal prediction are probabilistic interpolation, i.e., predict the true process $Y(\mathbf{s}_0,t)$ at unobserved spatial location $\mathbf{s}_0$, and forecasting, i.e., predict $Y(\mathbf{s}_0,t_{K+u})$ at unobserved location $\mathbf{s}_0$ at a future time point t_{K+u}.

Optimal Predictor for Probabilistic Interpolation

- The optimal predictor can written as:

$$\hat{Y}_{\tau}^{opt}((\mathbf{s}_0, t)|\mathbf{Z}_{N,K}) = \underset{\hat{Y}}{\text{argmin}}\, R_1(\hat{Y}_{\tau}(\mathbf{s}_0, t)|\mathbf{Z}_{N,K}),$$

  where $R_1(\cdot)$ represents the true risk function necessary for obtaining the τ-th quantile prediction.

- An estimation for $R_1(\cdot)$ can be expressed through the quantile loss function, defined as:

$$R_1^{emp}(\hat{Y}_{\tau}(\mathbf{s}, t)|\mathbf{Z}_{N,K}) = \frac{1}{NK}\sum_{n=1}^{N}\sum_{k=1}^{K}\rho_{\tau}(\hat{Y}_{\tau}(\mathbf{s}_n, t_k) - Z(\mathbf{s}_n, t_k)),$$

  where $\rho_\tau(v) = v(\tau - I(v < 0))$ and $\tau \in (0, 1)$ is quantile level.

Basis Functions

- Similar to the approach outlined in DeepKriging by Chen et al. (2022) we formulate this as a regression problem with embedded inputs from $(\mathbf{s}, t)$ as covariates and $Y(\cdot, \cdot)$ as response.
- We employ the Wendland compactly supported basis functions defined via $B_1(d) = \frac{(1-d)^6}{3}(35d^2 + 18d + 3)\mathbf{1}\{0 \leq d \leq 1\}$ to represent the spatial locations. The spatial basis functions are then defined as $\phi_i(\mathbf{s}) = B_1(\|\mathbf{s} - u_i\| / \theta)$ with θ as the bandwidth parameter and anchor points (spatial locations) $\{u_1, u_2, ..., u_G\}$.
- To represent the temporal bases, we utilize Gaussian radial basis functions across the time domain. The temporal bases are subsequently formulated as: $\psi_j(t) = \exp(-0.5(t - v_j)^2/(\kappa^2))$ with anchor points (time points) $v \in \{v_1, v_2, ..., v_H\}$ and scale set to $\kappa = |v_1 - v_2|$.

Space-Time.DeepKriging: DNN for Interpolation

- We use a single-output deep neural network structure (**Space-Time.DeepKriging**) to build the spatio-temporal DeepKriging framework with the stacked basis functions as inputs.

- Hence $\hat{Y}_\tau(\cdot,\cdot)$ can be expressed through the DNN as:

$$\hat{Y}_\tau(\mathbf{s}, t) = \Psi(\tau, f_{NN_\tau}(\mathbf{X}_\phi(\mathbf{s}, t))),$$

where $\mathbf{X}_\phi(\mathbf{s}, t)$ is the set of stacked basis functions, $\Psi(\cdot,\cdot)$ is the activation function of the output layer.

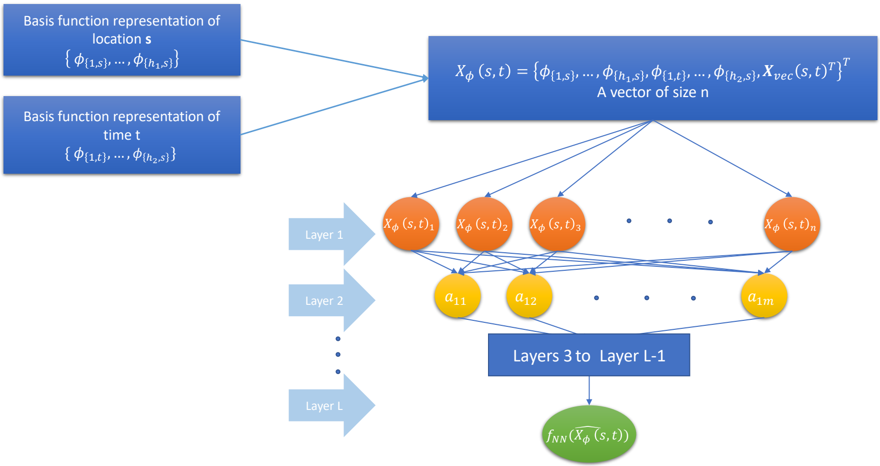
The output Layer Activation $\Psi(\cdot, \cdot)$

- In theory, quantile regression lines are expected not to intersect; however, unconstrained optimization of $R_1^{emp}(\hat{Y}_\tau(\mathbf{s}, t)|\mathbf{Z}_{N,K})$ may inadvertently introduce crossing issues.

- To avoid quantile cross-over, we propose the following activation function for the output layer:

$$\Psi(\tau, x) = \begin{cases} x & \text{for } \tau = 0.5 \\ f_{Constant} + \frac{\lambda(\tau - 0.5)}{1+e^{-x}} & \text{for } \tau > 0.5 \\ f_{Constant} - \frac{\lambda(0.5 - \tau)}{1+e^{-x}} & \text{for } \tau < 0.5, \end{cases}$$

Space-time.DeepKriging Skeleton for Interpolation



Optimal Predictor for Forecasting

- The optimal predictor can written as:

$$\hat{Y}^{opt}_{\tau}((\mathbf{s}_0, t_{K+u})|\mathbf{Z}_{N,K}) = \underset{\hat{Y}}{\text{argmin}}\, R_2(\hat{Y}_{\tau}(\mathbf{s}_0, t)|\mathbf{Z}_{N,K}),$$

where $R_2(\cdot)$ represents the true risk function.

- We can estimate $R_2(\cdot)$ as:

$$R_2^{emp}(\hat{Y}_{\tau}(\mathbf{s}_0, t)|\mathbf{Z}_{N,K}) = \frac{1}{K}\sum_{k=1}^{K} \rho_{\tau}(\hat{Y}_{\tau}(\mathbf{s}_0, t_k) - \mathbf{X}_k^{NN}),$$

where $\mathbf{X}^{NN} = \{\widehat{f_{NN_\tau}}(\mathbf{X}_{\phi}(\mathbf{s_0}, t_1)), \ldots, \widehat{f_{NN_\tau}}(\mathbf{X}_{\phi}(\mathbf{s_0}, t_K))\}^T$ are the predictions from **Space-Time.DeepKriging** for location $\mathbf{s}_0$ at all observed time points.

QLSTM for Probabilistic Forecasting

- We use the Long short-term memory (LSTM) network to perform quantile based forecast of the time series at time point t_{K+u} (We call it **QLSTM**).
- Here $\hat{Y}_\tau(\mathbf{s}_0, t) = \widehat{f_{NN_\tau}^{LSTM}(\mathbf{s}_0}, t)$, where $\widehat{f_{NN_\tau}^{LSTM}(\mathbf{s}_0}, t)$ is a multi-layer stacked LSTM network.

Convolutional LSTM: **QConvLSTM**

- Although **QLSTM** is highly effective for capturing temporal dependence, it does not use information from other locations.

- For space-time data, we propose the convolutional LSTM which includes data from other locations by passing the CNN layer as the input to the LSTM layer. (We call it **QConvLSTM**)

Expression of $R_2(\cdot)$ for **QConvLSTM**

- For this network $R_2(\cdot)$ can be written as:

$$R_2^{emp}(\hat{Y}_\tau(\mathbf{s}_0, t)|\mathbf{Z}_{N,K}) = \frac{1}{K}\sum_{k=1}^{K} \rho_\tau(f_{NN_\tau}^{Conv}(\mathbf{s}_0, t_k) - \mathbf{X}_k^{NN}),$$

  where $f_{NN_\tau}^{Conv}(\mathbf{s}_0, t_k)$ is the output of **QConvLSTM**.

- The sole distinction between $f_{NN_\tau}^{Conv}(\mathbf{s}_0, t_k)$ and $f_{NN_\tau}^{LSTM}(\mathbf{s}_0, t)$ lies in the former's utilization of matrix inputs $\mathbf{X}^{NN_{CONV}}$, as provided below: $\mathbf{X}^{NN_{CONV}} = \{\mathcal{A}(\mathbf{s}_0, t_1), \ldots, \mathcal{A}(\mathbf{s}_0, t_K)\}$. The matrix $\mathcal{A}(\mathbf{s}_0, t) = \{\widehat{f_{NN_\tau}}(\mathbf{X}_\phi(\mathbf{s}_j, t)) : \mathbf{s}_j \in N_{\mathbf{s}_0}\}$ (where $N_{\mathbf{s}_0}$ is a gridded neighbourhood of $\mathbf{s}_0$), is a $r \times r$ matrix with elements $[X_t(i,j)]_{i,j\in\{1,\ldots,r\}}$.

Simulation Studies

- The proposed **Space-Time.DeepKriging** won the KAUST competition in large-scale prediction on 100k and 1M space-time locations with double digit improvement in percentage for MSPE over competing methods such as the Veccia's approximation and block composite likelihood.

- We compare the method on a simulated nonstationary field with 50k space-time locations with other competitive methods that can be applied for large-scale interpolation and forecasting.

Numeric Results on Simulated Data

Table: Average MSPE of prediction for simulated data. Here SE stands for standard error of the predictions.

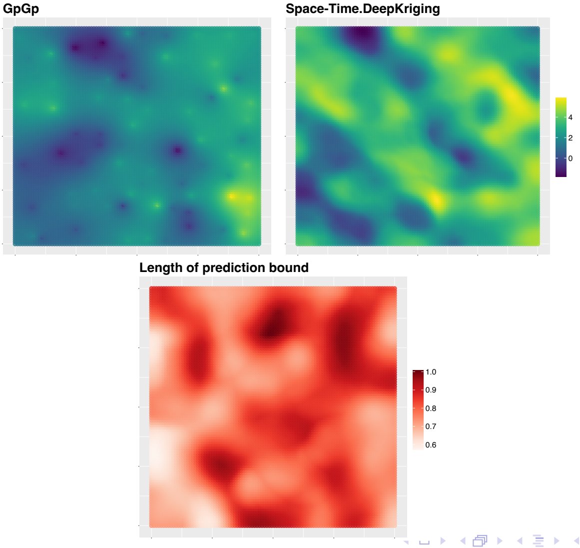
Models	MSPE	SE
Space-Time.DeepKriging	0.167	0.073
GpGp	0.746	0.288

Table: Average MSPE, MPIW and PICP of forecast for simulated data.

Models	Avg.MSPE	SE	Avg.MPIW	SE	Avg.PICP
QConvLSTM	0.267	0.219	1.462	0.126	90.39
ARIMA	0.277	0.278	2.262	0.082	90.72
QLSTM	0.392	0.523	1.558	0.316	89.94
GpGp	0.839	0.358	-	-	-

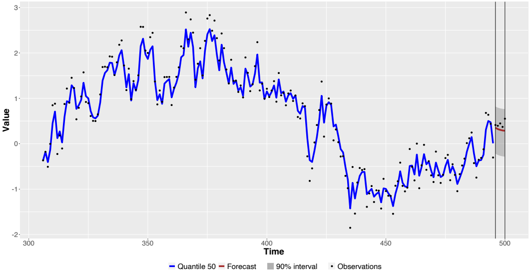
Interpolation on Simulated Data

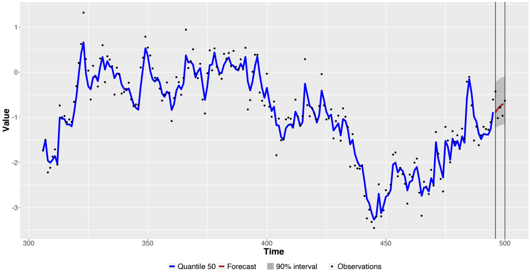
- Interpolation on unit square



Forecast on Simulated Data

- Forecasting at specific observed locations using **QConvLSTM**.

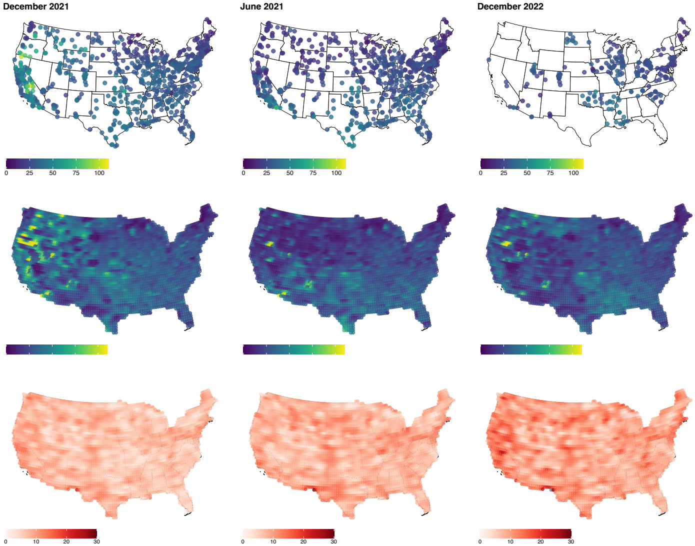




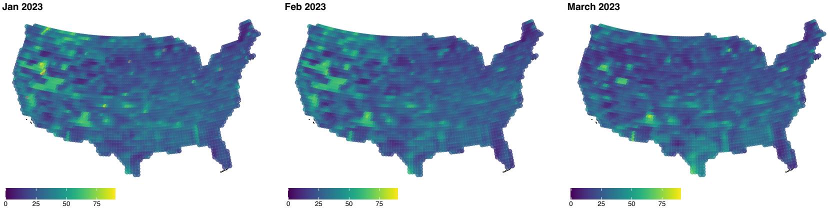
Application on $PM_{2.5}$

- We also apply our method to the $PM_{2.5}$ data over USA.
  - Time period: from January 1998 to December 2022.
  - Time resolution: monthly, in total 286 months.
  - Spatial region: The United States of America.
  - Spatial dimension: There were on average 1900 weather stations per month.

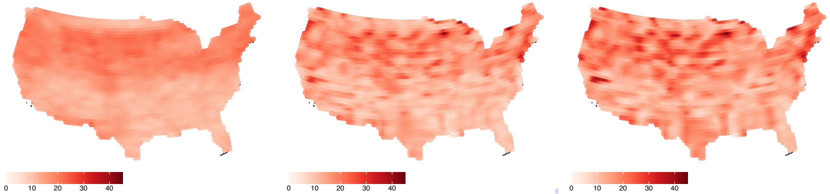
Interpolation: $PM_{2.5}$



Forecast: $PM_{2.5}$

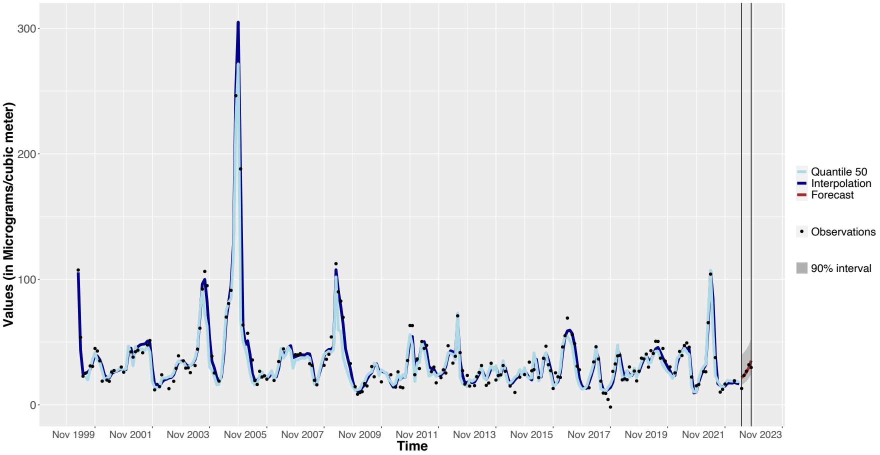


(a)



$PM_{2.5}$ Forecast for San Francisco using **QConvLSTM**

- Forecast period for the last six observed months up to December 2022:



Conclusion

- In this study, without making any parametric assumptions about the underlying distribution of the data, we have established a novel, easy-to-use methodology for interpolation as well as forecasting for spatio-temporal processes.
- In order to further enhance our deep learning-based spatio-temporal modeling architecture, we have additionally included semi-parametric quantile-based prediction intervals.
- Our proposed method for spatio-temporal interpolation and forecasting is valid for general class of non-Gaussian and nonstationary spatio-temporal processes.
- Our proposed approach can be easily extended to large datasets with minimum hardware support.

PhD Completed Projects: First Author

- Nag, Pratik, Ying Sun, and Brian J. Reich. "Spatio-temporal DeepKriging for Interpolation and Probabilistic Forecasting." Spatial Statistics 57 no. 100773 (2023). https://doi.org/10.1016/j.spasta.2023.100773.
- Nag, Pratik, Ying Sun, and Brian J. Reich. "Bivariate DeepKriging for large-scale spatial interpolation of wind fields." arXiv preprint arXiv:2307.08038 (2023). (In revision at Technometrics).
- Nag, Pratik, Yiping Hong, Sameh Abdulah, Ghulam A. Qadir, Marc G. Genton, and Ying Sun. "Efficient Large-scale Nonstationary Spatial Covariance Function Estimation Using Convolutional Neural Networks." arXiv preprint arXiv:2306.11487 (2023). (In revision at the Journal of Computational and Graphical Statistics (JCGS)).
- Hazra, Arnab, Pratik Nag, Rishikesh Yadav, and Ying Sun. "Exploring the Efficacy of Statistical and Deep Learning Methods for Large Spatial Datasets: A Case Study." Journal of Agricultural, Biological and Environmental Statistics (2024): 1-24.

PhD Completed Projects: Collaboration

- Cao, Qinglei, Sameh Abdulah, Rabab Alomairy, Yu Pei, Pratik Nag, George Bosilca, Jack Dongarra et al. "Reshaping geostatistical modeling and prediction for extreme-scale environmental applications." In SC22: International Conference for High Performance Computing, Networking, Storage and Analysis, pp. 1-12. IEEE, 2022.

- Abdulah, Sameh, Alamri, Faten, Nag, Pratik, Sun, Ying, Ltaief, Hatem, Keyes, David E., and Marc G. Genton. "The Second Competition on Spatial Statistics for Large Datasets." Journal of Data Science 20 no. 4 (2022):439-460. https://doi.org/10.6339/22-JDS1076.

Thank You

Pratik Nag

pratik187

Passionate about data-driven insights, pursuing my PhD in KAUST, and committed to leveraging technology for positive impact.