

國立東華大學應用數學系

碩士論文

基於結構化狀態空間擴散模型與自適應固定秩克利金法之未知區域時空預測

*Spatiotemporal Prediction of Unknown Areas Based on Structured State Space Diffusion and Adaptive Fixed Rank Kriging*



研究生：許堯智  
指導教授：黃灝勻 博士

中華民國 115 年 6 月



國立東華大學應用數學系

碩士論文

基於結構化狀態空間擴散模型與自適應固定秩克利金法之未知區域時空預測

*Spatiotemporal Prediction of Unknown Areas Based on Structured State Space Diffusion and Adaptive Fixed Rank Kriging*



研究生：許堯智  
指導教授：黃灝勻 博士

中華民國 115 年 6 月



# 學位考試委員會審定書

國立東華大學 \_\_\_\_\_ 系所

研究生 \_\_\_\_\_ 君所提之論文

---

經本委員會審查並舉行口試，認為

符合碩士學位標準。

學位考試委員會召集人 \_\_\_\_\_ 簽章

指導教授 \_\_\_\_\_ 簽章

委員 \_\_\_\_\_ 簽章

委員 \_\_\_\_\_ 簽章

委員 \_\_\_\_\_ 簽章

系主任 \_\_\_\_\_ 簽章  
(所長)

中華民國 \_\_\_\_\_ 年 \_\_\_\_\_ 月 \_\_\_\_\_ 日



國立東華大學  
NATIONAL DONG HWA UNIVERSITY

學位論文原創性聲明書  
DECLARATION OF THESIS/DISSERTATION ORIGINALITY

學位論文題目：基於結構化狀態空間擴散模型與自適應固定秩克利金法之未知區域時空預測

Thesis/Dissertation Title : *Spatiotemporal Prediction of Unknown Areas Based on Structured State Space Diffusion and Adaptive Fixed Rank Kriging*

本人在此聲明，所呈交的學位論文是在指導教授黃灝勻的指導下，由個人獨立研究所完成之最終版本。本人對論文內容負責，除了文中已經標註引用處的內容外，論文不包含任何其他已經發表或撰寫過的研究成果。對本研究及學位論文做出重要貢獻的個人和組織，均已在文中以明確方式標明。

該論文內容如有違反學術道德或學術規範的行為，如造假、變造、抄襲、研究成果重複發表或未適當引註、以違法或不當手段影響論文審查、不當作者列名等，本人願意承擔由此而產生的法律責任和法律後果。

I declare that the thesis/dissertation herein is the final version of my work, which is composed and accomplished individually under the guidance of my supervisor, Prof. Hao-Yun Huang. I am responsible for the contents of this thesis/dissertation: It contains no research result that was previously published or written by another person. Information derived from published and unpublished work of others has been acknowledged in the text, and a list of references is given . Any contribution made by other individual or organization is explicitly acknowledged in the thesis/dissertation.

If any research misconduct, including fabrication, falsification, or plagiarism in proposing, performing, or reviewing research, or in reporting research results, is discovered in my thesis/dissertation, I am willing to bear corresponding legal responsibilities and all the results therefrom.

聲明人 Declarant :

日期 Date : \_\_\_\_\_(yyyy/mm/dd)

1140303修訂



## 摘要

隨著感測技術與資料蒐集系統之發展，時空資料於交通、氣象與環境監測等領域中大量累積，然而實務上仍普遍存在觀測缺漏與未觀測地點之問題，對後續分析與預測造成挑戰。為兼顧時間序列之動態特性與空間資料之相關結構，本研究提出一種整合式時空預測框架  $SSSD^{S4+AFRK}$ ，結合結構化狀態空間擴散模型（Structured State Space Diffusion with S4 layers,  $SSSD^{S4}$ ）與自適應固定秩克利金法（Adaptive Fixed Rank Kriging, AFRK），以提升缺失資料重建與未來預測之準確性。

在方法上，本研究首先利用  $SSSD^{S4}$  擷取時間序列之長期依賴與非線性動態特徵，並於模型訓練過程中引入 AFRK 對擴散過程中的雜訊估計進行空間結構調整，以強化時空一致性；於推論階段，則結合時間預測結果與空間基底函數進行未觀測地點之空間插值與未來預測。

實驗採用 Weather2K、MERRA-2 與 2b-8 等多樣化時空資料集，並與多種基準模型（包含 TFT、VAR、SVGP 與 STDK）進行比較。結果顯示，所提出之  $SSSD^{S4+AFRK}$  在多數情境下，特別是在未觀測地點之未來預測任務中，能顯著降低平均平方預測誤差（MSPE），並於高維度及高變異資料中展現穩定且優異之預測效能。相較之下，僅考慮時間或空間資訊之模型，其預測表現較易受到資料分布與變異性影響。

本研究證實透過深度時間序列模型與空間統計方法之整合，可有效提升時空資料在缺失與外插情境下之預測能力，並提供一具擴展性之建模框架，對於未來大規模時空資料分析具有實務應用潛力。

**關鍵字：**時空預測、結構化狀態空間模型、擴散模型、自適應固定秩克利金、空間插值、時間序列分析



# Abstract

With the advancement of sensing technologies and data acquisition systems, large volumes of spatiotemporal data have been accumulated in domains such as transportation, meteorology, and environmental monitoring. However, in practice, missing observations and unobserved locations remain prevalent, posing significant challenges for subsequent analysis and forecasting. To address both the dynamic nature of temporal sequences and the spatial dependency structure of data, this study proposes an integrated spatiotemporal forecasting framework,  $\text{SSSD}^{\text{S4+AFRK}}$ , which combines Structured State Space Diffusion with S4 layers ( $\text{SSSD}^{\text{S4}}$ ) and Adaptive Fixed Rank Kriging (AFRK), aiming to improve the accuracy of missing data reconstruction and future prediction.

Methodologically, the proposed approach first leverages  $\text{SSSD}^{\text{S4}}$  to capture long-range dependencies and nonlinear temporal dynamics in time series data. During training, AFRK is incorporated to adjust spatial structures in the noise estimation process of the diffusion model, thereby enhancing spatiotemporal consistency. During inference, the model further integrates temporal predictions with spatial basis functions to perform spatial interpolation and future forecasting at unobserved locations.

Experiments are conducted on diverse spatiotemporal datasets, including Weather2K, MERRA-2, and 2b-8, and the proposed method is compared against several baseline models, including TFT, VAR, SVGP, and STDK. Results demonstrate that  $\text{SSSD}^{\text{S4+AFRK}}$  consistently reduces mean squared prediction error (MSPE) under most settings, particularly in forecasting tasks at unobserved locations. Moreover, it exhibits stable and superior performance in high-dimensional and highly variable data regimes. In contrast, models that consider only temporal or spatial information are more susceptible to data distribution shifts and variability.

Overall, the findings validate that integrating deep temporal sequence models with spatial statistical methods can effectively enhance forecasting performance under missing and extrapolation scenarios. Furthermore, the proposed framework provides a scalable modeling paradigm with strong potential for large-scale spatiotemporal data analysis in real-world applications.

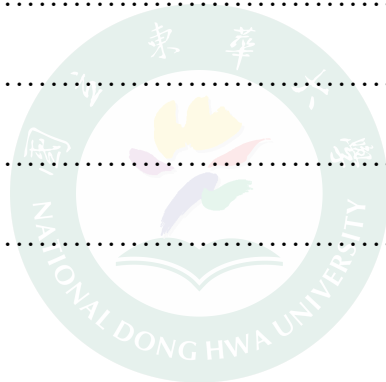
**Keywords:** spatiotemporal forecasting, Structured State Space Models, Diffusion Models, Adaptive Fixed Rank Kriging, spatial interpolation, time series analysis



# 目錄

學位考試委員會審定書 .....	i
學位論文原創性聲明書 .....	ii
摘要 .....	iii
Abstract .....	iv
目錄 .....	v
<b>1 引言 .....</b>	<b>1</b>
<b>2 相關研究 .....</b>	<b>1</b>
2.1 時間序列模型 .....	2
2.1.1 向量自迴歸模型 .....	2
2.1.2 時序融合變換器 .....	3
2.1.3 狀態空間模型 .....	4
2.1.4 結構化狀態空間模型 .....	5
2.1.5 擴散模型 .....	6
2.1.6 基於 S4 層的結構化狀態空間擴散模型 .....	8
2.2 空間統計學 .....	9
2.2.1 克利金法 .....	9
2.2.2 固定秩克利金法 .....	10
2.2.3 自適應固定秩克利金法 .....	10
2.3 時空模型 .....	11
2.3.1 高斯過程 .....	11
2.3.2 時空深度克利金法 .....	12
<b>3 研究方法 .....</b>	<b>13</b>
3.1 問題定義 .....	13
3.2 時空建模方法 .....	14
3.2.1 基於 SSSD <sup>S4</sup> 之時間建模 .....	14
3.2.2 基於 AFRK 之空間建模 .....	15
3.3 模型訓練與推論演算法 .....	16

<b>4</b>	<b>實驗</b> .....	<b>17</b>
4.1	資料集.....	17
4.1.1	Weather2K.....	17
4.1.2	MERRA-2.....	18
4.1.3	2b-8.....	18
4.2	實驗環境與計算資源.....	19
4.3	實驗設計.....	19
<b>5</b>	<b>實驗結果</b> .....	<b>23</b>
<b>6</b>	<b>結論</b> .....	<b>28</b>
	<b>參考文獻</b> .....	<b>28</b>
	<b>附錄</b> .....	<b>33</b>
<b>A</b>	<b>Weather2K 資料集</b> .....	<b>33</b>
<b>B</b>	<b>MERRA-2 資料集</b> .....	<b>34</b>
<b>C</b>	<b>2b-8 資料集</b> .....	<b>35</b>



# 1 引言

隨著觀測技術與數位基礎設施的快速進步，研究人員與大眾可以更輕易地透過網路平臺、即時感測器以及各類公開資料庫，獲取交通流量、水質監測或衛星遙測等多樣化的資料。然而，在資料蒐集過程中，儀器仍可能因故障、失靈或維護停機而導致觀測值缺漏，進而形成時空資料的不完整性。當我們希望針對缺漏位置進行未來的趨勢預測時，如何兼顧時間序列的特徵與空間資料的相關結構成爲一個重要議題 (Decorte et al., 2024)。

爲了解決上述問題，本研究提出一種結合時間序列預測與空間統計建模的方法。我們利用結構化狀態空間擴散模型 (Structured State Space Diffusion, SSSD) (Alcaraz and Strodthoff, 2023) 捕捉資料的時間相依性，同時結合自適應固定秩克利金法 (Adaptive Fixed Rank Kriging, AFRK) (Tzeng and Huang, 2018) 用於刻畫空間相關性。透過此種時空結合框架，提升對缺失資料的重建能力，也能強化未來趨勢的預測準確性。

## 2 相關研究

時空資料分析在交通工程、水文監測與環境科學等領域中皆扮演關鍵角色。隨著觀測技術與感測設備的普及，大規模且高解析度的時空資料得以持續蒐集。然而，實際資料收集過程中仍可能因感測器故障、通訊中斷或維護作業而產生缺失值與離群值，增加資料分析與預測建模的困難度 (Decorte et al., 2024; Little and Rubin, 2002)。過去相關研究多分別針對時間序列模型或空間模型進行探討，但由於時空資料同時具有時間依賴與空間相關結構，近年研究逐漸發展出整合時間與空間資訊的時空建模方法，以提升預測與推估的準確性 (N. Cressie and Wikle, 2011; Shi et al., 2015)。

在時間建模方面，傳統方法如自迴歸整合移動平均模型 (Autoregressive Integrated Moving Average, ARIMA) (Box and Jenkins, 1976) 能有效捕捉單變量時間依賴性，而向量自迴歸模型 (Vector Autoregression, VAR) (Sims, 1980) 則進一步擴展至多變量系統，描述變數間的動態相互依賴關係。然而，這些線性模型在處理複雜非線性結構時仍有限制 (Primiceri, 2005)。

繼之而起的長短期記憶網路 (Long Short-Term Memory, LSTM) (Hochreiter and Schmidhuber, 1997) 雖強化了對非線性動態的捕捉，但在處理長時序資料時，仍受限

於計算效率及遠期資訊遺失的問題 (Bengio, Simard, and Frasconi, 1994)。相較之下，時序融合變換器 (Temporal Fusion Transformers, TFT) (Lim et al., 2020) 導入了特徵選擇機制與注意力架構，能針對不同類型的輸入變數進行整合，提升了模型對於複雜時空特徵的辨識能力與預測的可解釋性。然而，TFT 在面對極端高維度數據或存在大規模連續缺失值的觀測序列時，其注意力機制往往難以維持穩定的表徵能力，且對於潛在隨機擾動的穩健性仍有提升空間 (Wu et al., 2022)。

在此基礎上，近期提出的狀態空間模型 (State Space Model, SSM) (Kalman, 1960) 及結構化狀態空間擴散模型 (Structured State Space Diffusion, SSSD) (Alcaraz and Strodthoff, 2023) 展現了在長時間序列建模上的優勢，透過結合 S4 層與擴散機率架構，使其能兼具高品質的資料填補、預測能力與時間成本控制。

在空間建模方面，空間統計學中的克利金 (Kriging) 方法 (N. A. C. Cressie, 1993) 被廣泛應用於空間插值與缺失值推論，但在面對大規模資料時，其計算成本高昂，限制了即時應用的可行性。為解決此問題，固定秩克利金 (Fixed Rank Kriging, FRK) (N. Cressie and Johannesson, 2008) 及其延伸方法自適應固定秩克利金 (Adaptive Fixed Rank Kriging, AFRK) 方法被提出 (N. Cressie and Johannesson, 2008; Tzeng and Huang, 2018)，有效降低計算複雜度並減少時間成本。

時間模型能刻劃資料的動態變化，而空間模型則強調鄰近地點間的相關結構。近期研究逐漸意識到將兩者結合的重要性，例如在交通需求預測、氣象模擬與環境監測中的應用。基於此背景，本研究提出結合 SSSD 與 AFRK 的時空整合方法，以提升在資料缺失情境下的預測準確度與穩健性。

## 2.1 時間序列模型

### 2.1.1 向量自迴歸模型

向量自迴歸模型 (Vector Autoregression model, VAR) 係由 Sims (1980) 提出，為單變量自迴歸模型 (AR) 之擴展，旨在捕捉多個隨時間變化變數間的動態相互依賴關係。與單變量模型不同，VAR 將系統中所有變數均視為內生變數 (endogenous variables)，並透過聯立方程式組描述變數間的滯後影響。

對於一個包含  $k$  個變數之  $p$  階向量自迴歸模型  $\text{VAR}(p)$ ，其數學表達式為：

$$\mathbf{y}_t = \mathbf{c} + \sum_{i=1}^p \Phi_i \mathbf{y}_{t-i} + \boldsymbol{\epsilon}_t, \quad (1)$$

其中， $\mathbf{y}_t \in \mathbb{R}^k$  為  $t$  時刻之觀測向量； $\mathbf{c}$  為截距項向量； $\Phi_i \in \mathbb{R}^{k \times k}$  為滯後算子矩陣，用以衡量  $t-i$  時刻之變數對當前狀態的影響強度； $\boldsymbol{\epsilon}_t \sim \mathcal{N}(0, \Omega)$  為雜訊向量。

VAR 模型之優勢在於其能透過脈衝響應函數（Impulse Response Function）與變異數分解（Variance Decomposition）分析變數間的動態交互影響 (Sims, 1980)。然而，隨著變數數量  $k$  或滯後階數  $p$  增加，參數數量將呈平方增長，易導致過度擬合問題，且傳統線性 VAR 模型難以捕捉複雜的非線性結構與參數的時變性 (Primiceri, 2005)。

### 2.1.2 時序融合變換器

鑒於傳統線性多變量模型在捕捉複雜非線性動態與高維特徵關聯上之侷限，近期研究多轉向利用深度學習架構以提升預測效能。為克服傳統循環架構在處理長距離依賴時之困境，並有效整合包含靜態背景資訊、已知未來變數與歷史觀測值在內的多樣化資訊來源，Lim et al. (2020) 提出了時序融合變換器（Temporal Fusion Transformers, TFT）。該模型係一種專為多步時序預測設計的深度學習架構，其設計重點在於透過專門的網路組件對不同性質的輸入變數進行權重分配，從而提升預測結果的解釋能力，補足傳統統計模型在處理大規模異質資料時之不足。

TFT 之實現原理依賴於門控殘差網路（Gated Residual Networks, GRN），該組件透過門控線性單元（Gated Linear Units, GLU）控制資訊流動，使模型能根據資料特性自動調整非線性轉換之深度。對於輸入向量  $\mathbf{a}$  與選用的背景向量  $\mathbf{c}$ ，其運算過程如下：

$$\text{GRN}_\omega(\mathbf{a}, \mathbf{c}) = \text{LayerNorm}(\mathbf{a} + \text{GLU}_\omega(\boldsymbol{\eta}_1)), \quad (2)$$

其中  $\boldsymbol{\eta}_1$  為經過權重矩陣轉換後之特徵向量。此機制確保了模型在處理不同複雜度的序列時具備高度彈性，有效避免了深層網路在簡單資料集上的過度擬合問題。

針對包含大量外部因子之時空資料，TFT 導入了變數選擇網路（Variable Selection Networks, VSN），旨在從眾多輸入特徵中篩選關鍵變數。透過為每個特徵分配權重  $\nu_t^{(i)}$ ，模型得以自動忽略冗餘資訊並聚焦於具影響力之因子，其整合後之特徵向量表

示如下：

$$\tilde{\boldsymbol{\xi}}_t = \sum_{i=1}^m \nu_t^{(i)} \tilde{\boldsymbol{\xi}}_t^{(i)}, \quad (3)$$

其中  $\tilde{\boldsymbol{\xi}}_t^{(i)}$  為處理後之特徵向量。此設計大幅提升了模型在面對多維度特徵輸入時的穩健性，並使研究者能直觀地量化各類變數對預測結果之貢獻。

在時間序列關係的捕捉上，TFT 採用改良之時序自我注意力機制（Temporal Self-Attention）來處理長期依賴關係。相較於標準 Transformer 架構，TFT 在注意力層中加入門控層進行殘差連接，並透過解碼器整合歷史與未來的時空背景資訊，進而找出對當前預測最具影響力的時間點。TFT 不僅在預測精確度上表現優異，更賦予了深度學習模型在特定時間步或特定特徵上之解釋能力 (Lim et al., 2020)。

### 2.1.3 狀態空間模型

狀態空間模型（State Space Model, SSM）是一類透過潛在狀態（latent state）向量來描述動態系統或序列資料的數學模型。其最初由 Kalman (1960) 在控制理論與濾波領域提出，用以解決線性動態系統的最佳濾波與預測問題。隨後，Gu, Goel, and Ré (2022) 等人將該概念推廣至深度學習架構中，用以處理長序列的時間序列建模，證明 SSM 相較於傳統 RNN 與 LSTM 在捕捉長期依賴性（long-range dependencies）與維持穩定梯度方面表現更優 (Gu, Goel, and Ré, 2022)。

給定一維輸入訊號序列  $\mathbf{u}(t)$  與一維輸出訊號序列  $\mathbf{y}(t)$ ，SSM 的基本形式為

$$\begin{aligned} \mathbf{x}'(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t); \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t), \end{aligned} \quad (4)$$

其中， $\mathbf{x}(t) \in \mathbb{R}^N$  為  $\mathbf{u}(t)$  映射的  $N$  維度潛在狀態； $\mathbf{x}'(t) = \frac{d}{dt}\mathbf{x}(t)$ ； $\mathbf{A} \in \mathbb{R}^{N \times N}$  為狀態矩陣（state matrix）； $\mathbf{B} \in \mathbb{R}^{N \times 1}$  與  $\mathbf{C} \in \mathbb{R}^{1 \times N}$  分別為輸入與輸出矩陣，用於描述輸入對狀態的影響以及狀態對輸出的映射；而  $\mathbf{D} \in \mathbb{R}$  則為前饋矩陣（feedthrough matrix），使輸入向量直接影響輸出結果，通常為零矩陣。在深度學習中，這些參數通常透過梯度下降進行學習。

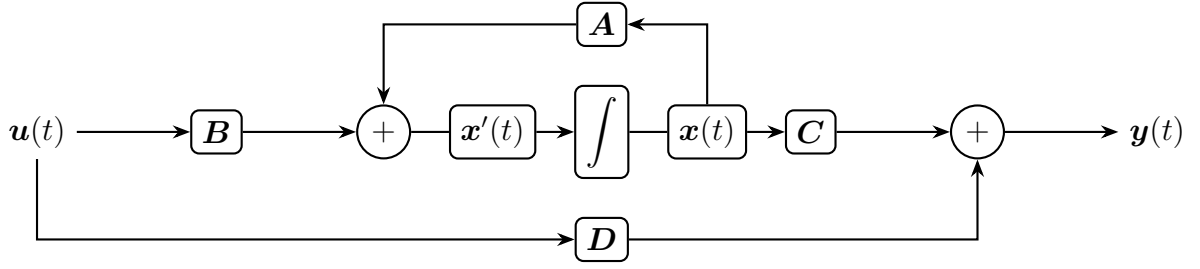


圖 1: 典型的狀態空間模型。

為解決 SSM 在實務中，梯度可能會出現隨序列長度呈指數增減的問題，Gu, Goel, and Ré (2022) 引入 HiPPO (High-order Polynomial Projection Operators) 矩陣 (Gu, Dao, et al., 2020)，用以取代原 (4) 中的隨機矩陣  $A$ 。

$$A_{nk} = - \begin{cases} (2n+1)^{1/2}(2k+1)^{1/2}, & \text{if } n > k; \\ n+1, & \text{if } n = k; \\ 0, & \text{if } n < k. \end{cases} \quad (5)$$

如 (5) 所示的 HiPPO 矩陣能夠隨時間變化衡量過去每個時間步的重要性，動態調整記憶並記住所有歷史。將原本的隨機矩陣  $A$  替換為 HiPPO 矩陣後，狀態  $x(t)$  可以有效記憶輸入序列  $u(t)$  的歷史資訊，同時避免梯度爆炸或消失。實驗結果顯示，這種設計不僅能提升模型的計算穩定性，也顯著改善長序列預測的性能 (Gu, Goel, and Ré, 2022)。

#### 2.1.4 結構化狀態空間模型

基於前一節的理論，Gu, Goel, and Ré (2022) 提出了結構化狀態空間模型 (Structured State Space Sequence Model, S4)。該模型旨在將連續時間狀態空間模型離散化，並應用於深度學習框架中以處理長序列資料。

S4 以 SSM 為基礎，將 (4) 離散化，以便將 SSM 應用於離散的輸入序列。定義步長為  $\Delta$ ，則離散化的 SSM 如下：

$$\begin{aligned} \mathbf{x}_k &= \bar{A}\mathbf{x}_{k-1} + \bar{B}\mathbf{u}_k; \\ \mathbf{y}_k &= \bar{C}\mathbf{x}_k, \end{aligned} \quad (6)$$

其中， $\overline{A}, \overline{B}, \overline{C}$  為離散近似形式的  $A, B, C$ 。

$$\begin{aligned}\overline{A} &= (I - \Delta/2 \cdot A)^{-1}(I + \Delta/2 \cdot A); \\ \overline{B} &= (I - \Delta/2 \cdot A)^{-1}\Delta B; \\ \overline{C} &= C.\end{aligned}\tag{7}$$

為降低矩陣運算成本，S4 可透過對角化將離散矩陣，表示為不同基底下的等價形式。

$$(A, B, C) \sim (V^{-1}AV, V^{-1}B, CV),\tag{8}$$

其中  $V$  為基底變換矩陣。同時，離散化 SSM 可展開為卷積形式，以提高平行運算效率：

$$\mathbf{y} = \overline{\mathbf{K}} * \mathbf{u};\tag{9}$$

$$\overline{\mathbf{K}} \in \mathbb{R}^L := (\overline{CB}, \overline{CAB}, \dots, \overline{CA}^{L-1}\overline{B}),\tag{10}$$

其中， $\overline{\mathbf{K}}$  為 SSM 卷積核， $L$  表示卷積長度。

為進一步降低運算複雜度，S4 透過 Normal Plus Low-Rank (NPLR) 方式，將  $\overline{A}$  矩陣參數化：

$$A = V\Lambda V^* - PQ^\top = V(\Lambda - (V^*P)(V^*Q)^*)V^*,\tag{11}$$

其中  $\Lambda$  為對角矩陣， $P, Q \in \mathbb{R}^{N \times r}$  為低秩 (low-rank) 矩陣， $V \in \mathbb{C}^{N \times N}$  為酉 (unitary) 矩陣。

綜合以上研究，S4 結合 HiPPO 矩陣、離散化、對角化、卷積與 NPLR 參數化，使模型在長序列任務中實現高效運算與優異表現 (Gu, Goel, and Ré, 2022)。

### 2.1.5 擴散模型

擴散模型 (Diffusion Model) 是透過建立正向擴散過程 (forward diffusion process) 與反向去噪過程 (reverse denoising process) 的對偶結構，學習資料的生成分佈 (Sohl-Dickstein et al., 2015)。其模型逐步向原始資料注入高斯雜訊，使其最終接近標準常

態分佈，並訓練模型學習該過程的反向映射以重建資料分佈。近年來，該方法被應用於時間序列插補，僅對缺失區段進行擴散與去噪操作，以在條件觀測下恢復完整序列 (Alcaraz and Strodthoff, 2023)。

令  $\mathbf{x}_0 \sim q(\mathbf{x}_0)$  表示原始資料樣本，正向過程被定義為一個固定參數的高斯馬可夫鏈，用以模擬逐步擾動的資料生成機制：

$$\begin{cases} q(\mathbf{x}_1|\mathbf{x}_0) = \prod_{t=1}^T q(\mathbf{x}_t|\mathbf{x}_{t-1}); \\ q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}), \end{cases} \quad (12)$$

其中， $\beta_t$  為控制雜訊強度的變異數； $\mathcal{N}$  為常態分佈。

為了重建資料，模型需學習該過程的反向映射。反向過程定義如下：

$$\begin{cases} p_\theta(\mathbf{x}_0) = p(\mathbf{x}_T) \prod_{t=1}^T p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t); \\ p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \Sigma_\theta(\mathbf{x}_t, t)), \end{cases} \quad (13)$$

其中， $p(\mathbf{x}_T) = \mathcal{N}(\mathbf{x}_T; 0, \mathbf{I})$  為標準常態分佈； $\mu_\theta$  與  $\Sigma_\theta$  為經參數  $\theta$  的神經網路所參數化的平均值與共變異數矩陣。

然而，直接對反向過程的均值  $\mu_\theta$  進行建模在實作上往往難以收斂。因此，Ho, Jain, and Abbeel (2020) 提出了一種稱為去噪擴散機率模型 (Denoising Diffusion Probabilistic Models, DDPM) 的參數化方式，將  $p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t)$  重新參數化為

$$\mu_\theta(\mathbf{x}_t, t) = \frac{1}{\sqrt{\alpha_t}} \left( \mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \right), \quad (14)$$

$$\Sigma_\theta(\mathbf{x}_t, t) = \sigma_t^2 \mathbf{I}, \quad \sigma_t^2 = \beta_t \text{ or } \sigma_t^2 = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t} \beta_t, \quad (15)$$

其中， $\alpha_t = 1 - \beta_t$ ； $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$ 。在此架構下， $\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)$  是用於估計  $\mathbf{x}_t$  於正向擴散過程中所加入的隨機高斯雜訊，並透過移除  $\mathbf{x}_t$  中該雜訊的估計值重建  $\mathbf{x}_{t-1}$ 。此參數化策略避免了對複雜高維資料分佈的直接建模，大幅簡化訓練目標並提升數值穩定性。因此，任意擴散步驟  $t$  之樣本可表述為原始資料  $\mathbf{x}_0$  與雜訊之線性組合

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t} \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I}), \quad (16)$$

使得模型能夠在訓練期間隨機抽樣時間步  $t$  與雜訊  $\boldsymbol{\epsilon}$ ，而無需迭代計算中間步驟。由於此表示方式將原先對資料生成分佈的直接擬合問題轉化為對高斯雜訊的估計，訓

練目標可進一步簡化為

$$\min_{\theta} \mathbb{E}_{t, x_0, \epsilon} \left[ \left\| \epsilon - \epsilon_{\theta} \left( \sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, t \right) \right\|^2 \right], \quad (17)$$

即僅需最小化預測雜訊與實際注入雜訊之間的平均平方誤差（Mean Squared Error, MSE）。

在時間序列插補應用中，Alcaraz and Strodthoff (2023) 進一步提出條件式擴散模型（Conditional Diffusion Model），僅對缺失片段施加擴散與去噪操作。於訓練階段，模型接收部分觀測序列作為條件輸入以學習重建完整序列；而於生成階段，藉由固定已觀測部分並執行反向去噪過程，即可在保持時間一致性的前提下補全缺失區段。此方法結合了擴散模型的穩定性與高品質生成能力，能有效處理具長期依賴或結構性缺失的時間序列資料。

### 2.1.6 基於 S4 層的結構化狀態空間擴散模型

結構化狀態空間擴散模型（Structured State Space Diffusion, SSSD）由 Alcaraz and Strodthoff (2023) 提出，旨在結合 DiffWave 架構擴散模型 (Kong et al., 2021) 的生成穩定性與 SSM 的長期依賴建模能力。該模型以條件擴散（conditional diffusion）的形式應用於時間序列插補任務，僅對缺失區段施加雜訊，而保留已觀測部分不受擾動，藉此避免資料洩漏並維持條件一致性。透過在每一步中學習反向去噪過程，SSSD 能在固定觀測條件下逐步重建缺失資料，展現高品質的插補與生成效果。

在此基礎上，Alcaraz and Strodthoff (2023) 進一步提出基於 S4 層的結構化狀態空間擴散模型（Structured State Space Diffusion with S4 Layers, SSSD<sup>S4</sup>），作為針對時間序列任務的改進變體。SSSD<sup>S4</sup> 延續 SSSD 的條件擴散設計，並將原本 DiffWave 架構中的雙向擴張卷積層（Bidirectional Dilated Convolution Layer）替換為 Structured State Space Layer (S4)，以提升模型對長期序列動態的建模能力，其結構如圖 2 所示。實驗結果顯示，該變體在多種缺失模式下均表現出更穩定且準確的插補效果，較以傳統卷積或 Transformer 基礎的擴散模型更穩定且具有更好插補表現 (Alcaraz and Strodthoff, 2023)。

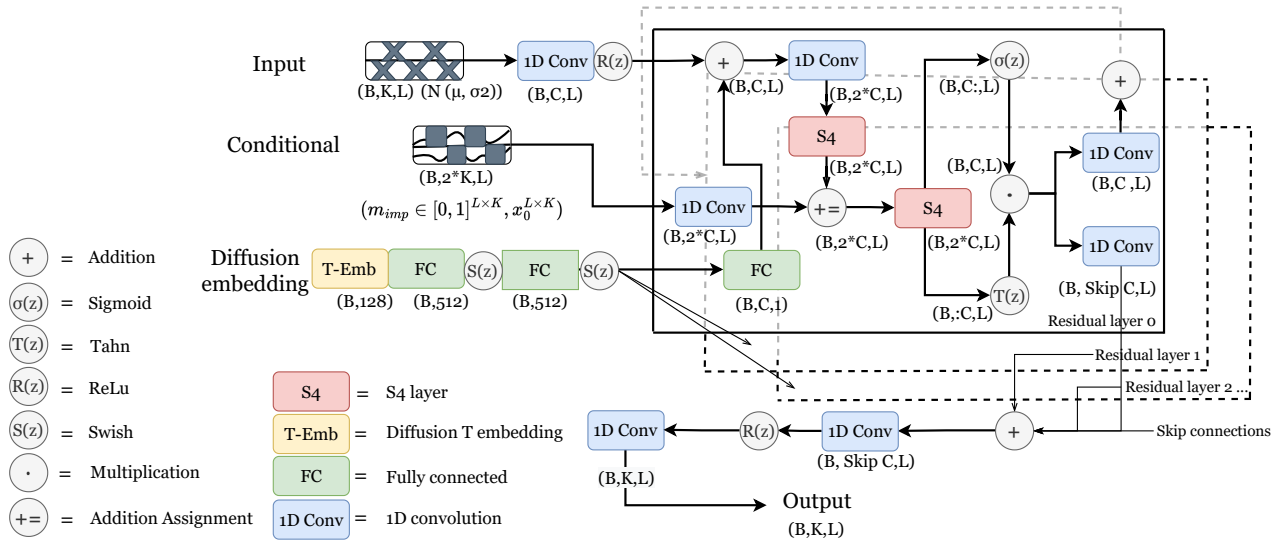


圖 2: SSSD<sup>S4</sup> 模型架構。

## 2.2 空間統計學

### 2.2.1 克利金法

克利金 (Kriging) 方法源自地質學中的空間統計學，是依據觀測資料進行線性內插，用以估計未觀測地點的隨機場值 (N. A. C. Cressie, 1993)。其假設位於  $\mathbf{s}$  的觀測值  $Z(\mathbf{s})$  可表示為：

$$Z(\mathbf{s}) = Y(\mathbf{s}) + \varepsilon(\mathbf{s}), \quad \mathbf{s} \in D \subset \mathbb{R}^D, \quad (18)$$

其中， $Y(\mathbf{s}) = \mu(\mathbf{s}) + \xi(\mathbf{s})$  是隨空間位置變化的線性均值結構； $\varepsilon(\mathbf{s})$  為零均值且與  $Y(\mathbf{s})$  不相關的隨機雜訊，其共變異數函數為  $C(\mathbf{s}, \mathbf{s}') = \text{Cov}(\varepsilon(\mathbf{s}), \varepsilon(\mathbf{s}'))$ ，可為非平穩的空間共變異數函數。

傳統克利金方法依賴對完整共變異數矩陣的逆運算，導致當觀測點數量  $n$  增大時，計算成本呈現指數級增長，形成顯著的計算瓶頸 (N. Cressie and Johannesson, 2008)。固定秩克利金 (Fixed Rank Kriging, FRK) 以及其後的自適應變體 (Adaptive Fixed Rank Kriging, AFRK) 即於此背景下提出，以有效降低大規模空間資料分析的

計算負擔。

### 2.2.2 固定秩克利金法

為降低克利金法的計算負擔，N. Cressie and Johannesson (2008) 提出固定秩克利金法 (Fixed Rank Kriging, FRK)，將隨機場以有限基底函數展開，使高維空間隨機效應近似為低維隨機係數：

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + \mathbf{f}(\mathbf{s})^\top \mathbf{w} + \xi(\mathbf{s}), \quad (19)$$

其中， $\mathbf{f}(\mathbf{s}) = (f_1(\mathbf{s}), \dots, f_K(\mathbf{s}))^\top$  為預先指定的  $K$  維基底函數向量， $K \leq n$ ； $\mathbf{w} \sim N(\mathbf{0}, \mathbf{M})$ ， $\mathbf{M}$  為未知的非負定矩陣； $\xi(\mathbf{s}) \sim \mathcal{N}(0, \sigma_\xi^2)$  為細尺度隨機雜訊。此時共變異數矩陣可表示為：

$$\text{Cov}(Y(\mathbf{s}), Y(\mathbf{s}')) = \mathbf{f}(\mathbf{s})^\top \mathbf{M} \mathbf{f}(\mathbf{s}') + \sigma_\xi^2 \mathbf{I}(\mathbf{s} = \mathbf{s}'). \quad (20)$$

由於  $\mathbf{f} \mathbf{M} \mathbf{f}^\top$  的秩通常遠小於觀測點數量  $n$ ，FRK 可有效降低共變異數矩陣逆運算的計算成本，特別適合處理大規模遙測與環境觀測資料。

### 2.2.3 自適應固定秩克利金法

在 FRK 的基礎上，Tzeng and Huang (2018) 進一步提出自適應固定秩克利金法 (Adaptive Fixed Rank Kriging, AFRK)。其想法是讓基底函數的解析度能夠依據資料分布自動調整，以更靈活地捕捉空間非均勻性，使模型能針對不同區域分配適當的空間解析度。

AFRK 所使用的基底函數為多解析度樣條基底函數 (multi-resolution spline basis functions, MRTS)，其係由薄板樣條 (thin-plate splines, TPS) 所構成。TPS 為一種常見的平滑樣條方法，透過最小化平方誤差與平滑懲罰項以獲得平滑函數 (Wahba and Wendelberger, 1980; Green and Silverman, 1993)。在 TPS 的基礎上，Tzeng and Huang (2018) 進一步透過特徵值分解建構出一組具有不同解析度的有序基底函數，並稱其為 MRTS。為了自動適應資料分布與空間非均勻性，AFRK 可依據特徵值大小自動選擇使用的基底數量，只保留對大部分變異有解釋力的基底，實現以較少的基底函數捕捉主要空間變異、提高計算效率。

在 AFRK 中，MRTS 定義為：

$$f_k(\mathbf{s}) = \begin{cases} 1, & \text{if } k = 1; \\ x_{k-1}, & \text{if } k = 2, \dots, d+1; \\ \lambda_{k-d-1}^{-1} \times \{\phi(\mathbf{s}) - \Phi \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{x}\}' \mathbf{v}_{k-d-1}, & \text{if } k = d+2, \dots, n, \end{cases} \quad (21)$$

其中， $f_k$  是  $\mathbf{f}$  中第  $k$  個基底； $\mathbf{X} \in \mathbb{R}^{N \times (d+1)}$  是設計矩陣，其每一列對應觀測位置  $\mathbf{s}_i$  的截距與座標， $\mathbf{x} = (1, \mathbf{s}')' = (1, x_1, \dots, x_d)'$ ；而  $\Phi$  定義為

$$J(\mathbf{f}) = \boldsymbol{\alpha}' \Phi \boldsymbol{\alpha}, \quad (22)$$

是  $n \times n$  矩陣，其  $(i, j)$  元素為  $\phi_j(\mathbf{s}_i)$ ， $\phi(\mathbf{s})$  定義如下：

$$\phi_i(\mathbf{s}) = \begin{cases} \frac{1}{12} \|\mathbf{s} - \mathbf{s}_i\|^3, & \text{if } d = 1; \\ \frac{1}{8\pi} \|\mathbf{s} - \mathbf{s}_i\|^2 \log(\|\mathbf{s} - \mathbf{s}_i\|), & \text{if } d = 2; \\ -\frac{1}{8} \|\mathbf{s} - \mathbf{s}_i\|, & \text{if } d = 3; \end{cases} \quad (23)$$

而  $\mathbf{v}_k$  為矩陣  $\mathbf{V}$  的第  $k$  列，且  $\mathbf{V} \text{diag}(\lambda_1, \dots, \lambda_n) \mathbf{V}'$  為  $\mathbf{Q}\Phi\mathbf{Q}$  的特徵分解，其中  $\mathbf{Q} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'$ 。此方法能依據資料密度調整基底，在資料較密集的區域產生較高解析度的基底，而在資料稀疏的區域則保持較平滑的結構。

藉由保留低秩近似的計算優勢，同時引入資料導向的多解析度基底，AFRK 能有效處理非均勻採樣與局部變異性顯著的空間資料。相較於傳統 FRK，AFRK 能在非均勻與非平穩的空間資料中展現最佳的預測表現。

## 2.3 時空模型

### 2.3.1 高斯過程

高斯過程 (Gaussian Process, GP) 為一種非參數貝氏模型，用於建立輸入空間至輸出值間之隨機函數關係 (Rasmussen and Williams, 2006)。GP 由均值函數與共變異數函數 (covariance function，或稱核函數 kernel function) 所定義，能在貝氏框架下提供預測分佈之均值與不確定性量化。然而，在標準 GP 中，後驗推論需對規模為  $N \times N$  之共變異數矩陣進行反矩陣運算或 Cholesky 分解，其計算複雜度為  $\mathcal{O}(N^3)$ ，因此難

以直接應用於大規模時空資料情境。

為克服計算瓶頸，Hensman, Fusi, and Lawrence (2013) 提出隨機變分高斯過程 (Stochastic Variational Gaussian Process, SVGP)。該方法引入一組數量為  $M \ll N$  之誘導點 (inducing points) 以近似完整後驗分佈。在 Gardner et al. (2021) 的實作框架中，變分分佈 (variational distribution) 被定義為具備全共變異數矩陣 (full covariance matrix) 之多元常態分佈。為了在優化過程中確保共變異數矩陣之正定性 (positive definite)，實務上將其參數化為均值向量與下三角矩陣 (lower triangle)，即透過 Cholesky 分解進行參數化 (Hensman, Matthews, and Ghahramani, 2014)。

在 SVGP 框架下，變分分佈的大小由誘導點數量決定，即變分均值之維度為  $M$ ，而變分共變異數矩陣之規模為  $M \times M$ 。透過  $M \ll N$  的誘導點以近似完整後驗，並以變分證據下界 (Evidence Lower Bound, ELBO) 作為目標函數，SVGP 將計算複雜度降至  $\mathcal{O}(M^3)$  (Gardner et al., 2021)。

### 2.3.2 時空深度克利金法

時空深度克利金法 (Spatio-temporal DeepKriging, STDK) 為近年提出之結合深度學習與空間統計之非參數化建模方法，主要用於大規模時空資料之內插與機率性預測問題 (Nag, Sun, and Reich, 2023)。相較於傳統高斯過程需事前假設共變異數函數，且在樣本數增加時面臨  $\mathcal{O}(N^3)$  計算複雜度之限制，STDK 採用資料驅動 (data-driven) 方式學習時空相關結構，藉此提升在大規模資料情境下之適用性。

STDK 透過嵌入層 (Embedding Layer) 將時空位置  $(\mathbf{s}, t)$  映射至高維特徵空間。令基底函數向量為

$$\phi(\mathbf{s}, t) = [\phi_1(\mathbf{s}, t), \dots, \phi_K(\mathbf{s}, t)]^\top, \quad (24)$$

其中  $\{\phi_k(\cdot)\}_{k=1}^K$  為具多尺度 (multi-resolution) 特性之基底函數。既有研究多採用具緊支撐性質之溫德蘭函數 (Wendland functions) 或徑向基底函數 (Radial Basis Functions, RBF)，以捕捉不同尺度之空間依賴關係 (Nag, Sun, and Reich, 2023)。

在此基礎上，嵌入後之特徵將輸入深度神經網路以進行非線性映射，其模型形式可表示為

$$Z(\mathbf{s}, t) = f_\theta(\phi(\mathbf{s}, t)) + \epsilon, \quad (25)$$

其中  $f_{\theta}(\cdot)$  為深度神經網路所表示之函數； $\theta$  為模型參數； $\epsilon$  為隨機誤差項。

在預測方面，STDK 為進行機率性預測（probabilistic forecasting），採用分位數損失函數（quantile loss）作為訓練目標，以估計條件分佈之不同分位數，進而構建預測區間。相較於傳統以均方誤差（MSE）為目標之方法，此作法能提供不確定性量化之資訊。

STDK 透過基底函數嵌入與深度神經網路之結合，避免對共變異數矩陣進行顯式建模與分解，使其計算複雜度可近似為  $\mathcal{O}(N)$ ，因此常被用作處理大規模時空資料之基準方法之一。

### 3 研究方法

本研究提出一個結合深度時間序列模型與空間統計建模的時空預測框架，旨在同時利用時間與空間資訊，以提升資料缺失情境下之重建品質與未來時間點之預測準確性。

此整合框架利用 SSSD<sup>S4</sup> 模型建立資料在時間維度上的動態結構，藉此取得資料在時間上的序列特徵，並同時透過 AFRK 使其在模型訓練過程中考慮空間相依性；隨後於推論階段，將 SSSD<sup>S4</sup> 所預測之時間特徵與未觀測地點之空間座標輸入 AFRK，綜合時空雙維度資訊估計未觀測位置之目標值。本節將依序說明問題定義、時空建模方法以及整體演算法流程。

#### 3.1 問題定義

考慮空間域中存在位置集合  $\mathcal{S}$ ，並可將其劃分為具備觀測資料之已觀測位置集合  $\mathcal{S}_{\text{observed}}$ ，以及無任何觀測紀錄之未觀測位置集合  $\mathcal{S}_{\text{unobserved}}$ 。對於已觀測位置  $\mathbf{s} \in \mathcal{S}_{\text{observed}}$ ，其於時間  $t \in \{1, \dots, T\}$  之目標變數均已完整觀測，將其觀測值記為  $y_t(\mathbf{s})$ ；相反地，未觀測位置  $\mathbf{s}^* \in \mathcal{S}_{\text{unobserved}}$  則完全缺乏此時間段內之任何歷史資料。本研究旨在建立一預測模型，利用已觀測位置之歷史時空特徵  $\mathbf{Y}_{1:T}(\mathcal{S}_{\text{observed}})$ ，實現對未觀測位置於未來時間點  $t > T$  之目標變數值  $\hat{y}_t(\mathbf{s}^*)$  的時間外插與空間推估。

此研究目標可形式化為建構一映射函數  $f(\cdot)$

$$\hat{y}_t(\mathbf{s}^*) = f(\mathbf{Y}_{1:T}(\mathcal{S}_{\text{observed}}), \mathbf{s}^*), \quad \mathbf{s}^* \in \mathcal{S}_{\text{unobserved}}, t > T, \quad (26)$$

其中， $\mathbf{Y}_{1:T}(\mathcal{S}_{\text{observed}})$  表示截至時間  $T$ ，由已觀測位置所組成的觀測序列集合； $\hat{y}_t(\mathbf{s}^*)$  則為對未觀測位置  $\mathbf{s}^*$  在未來時間  $t$  的估計值。

## 3.2 時空建模方法

為實現上述映射目標，本研究先以 SSSD<sup>S4</sup> 擷取時間維度之動態依賴結構，再引入 AFRK 將時間特徵與空間座標融合，完成未觀測位置之空間插值與未來預測。

### 3.2.1 基於 SSSD<sup>S4</sup> 之時間建模

為有效捕捉資料於時間序列上的相依結構，本研究採用 SSSD<sup>S4</sup> 作為時間特徵模型。其目的為藉由擴散模型與 S4 層，從不完整或含有雜訊的時間序列中，學習兼具長期依賴與局部動態之潛在時間特徵。

以已觀測位置  $\mathcal{S}_{\text{observed}}$  之觀測資料作為訓練基礎，對資料進行標準化處理，使得

$$\tilde{y}_t(\mathbf{s}) = \frac{y_t(\mathbf{s}) - \mu(\mathbf{s})}{\sigma(\mathbf{s})}, \quad \mathbf{s} \in \mathcal{S}_{\text{observed}}, \quad (27)$$

其中， $\mu(\mathbf{s})$  與  $\sigma(\mathbf{s})$  分別表示位置  $\mathbf{s}$  在整段時間序列上的平均值與標準差，即

$$\mu(\mathbf{s}) = \frac{1}{T} \sum_{t=1}^T y_t(\mathbf{s}), \quad (28)$$

$$\sigma(\mathbf{s}) = \sqrt{\frac{1}{T} \sum_{t=1}^T (y_t(\mathbf{s}) - \mu(\mathbf{s}))^2}. \quad (29)$$

隨後，將標準化後的輸入序列  $\tilde{y}_t(\mathbf{s})$  送入 SSSD<sup>S4</sup> 模型，以學習其時間依賴結構。

依據 Ho, Jain, and Abbeel (2020) 之理論，擴散模型於參數優化階段係利用神經網路輸出之雜訊預測項  $\epsilon_\theta$  近似正向擴散過程中的真實高斯雜訊  $\epsilon$ ，並透過最小化下式更新模型參數

$$\mathcal{L} = \|\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, t)\|^2. \quad (30)$$

本研究於反向擴散過程中，進一步將  $\epsilon_\theta$  依其空間鄰接結構進行調整。具體而言，

於梯度更新前，利用 AFRK 對  $\epsilon_\theta$  於相同空間位置進行預測，以使其在空間層面具備一致性與平滑性，提升反向擴散時的空間結構擬合能力。

$$\hat{\epsilon}_\theta = \text{AFRK}(\epsilon_\theta) = \mu_{\epsilon_\theta} + \mathbf{f}_{\epsilon_\theta}^\top \hat{\mathbf{w}}_{\epsilon_\theta} + \hat{\xi}_{\epsilon_\theta}. \quad (31)$$

經 AFRK 調整之 SSSD<sup>S4</sup> 模型預測輸出可表示為

$$\hat{y}_t(\mathbf{s}) = g_{\theta_{\text{S4+AFRK}}}(\tilde{y}_{1:t}(\mathbf{s})), \quad \mathbf{s} \in \mathcal{S}_{\text{observed}}, \quad (32)$$

其中， $g_{\theta_{\text{S4+AFRK}}}(\cdot)$  為採用 S4 結構並經 AFRK 調整之 SSSD 模型，而  $\hat{y}_t(\mathbf{s})$  為模型於標準化尺度下的預測結果。由於模型係以標準化後之資料進行訓練與預測，因此最終所得之預測結果需透過反標準化轉換回原始資料尺度，其形式為

$$\hat{y}_t(\mathbf{s}) = \sigma(\mathbf{s})\hat{y}_t(\mathbf{s}) + \mu(\mathbf{s}), \quad (33)$$

其中， $\mu(\mathbf{s})$  與  $\sigma(\mathbf{s})$  分別表示位置  $\mathbf{s}$  在整段時間序列上的平均值與標準差。透過此轉換，可將模型輸出還原至原始量測尺度，以利後續分析與空間插值應用。

SSSD 訓練以 MSE 作為損失函數，用以迭代更新參數  $\theta_{\text{S4+AFRK}}$ ，確保模型所萃取之時間特徵具有穩定性與預測效度。於缺失值填補與多步未來預測之任務中，SSSD<sup>S4</sup> 所學得之時間表徵能提供具鑑別能力之序列表達，而 AFRK 於訓練過程中所施加之空間補強則進一步提升整體時空推估之準確性與可靠度。

### 3.2.2 基於 AFRK 之空間建模

在本研究的整合架構中，AFRK 將 SSSD<sup>S4</sup> 所產生之已觀測位置的空間預測值  $\hat{y}_t(\mathbf{s})$ ， $\mathbf{s} \in \mathcal{S}_{\text{observed}}$  視為時間端的輸入資訊。為消除不同位置資料尺度之差異，先對同一時間點  $t$  之預測結果進行空間維度標準化

$$\tilde{y}_t(\mathbf{s}) = \frac{\hat{y}_t(\mathbf{s}) - \mu_t}{\sigma_t}, \quad \mathbf{s} \in \mathcal{S}_{\text{observed}}, \quad (34)$$

其中， $\mu_t$  與  $\sigma_t$  分別為時間點  $t$  於所有已觀測位置  $\mathcal{S}_{\text{observed}}$  之空間平均與標準差。

而後，以標準化後之  $\tilde{y}_t(\mathbf{s})$  作為 AFRK 之輸入，對未觀測位置  $\mathbf{s}^* \in \mathcal{S}_{\text{unobserved}}$  進行

條件推估。於標準化尺度下，未觀測位置的預測可表示為

$$\hat{y}_t(\mathbf{s}^*) = \mu(\mathbf{s}^*) + \mathbf{f}(\mathbf{s}^*)^\top \hat{\mathbf{w}}_t + \hat{\xi}_t(\mathbf{s}^*), \quad (35)$$

其中  $\hat{\mathbf{w}}_t$  與  $\hat{\xi}_t(\mathbf{s}^*)$  係由 AFRK 依據已觀測位置資訊  $\tilde{y}_t(\mathbf{s})$  所估計之空間條件分布參數。

最後，將空間推估結果反標準化，還原至原始量測尺度

$$\hat{y}_t(\mathbf{s}^*) = \sigma_t \hat{y}_t(\mathbf{s}^*) + \mu_t. \quad (36)$$

藉此完成全域時空場之重建與預測。

### 3.3 模型訓練與推論演算法

本節總結上述時空整合模型之運作流程。演算法 1 說明 SSSD<sup>S4</sup> 結合 AFRK 之訓練過程；演算法 2 則展示推論之完整步驟。

---

**Algorithm 1** SSSD-AFRK 訓練階段。

---

**Require:**  $y_t(\mathbf{s}), \mathbf{s} \in \mathcal{S}_{\text{observed}}$ ; Max steps  $T$ ; Learning rate  $\eta$

- 1:  $\tilde{y}_0(\mathbf{s}) = (y_t(\mathbf{s}) - \mu(\mathbf{s})) / \sigma(\mathbf{s})$
  - 2: **repeat**
  - 3:  $t \sim \text{Uniform}(\{1, \dots, T\}), \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$
  - 4:  $\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \tilde{y}_0 + \sqrt{1 - \bar{\alpha}_t} \boldsymbol{\epsilon}$
  - 5:  $\mathbf{h} = \text{S4-Layer}(\mathbf{x}_t, t)$
  - 6:  $\boldsymbol{\epsilon}_\theta = \mathbf{W}_o \mathbf{h} + \mathbf{b}_o$
  - 7:  $\hat{\boldsymbol{\epsilon}}_\theta = \text{AFRK}(\boldsymbol{\epsilon}_\theta) = \mu_{\boldsymbol{\epsilon}_\theta} + \mathbf{f}_{\boldsymbol{\epsilon}_\theta}^\top \hat{\mathbf{w}}_{\boldsymbol{\epsilon}_\theta} + \hat{\xi}_{\boldsymbol{\epsilon}_\theta}$
  - 8:  $\mathcal{L} = \|\boldsymbol{\epsilon} - \hat{\boldsymbol{\epsilon}}_\theta\|^2$
  - 9:  $\theta \leftarrow \theta - \eta \nabla_\theta \mathcal{L}$
  - 10: **until** convergence
- 

其中， $\mathbf{h}$  為經 S4-Layer 處理後之潛在時序特徵； $\mathbf{W}_o$  與  $\mathbf{b}_o$  分別為輸出投影層 (Output Projection Layer) 之可學習權重矩陣與偏差向量。

---

**Algorithm 2** SSSD-AFRK 推論階段。

---

**Require:** Known samples  $y_t(\mathbf{s})$ ; Unknown locations  $\mathcal{S}_{\text{unobserved}} = \{\mathbf{s}_1^*, \dots, \mathbf{s}_m^*\}$ ; Model  $\theta$

- 1: {Stage 1: Temporal Sequence Generation (at  $\mathcal{S}_{\text{observed}}$ )}
- 2:  $\tilde{y}_0(\mathbf{s}) = (y_t(\mathbf{s}) - \mu(\mathbf{s}))/\sigma(\mathbf{s})$
- 3:  $\mathbf{x}_T \sim \mathcal{N}(0, \mathbf{I})$
- 4: **for**  $t = T, \dots, 1$  **do**
- 5:    $\mathbf{h} = \text{S4-Layer}(\mathbf{x}_t, t)$
- 6:    $\boldsymbol{\epsilon}_\theta = \mathbf{W}_o \mathbf{h} + \mathbf{b}_o$
- 7:    $\mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}} \left( \mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\alpha_t}} \boldsymbol{\epsilon}_\theta \right) + \sigma_t \mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(0, \mathbf{I})$
- 8: **end for**
- 9:  $\hat{y}_t(\mathbf{s}) = \sigma(\mathbf{s}) \mathbf{x}_0 + \mu(\mathbf{s}), \quad \mathbf{s} \in \mathcal{S}_{\text{observed}}$
- 10: {Stage 2: Spatial Interpolation via AFRK (at  $\mathcal{S}_{\text{unobserved}}$ )}
- 11:  $\tilde{y}_t(\mathbf{s}) = (\hat{y}_t(\mathbf{s}) - \mu_t)/\sigma_t$
- 12:  $\hat{y}_t(\mathbf{s}^*) = \text{AFRK}(\tilde{y}_t(\mathbf{s}) \mid \mathbf{s}^*), \quad \forall \mathbf{s}^* \in \mathcal{S}_{\text{unobserved}}$
- 13:  $\hat{y}_t(\mathbf{s}^*) = \sigma_t \hat{y}_t(\mathbf{s}^*) + \mu_t$
- 14: **return** Full field  $\hat{y}_t(\mathcal{S}_{\text{observed}} \cup \mathcal{S}_{\text{unobserved}})$

---

## 4 實驗

本章介紹時空整合框架的實驗設計與設定，用於評估其在未觀測地點之未來時空預測任務中的效能與可行性。實驗採用兩個主要資料集，並詳細說明模型設定、訓練與推理流程，以及資料集的特性與分割方式，以呈現框架在不同資料條件下的表現。

### 4.1 資料集

本研究選用兩個具代表性的氣象資料集，分別涵蓋地面觀測站尺度與全球再分析尺度；以及一個模擬資料集，以評估模型於不同時空動態條件下的適用性與泛化能力。

#### 4.1.1 Weather2K

Weather2K 是近期提出的多變量地面觀測基準資料集，收錄中國境內數千個氣象站的觀測資料，時間解析度為 3 小時，涵蓋氣溫、氣壓、濕度、風速等近地表氣象變數 (Zhu et al., 2023)。開源版本 Weather2K-R 包含 1,866 個氣象站及 13,632 個連續時間步，時間序列完整且定期採樣，無缺失值。資料同時提供位置常數（經緯度與海拔），

方便時空模式分析。

Weather2K 資料集呈現空間分佈異質及測站密度不均，涵蓋都市、平原與山區等多種地理環境與氣候條件，提供豐富的時空變化訊號。這些特性使 Weather2K 適合用於時空插值、短期序列預測，以及模型泛化能力與穩健性評估。

本研究選取 Weather2K-R 資料集中自 2021 年 07 月 01 日 00 時至 2021 年 08 月 31 日 21 時之觀測序列。實驗選取 200 個觀測點，並選取氣溫 (Air Temperature) 變數作為資料來源。詳細變數說明參閱附錄 A。

#### 4.1.2 MERRA-2

MERRA-2 (Modern-Era Retrospective Analysis for Research and Applications, Version 2) 由 NASA Goddard Earth Sciences Data and Information Services Center (GES DISC) 提供，是一套全球大氣再分析資料集，透過數值天氣預報模式與以衛星觀測為主的多源觀測資料，自 1980 年起以資料同化技術重建全球大氣狀態 (GMAO, 2015)。本研究採用其中逐時、單層、瞬時同化診斷資料產品 M2I1NXASM (Version 5.12.4) 作為分析資料集。

MERRA-2 由 NASA Goddard Space Flight Center 所屬的 Distributed Active Archive Center (DAAC) 典藏與管理，提供全球尺度且高一致性的再分析資料。其資料同化流程整合衛星、地面及遙測觀測，並在氣膠、輻射收支及水文循環等物理過程上較前代 MERRA 有顯著改進，廣泛應用於氣候趨勢分析、極端事件研究、能量平衡診斷與數值模式評估。M2I1NXASM 為逐時瞬時資料產品，提供高解析度且具代表性的氣候訊號，適合用於短期時空預測與統計特徵分析。

本研究自 MERRA-2 之 M2I1NXASM (Version 5.12.4) 資料產品中選取自 2023 年 12 月 11 日 00 時至 2023 年 12 月 31 日 23 時，與 Surface Skin Temperature 變數作為資料來源。詳細變數說明參閱附錄 B。實驗區域界定於北緯 26.0° 至 48.0° 與西經 70.0° 至 123.0° 之矩形空間範圍內，並選取該區域內 200 個觀測點作為資料來源。

#### 4.1.3 2b-8

2b-8 資料集源自第二屆空間統計競賽 (The Second Competition on Spatial Statistics)，旨在探討大規模空間與時空數據分析中的預測精度與計算效率議題 (Abdulah, Alamri, Nag, et al., 2022)。該競賽資料集提供了一系列經過嚴格採樣與模擬之空間數據，專門

設計用於評估複雜時空模型的插值效能，特別是在面對高維度、大規模且具備複雜空間相關性之數據情境下的表現。

本研究選取 2b-8 資料集 (Abdulah, Alamri, Ltaief, et al., 2022) 中 200 個觀測點作為資料來源。詳細資料集說明參閱附錄 C。

## 4.2 實驗環境與計算資源

為確保所提出之時空整合框架在大規模資料上的訓練與推理實驗具有可行性與可靠性，本研究建立統一的運算環境，並整合多種現有軟體套件以支援模型開發、訓練及評估。

SSSD 已由原作者以 Python 實現 (AI4HealthUOL, 2023)，可有效捕捉時間序列資料中的長程依賴結構。AFRK 則由 Wen-Ting Wang 以 R 語言實現，命名為 autoFRK (Tzeng, Huang, Wang, Nychka, et al., 2021)，提供穩定且可擴展的空間插值能力。為整合上述功能，本研究將 autoFRK 重新實現並封裝為 Python 套件 (Tzeng, Huang, Wang, and Hsu, 2025)，並將其演算法與 SSSD 原始碼結合，在 Python 的 PyTorch 框架下完成完整的時空整合運算流程。

所有實驗，包括模型訓練、驗證與推理，皆在 Taiwania 2 高效能運算平臺 (NCHC, 2018) 上執行。Taiwania 2 提供高效能 GPU 運算資源，以及大容量記憶體與高速儲存系統，使本研究能夠有效處理高解析度資料與長時間序列，同時確保運算的穩定性與結果的重現性。

## 4.3 實驗設計

為清楚呈現 SSSD<sup>S4</sup> 與 autoFRK 的設定，以下將訓練與推理的超參數整理如表 1。

表 1: 模型超參數設定。

模型	超參數	數值
<b>Training Configuration</b>		
	Batch size	40
	Learning rate	0.001

模型	超參數	數值
	Only generate missing	True
	Masking	Forecast
	Missing $k$	—
<b>SSSD<sup>S4</sup></b>		
WaveNet	Input channels	1
	Output channels	1
	Residual layers	32
	Residual channels	64
	Skip channels	64
Diffusion step embedding	Input dimension	128
	Hidden dimension	256
	Output dimension	256
S4	Max sequence length	—
	State dimension	128
	Dropout	0.1
	Bidirectional	True
	Layer normalization	True
Diffusion	Diffusion steps ( $T$ )	100
	$\beta_0$	0.0001
	$\beta_T$	0.05
<b>autoFRK</b>		
	Method	Fast
	Thin-plate spline method	Rectangular

表 2 列出實驗之參數與資料集設定，用於比較各模型之表現。輸入序列表示為  $\mathbf{X} \in \mathbb{R}^{N \times T \times C}$ ，其中  $N$  代表空間觀測點數量 (Number of spatial locations)、 $T$  為時間序列長度 (Time sequence length)、 $C$  則對應輸入變數維度或通道數 (Input channels)。資料集中的訓練集與測試集時間步比例分別為 0.9 與 0.1，而觀測地點與未觀測地點比例則為 0.8 與 0.2。

實驗設計包含訓練階段是否整合 AFRK，透過對比基準模型 SSSD<sup>S4</sup> 與訓練階段整合 AFRK 之 SSSD<sup>S4+AFRK</sup> 模型，評估於訓練過程中加入 AFRK 是否有助於增強空間依賴性，並比較兩者於推論階段對未知觀測點之填補影響。此外，本研究亦納入 TFT、VAR、SVGP 與 STDK 等模型，作為多元基準以評估不同模型架構與訓練策略在時空預測任務上的相對表現。

表 2: 實驗參數與資料集設定 (Weather2K / MERRA-2 / 2b-8)。

參數	數值
迭代次數	500
已觀測位置	160
未觀測位置	40
時序長度	448 / 456 / 90
缺失時序 $k$	48 / 48 / 10

資料集的已觀測與未觀測測站站點分布如圖 3、圖 4 與圖 5 所示，其中經抽樣後，藍色圓點表示已觀測位置，綠色三角形表示未觀測位置。

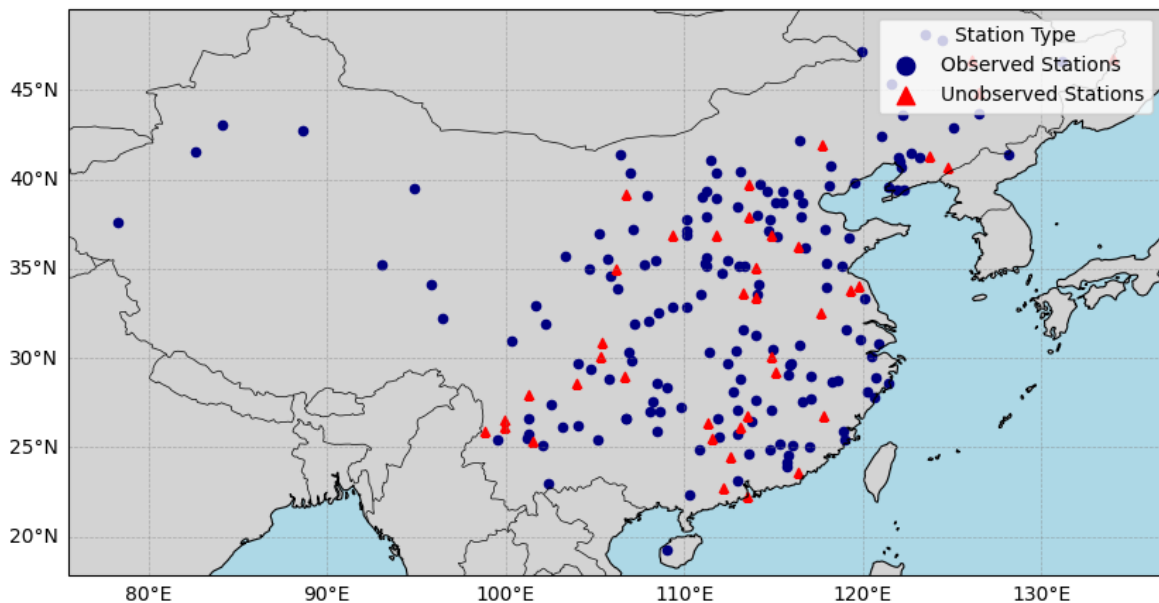


圖 3: Weather2K 3-Hourly 資料站點分布圖。

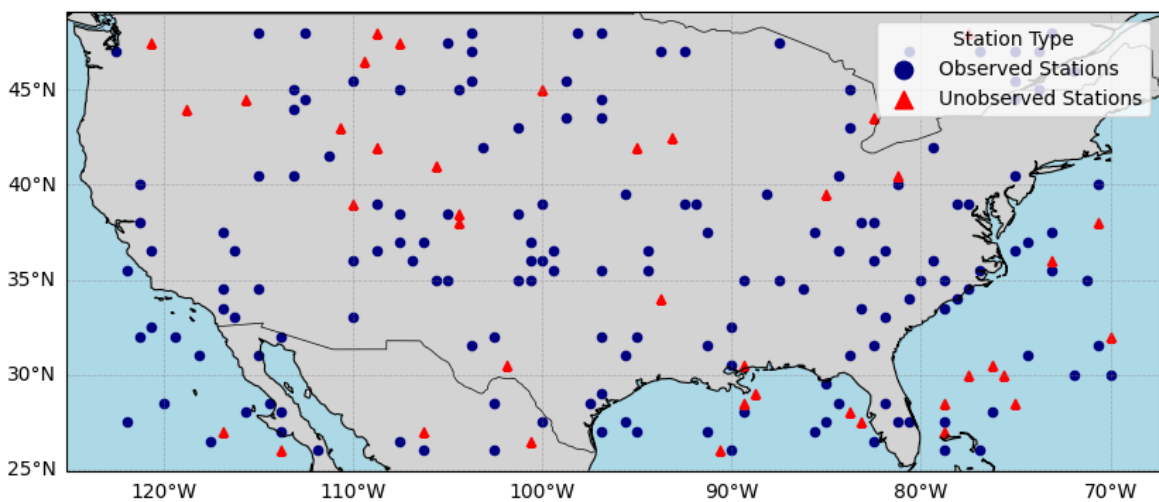


圖 4: MERRA-2 Hourly 資料站點分布圖。

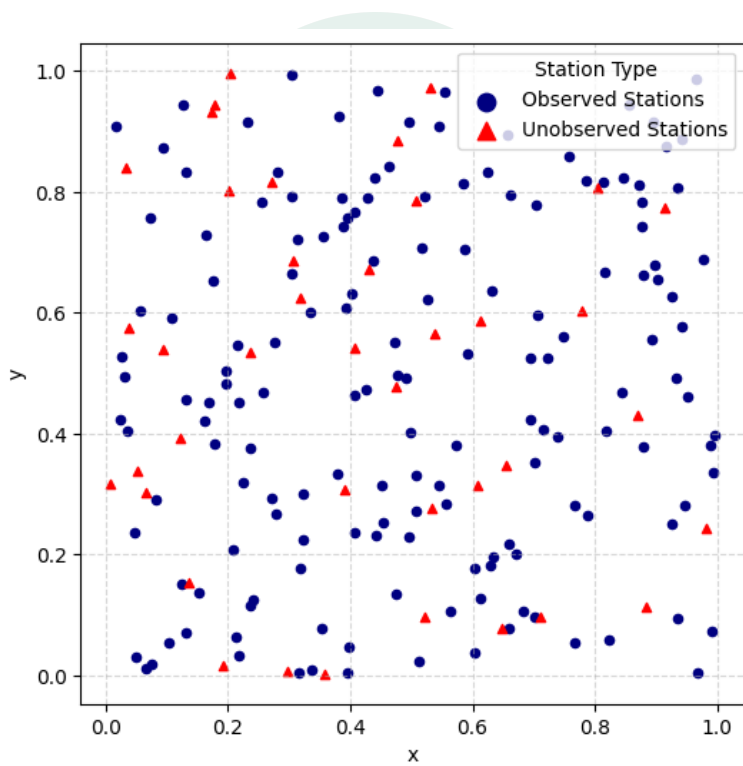


圖 5: 2b-8 資料站點分布圖。

## 5 實驗結果

本研究採用平均平方預測誤差 (Mean Squared Prediction Error, MSPE) 作為主要評估指標，用以衡量模型的預測精準度。表 3 呈現以  $SSSD^{S4+AFRK}$  與  $SSSD^{S4}$  訓練之模型，並於推論階段使用 AFRK 進行空間欄位填補之整合式時空框架，在不同資料集上進行 30 次實驗之平均 MSPE 表現，同時列出 AFRK 在推論階段選取之 MRTS 基底數量。表中亦包含多項基準模型之結果，部分模型僅可預測時間序列，故使用 AFRK 填補未觀測位置，用以比較各模型在時空預測任務中的相對效能。

表 3: 各模型於不同資料集與預測情境下之 MSPE 與推論階段選取之 MRTS 基底數量。

訓練策略	$SSSD^{S4+AFRK}$	$SSSD^{S4}$	TFT	VAR	SVGP	STDK
<b>Weather2K</b>						
未觀測地點未來	<b>19.1587</b>	19.2287	22.2616	30.1737	31.1692	30.5259
未觀測地點過去	4.0865	<b>4.0800</b>	4.0913	4.0967	20.5138	23.0728
已觀測地點未來	<b>15.0609</b>	15.0880	19.1520	23.4325	29.7829	38.8186
# of MRTS Basis	123	123	123	123	—	—
<b>MERRA-2</b>						
未觀測地點未來	<b>10.0391</b>	10.2388	12.8345	235.2112	29663.98	178.5403
未觀測地點過去	6.7505	<b>6.7384</b>	6.7403	6.7460	18013.08	174.8610
已觀測地點未來	<b>7.9678</b>	8.1248	13.8354	331.8714	27667.48	112.6103
# of MRTS Basis	123	123	123	123	—	—
<b>2b-8</b>						
未觀測地點未來	0.8761	0.8344	0.9071	1.0477	0.9153	<b>0.8253</b>
未觀測地點過去	0.7835	<b>0.7834</b>	0.7843	0.7846	0.8601	0.8809
已觀測地點未來	1.1884	1.0674	1.6772	2.4342	0.9362	<b>0.9093</b>
# of MRTS Basis	23	23	23	23	—	—

表 3 顯示， $SSSD^{S4+AFRK}$  與  $SSSD^{S4}$  兩種設定在不同資料集上之表現具有一致性差異，其中加入 AFRK 之模型在多數未觀測地點之未來預測任務中取得較低之 MSPE，顯示 AFRK 對於空間結構之補償能力，有助於提升模型在空間外插 (spatial

extrapolation) 情境下之預測表現。

就不同資料集進行分析，在 Weather2K 資料集中， $SSSD^{S4+AFRK}$  於未觀測地點之未來預測任務表現改善最為明顯，反映在高時空變異性資料下的精準度提升。

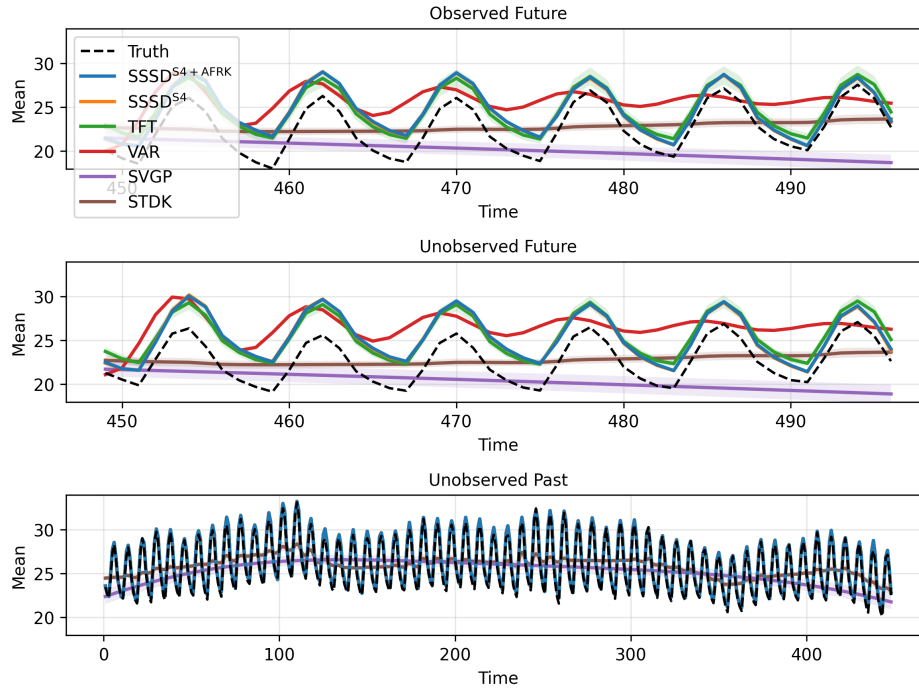


圖 6: Weather2K 資料集於不同預測情境下之預測平均值比較。

圖 6 與圖 7 展示了在已觀測地點未來、未觀測地點未來、未觀測地點過去情境下之時間序列預測平均值與 MSPE 比較，以及各模型的 95% 信賴區間。結果顯示，在未觀測地點未來與未觀測地點過去兩類任務中， $SSSD^{S4+AFRK}$  相對其他基線模型 (如 TFT、VAR、SVGP 與 STDK) 能更貼近真值並取得較低之 MSPE，顯示其在捕捉時空動態變化方面具有優勢。相較之下，在已觀測地點之預測任務中， $SSSD^{S4}$  與  $SSSD^{S4+AFRK}$  兩種模型之預測結果相似，反映 AFRK 於  $SSSD^{S4+AFRK}$  主要貢獻於模型對未觀測區域之學習與補償能力。

在 MERRA-2 資料集中， $SSSD^{S4+AFRK}$  模型同樣在未觀測地點未來預測上取得較佳 MSPE，且其誤差改善幅度相較 Weather2K 更為穩定，顯示 AFRK 在較大尺度與較高維度氣候資料中仍具一致性之空間建模能力。此外，在已觀測地點之預測任務中， $SSSD^{S4+AFRK}$  亦相較於  $SSSD^{S4}$  呈現誤差下降，顯示 AFRK 對於空間結構之建模不僅有助於未觀測區域之推論，亦能提升已觀測區域之預測表現。

圖 8 與圖 9 已移除於表 3 中表現較差的 SVGP 模型。由圖 8 可見， $SSSD^{S4}$ 、

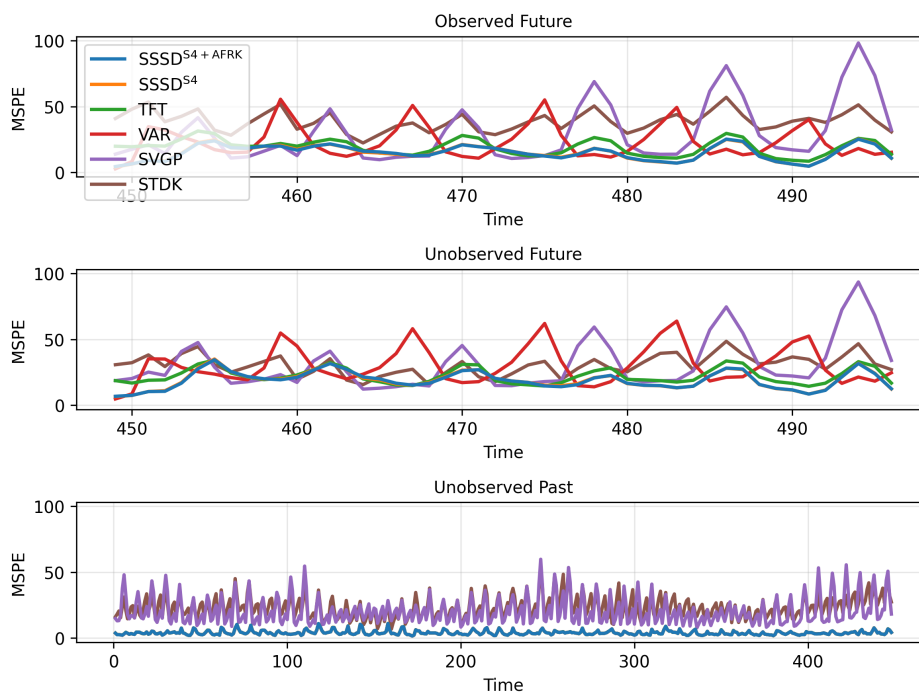


圖 7: Weather2K 資料集於不同預測情境下之預測 MSPE 比較。

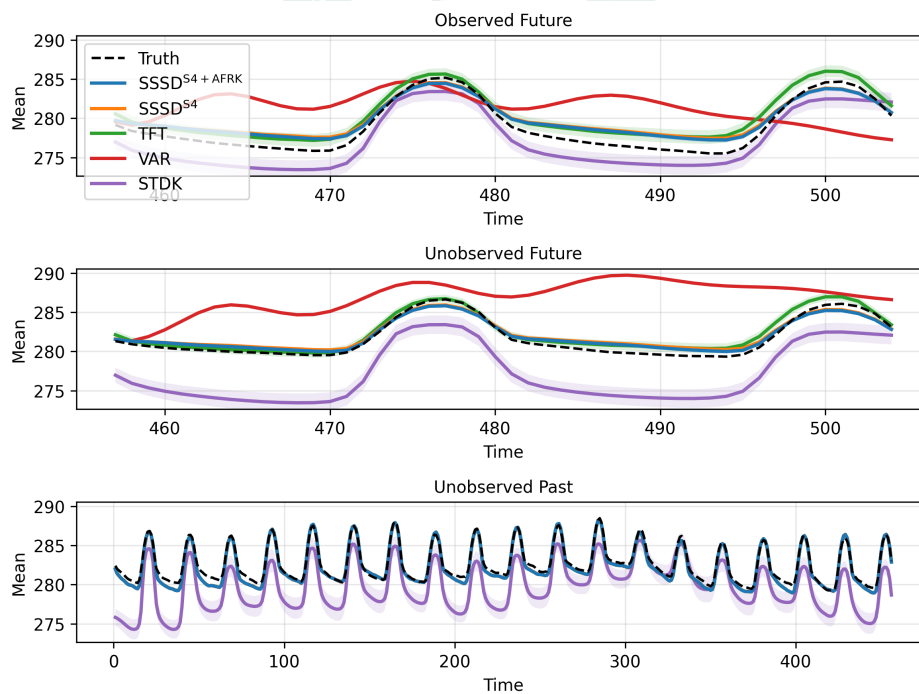


圖 8: MERRA-2 資料集於不同預測情境下之預測平均值比較。

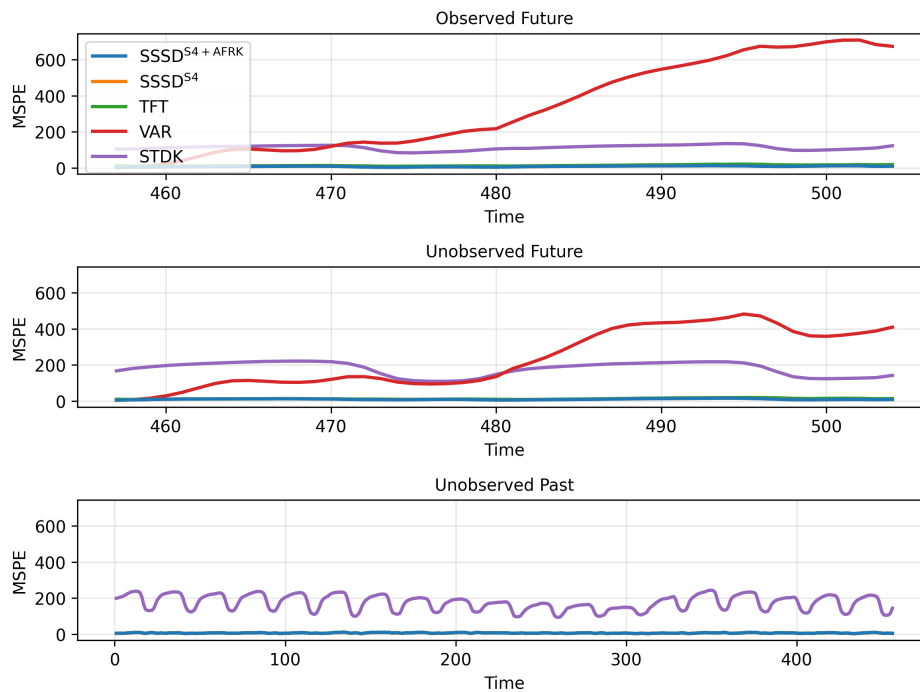


圖 9: MERRA-2 資料集於不同預測情境下之預測 MSPE 比較。

SSSD<sup>S4+AFRK</sup> 與 TFT 三種模型經由 AFRK 進行空間插值後，在不同預測情境下皆能貼近真值。而從未來預測的後段預測結果也能發現，SSSD<sup>S4+AFRK</sup> 的預測平均值較 TFT 更為貼近真值，顯示 SSSD<sup>S4+AFRK</sup> 在捕捉長距離時空依賴關係方面具有優勢。

從圖 10 可見，SSSD<sup>S4+AFRK</sup> 在未觀測地點未來預測任務中取得較低之 MSPE，且其誤差曲線多數穩定低於 SSSD<sup>S4</sup> 與 TFT，顯示在高維氣候資料上 AFRK 所建構之多層次空間基底，能有效協助主模型捕捉長距離時空依賴關係。

對於 2b-8 模擬資料集而言，由於其空間變異相對平滑，各模型間整體性能相近。在未觀測地點未來預測情境下，STDK 取得了最優異的 MSPE，顯示在規律性高且低變異性資料上，專門設計之深度時空卷積模型能有效泛化。而 SSSD<sup>S4+AFRK</sup> 與 SSSD<sup>S4</sup> 之差異則非常有限，表示 AFRK 在此資料集上的邊際效益較小。

整體而言，在空間結構簡單且規律的模擬資料下，AFRK 雖然無顯著優勢，仍能達到與其他空間建模模型相當的準確度。

在各基準模型中，TFT 在純時序預測任務中仍展現出一定競爭力，且誤差波動較平緩。VAR 模型在部分資料集（如 Weather2K）能提供合理預測，但面對高度時空變異（如 MERRA-2）時，其誤差明顯放大。STDK 與 SVGP 雖具備理論之空間建模能

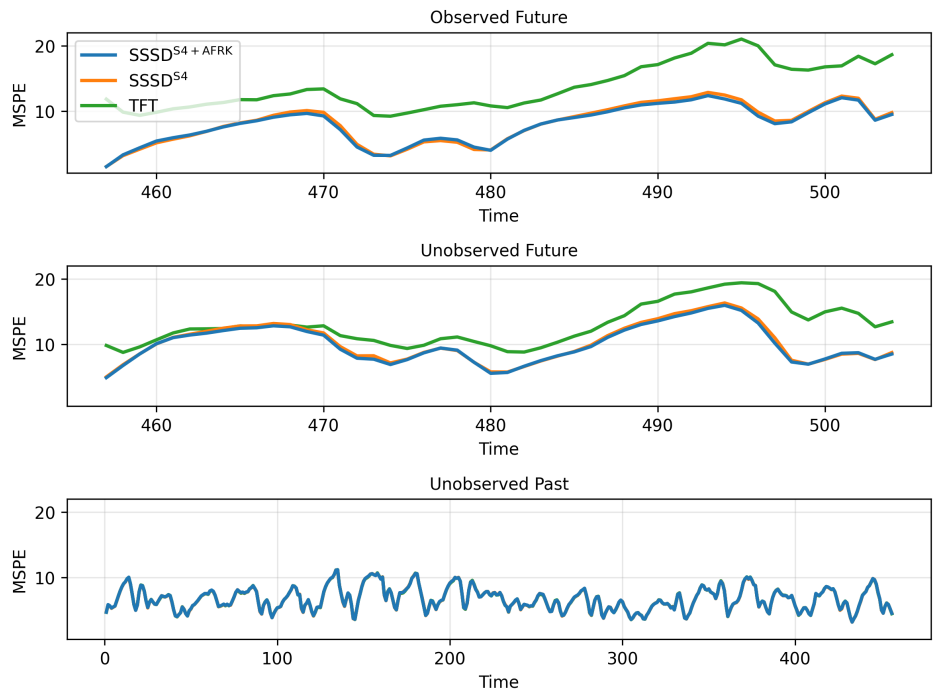


圖 10: MERRA-2 資料集於不同預測情境下 SSSD<sup>S4</sup>、SSSD<sup>S4</sup> + AFRK 與 TFT 之 MSPE 比較。

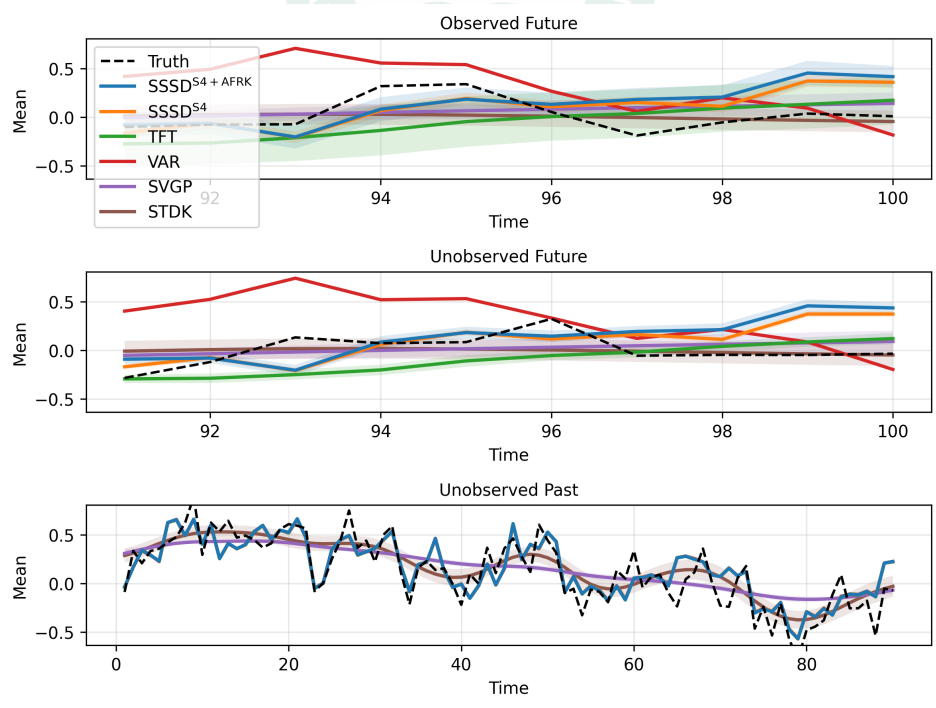


圖 11: 2b-8 資料集於不同預測情境下之預測平均值比較。

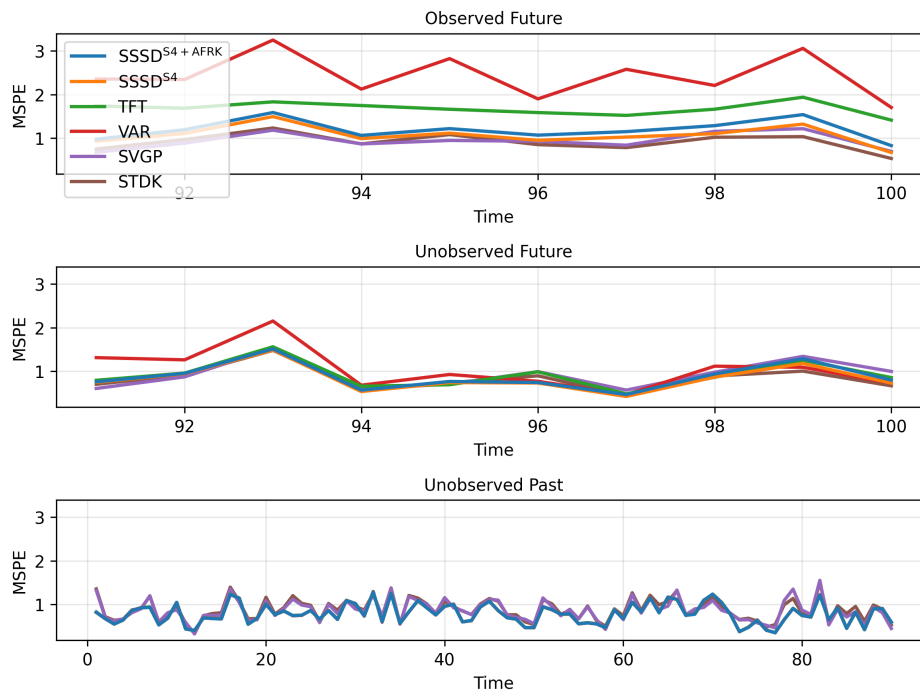


圖 12: 2b-8 資料集於不同預測情境下之預測 MSPE 比較。

力，但在高維時空預測情境下穩定性不足，例如 SVGP 在 Weather2K 與 MERRA-2 資料中預測曲線過度平滑且誤差大幅飆升，顯示其核方法在大規模資料上的近似推論難以有效收斂。

綜合以上，本結果充分證實本研究提出之  $SSSD^{S4+AFRK}$  時空整合框架，相較於僅考慮時間或空間單一結構之模型能更精準地捕捉真實環境資料中的複雜空間依賴性，並在不同資料分佈與變異性條件下保持穩定低誤差。

## 6 結論

本研究針對具地理空間特徵之時序資料提出一種整合式時空預測框架  $SSSD^{S4+AFRK}$ ，透過結合深度狀態空間模型與空間基底函數，有效提升對未觀測位置之預測能力。實驗結果顯示，加入 AFRK 的模型在處理空間外插（未觀測地點未來預測）任務上，平均平方預測誤差顯著降低。同時，其對已觀測地點未來與未觀測地點過去任務亦具有一定提升，證明 AFRK 不僅擅長捕捉缺失空間資訊，亦能強化主模型之整體預測精度。與基準模型比較，本方法在高維度與高變異資料（如 Weather2K、MERRA-2）下表現更為優越且穩定，凸顯時空聯合建模之必要性。未來可進一步拓展 AFRK 的設計，或引入其他時空特徵，以進一步優化複雜環境中之預測效能。

## 參考文獻

- Abdulah, Sameh, Faten Alamri, Hatem Ltaief, et al. (2022). *Data for "The Second Competition on Spatial Statistics for Large Datasets"*. KAUST Research Repository. DOI: 10.25781/KAUST-4ADYZ. URL: <https://doi.org/10.25781/KAUST-4ADYZ>.
- Abdulah, Sameh, Faten Alamri, Pratik Nag, et al. (2022). *The Second Competition on Spatial Statistics for Large Datasets*. arXiv: 2211.03119 [stat.OT]. URL: <https://arxiv.org/abs/2211.03119>.
- AI4HealthUOL (2023). *SSSD: Official Implementation of Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models*. <https://github.com/AI4HealthUOL/SSSD>. Accessed: 2025-11-28.
- Alcaraz, Juan Miguel Lopez and Nils Strodthoff (2023). *Diffusion-based Time Series Imputation and Forecasting with Structured State Space Models*. arXiv: 2208.09399 [cs.LG]. URL: <https://arxiv.org/abs/2208.09399>.
- Bengio, Y., P. Simard, and P. Frasconi (1994). "Learning long-term dependencies with gradient descent is difficult". In: *IEEE Transactions on Neural Networks* 5.2, pp. 157–166. DOI: 10.1109/72.279181.
- Box, George E. P. and Gwilym M. Jenkins (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day series in time series analysis and digital processing. Holden-Day. ISBN: 9780816211043. URL: <https://books.google.com.tw/books?id=1WVHAAAAMAAJ>.
- Cressie, Noel and Gardar Johannesson (2008). "Fixed rank kriging for very large spatial data sets". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 70.1, pp. 209–226. DOI: <https://doi.org/10.1111/j.1467-9868.2007.00633.x>. eprint: <https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9868.2007.00633.x>. URL: <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2007.00633.x>.
- Cressie, Noel and Christopher K. Wikle (2011). *Statistics for Spatio-Temporal Data*. CourseSmart Series. Wiley. ISBN: 9780471692744. URL: <https://books.google.com.tw/books?id=-k0C6D0DiNYC>.
- Cressie, Noel A. C. (1993). *Statistics for Spatial Data*. John Wiley & Sons, Inc. DOI: 10.1002/9781119115151. URL: <https://doi.org/10.1002/9781119115151>.
- Decorte, Thomas et al. (2024). "Missing Value Imputation of Wireless Sensor Data for Environmental Monitoring". In: *Sensors* 24.8. ISSN: 1424-8220. DOI: 10.3390/s24082416. URL: <https://www.mdpi.com/1424-8220/24/8/2416>.

- Gardner, Jacob R. et al. (2021). *GPyTorch: Blackbox Matrix-Matrix Gaussian Process Inference with GPU Acceleration*. arXiv: 1809.11165 [cs.LG]. URL: <https://arxiv.org/abs/1809.11165>.
- Global Modeling and Assimilation Office (GMAO) (2015). *MERRA-2 inst1\_2d\_asm\_Nx: 2d, 1-Hourly, Instantaneous, Single-Level, Assimilation, Single-Level Diagnostics V5.12.4*. Accessed: 2025-11-01. Greenbelt, MD, USA. DOI: 10.5067/3Z173KIE2TPD.
- Gneiting, Tilmann (2002). “Nonseparable, Stationary Covariance Functions for Space-Time Data”. In: *Journal of the American Statistical Association* 97.458, pp. 590–600. ISSN: 01621459. URL: <http://www.jstor.org/stable/3085674> (visited on 04/17/2026).
- Green, P. J. and B. W. Silverman (1993). *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. 1st ed. Chapman and Hall/CRC. DOI: 10.1201/b15710.
- Gu, Albert, Tri Dao, et al. (2020). *HiPPO: Recurrent Memory with Optimal Polynomial Projections*. arXiv: 2008.07669 [cs.LG]. URL: <https://arxiv.org/abs/2008.07669>.
- Gu, Albert, Karan Goel, and Christopher Ré (2022). *Efficiently Modeling Long Sequences with Structured State Spaces*. arXiv: 2111.00396 [cs.LG]. URL: <https://arxiv.org/abs/2111.00396>.
- Hensman, James, Nicolo Fusi, and Neil D. Lawrence (2013). *Gaussian Processes for Big Data*. arXiv: 1309.6835 [cs.LG]. URL: <https://arxiv.org/abs/1309.6835>.
- Hensman, James, Alex Matthews, and Zoubin Ghahramani (2014). *Scalable Variational Gaussian Process Classification*. arXiv: 1411.2005 [stat.ML]. URL: <https://arxiv.org/abs/1411.2005>.
- Ho, Jonathan, Ajay Jain, and Pieter Abbeel (2020). *Denoising Diffusion Probabilistic Models*. arXiv: 2006.11239 [cs.LG]. URL: <https://arxiv.org/abs/2006.11239>.
- Hochreiter, Sepp and Jürgen Schmidhuber (Nov. 1997). “Long Short-Term Memory”. In: *Neural Computation* 9.8, pp. 1735–1780. ISSN: 0899-7667. DOI: 10.1162/neco.1997.9.8.1735. eprint: <https://direct.mit.edu/neco/article-pdf/9/8/1735/813796/neco.1997.9.8.1735.pdf>. URL: <https://doi.org/10.1162/neco.1997.9.8.1735>.
- Kalman, R. E. (Mar. 1960). “A New Approach to Linear Filtering and Prediction Problems”. In: *Journal of Basic Engineering* 82.1, pp. 35–45. ISSN: 0021-9223. DOI: 10.1115/1.3662552. eprint: [https://asmedigitalcollection.asme.org/fluidsengineering/article-pdf/82/1/35/5518977/35\\_1.pdf](https://asmedigitalcollection.asme.org/fluidsengineering/article-pdf/82/1/35/5518977/35_1.pdf). URL: <https://doi.org/10.1115/1.3662552>.

- Kong, Zhifeng et al. (2021). *DiffWave: A Versatile Diffusion Model for Audio Synthesis*. arXiv: 2009.09761 [eess.AS]. URL: <https://arxiv.org/abs/2009.09761>.
- Lim, Bryan et al. (2020). *Temporal Fusion Transformers for Interpretable Multi-horizon Time Series Forecasting*. arXiv: 1912.09363 [stat.ML]. URL: <https://arxiv.org/abs/1912.09363>.
- Little, Roderick J. A. and Donald B. Rubin (2002). *Statistical Analysis with Missing Data*. John Wiley & Sons, Inc. DOI: 10.1002/9781119013563. URL: <https://doi.org/10.1002/9781119013563>.
- Nag, Pratik, Ying Sun, and Brian J Reich (2023). *Spatio-temporal DeepKriging for Interpolation and Probabilistic Forecasting*. arXiv: 2306.11472 [stat.ML]. URL: <https://arxiv.org/abs/2306.11472>.
- National Center for High-Performance Computing (NCHC) (2018). *Taiwania 2 / Taiwan Computing Cloud (TWCC) High-Performance AI Cloud Platform*. National Center for High-Performance Computing, Taiwan. Supercomputer: Taiwania 2; 9 PFLOPS, 2,016 NVIDIA Tesla V100 GPUs; Operated via TWCC.
- Primiceri, Giorgio E. (July 2005). “Time Varying Structural Vector Autoregressions and Monetary Policy”. In: *The Review of Economic Studies* 72.3, pp. 821–852. ISSN: 0034-6527. DOI: 10.1111/j.1467-937X.2005.00353.x. eprint: <https://academic.oup.com/restud/article-pdf/72/3/821/18344172/72-3-821.pdf>. URL: <https://doi.org/10.1111/j.1467-937X.2005.00353.x>.
- Rasmussen, Carl Edward and Christopher K. I. Williams (2006). *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press. URL: <https://gaussianprocess.org/gpml/>.
- Shi, Xingjian et al. (2015). “Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting”. In: *Advances in Neural Information Processing Systems*. Ed. by C. Cortes et al. Vol. 28. Curran Associates, Inc. URL: [https://proceedings.neurips.cc/paper\\_files/paper/2015/file/07563a3fe3bbe7e3ba84431ad9d055af-Paper.pdf](https://proceedings.neurips.cc/paper_files/paper/2015/file/07563a3fe3bbe7e3ba84431ad9d055af-Paper.pdf).
- Sims, Christopher A. (1980). “Macroeconomics and Reality”. In: *Econometrica* 48.1, pp. 1–48. DOI: 10.2307/1912017.
- Sohl-Dickstein, Jascha et al. (2015). *Deep Unsupervised Learning using Nonequilibrium Thermodynamics*. arXiv: 1503.03585 [cs.LG]. URL: <https://arxiv.org/abs/1503.03585>.

- Tzeng, ShengLi and Hsin-Cheng Huang (2018). “Resolution Adaptive Fixed Rank Kriging”. In: *Technometrics* 60.2, pp. 198–208. DOI: 10.1080/00401706.2017.1345701. eprint: <https://doi.org/10.1080/00401706.2017.1345701>. URL: <https://doi.org/10.1080/00401706.2017.1345701>.
- Tzeng, ShengLi, Hsin-Cheng Huang, Wen-Ting Wang, and Yao-Chih Hsu (2025). *autoFRK-python: Automatic Fixed Rank Kriging. The Python version with PyTorch*. Python package version 1.2.3. URL: <https://pypi.org/project/autoFRK/>.
- Tzeng, ShengLi, Hsin-Cheng Huang, Wen-Ting Wang, Douglas Nychka, et al. (2021). *autoFRK: Automatic Fixed Rank Kriging*. R package version 1.4.3. URL: <https://CRAN.R-project.org/package=autoFRK>.
- Wahba, Grace and James Wendelberger (1980). “Some New Mathematical Methods for Variational Objective Analysis Using Splines and Cross Validation”. In: *Monthly Weather Review* 108.8, pp. 1122–1143. DOI: 10.1175/1520-0493(1980)108<1122:SNMMFV>2.0.CO;2. URL: [https://journals.ametsoc.org/view/journals/mwre/108/8/1520-0493\\_1980\\_108\\_1122\\_snmmfv\\_2\\_0\\_co\\_2.xml](https://journals.ametsoc.org/view/journals/mwre/108/8/1520-0493_1980_108_1122_snmmfv_2_0_co_2.xml).
- Wu, Haixu et al. (2022). *Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting*. arXiv: 2106.13008 [cs.LG]. URL: <https://arxiv.org/abs/2106.13008>.
- Zhu, Xun et al. (2023). *Weather2K: A Multivariate Spatio-Temporal Benchmark Dataset for Meteorological Forecasting Based on Real-Time Observation Data from Ground Weather Stations*. arXiv: 2302.10493 [cs.LG]. URL: <https://arxiv.org/abs/2302.10493>.

## A Weather2K 資料集

Weather2K-R 資料集所包含之變數如下。本研究使用其地表溫度 (Air Temperature) 變數作為主要分析對象。

表 4: Weather2K 變數表。

變數名稱	縮寫	單位
Latitude	lat	degrees east
Longitude	lon	degrees north
Altitude	alt	m
Air Pressure	ap	hPa
Air Temperature	t	°C
Maximum Temperature	mxt	°C
Minimum Temperature	mnt	°C
Relative Humidity	rh	%
Precipitation in 3h	p3	mm
Wind Direction	wd	degrees
Wind Speed	ws	m s <sup>-1</sup>
Maximum Wind Direction	mwd	degrees
Maximum Wind Speed	mws	m s <sup>-1</sup>

## B MERRA-2 資料集

MERRA-2 資料集所包含之變數如下。本研究使用其 Surface Skin Temperature 變數作為主要分析對象。

表 5: MERRA-2 變數表。

變數名稱	縮寫	單位
Longitude	lon	degrees east
Latitude	lat	degrees north
Time	time	minutes since 2024-06-01 00:00:00
2-Meter Air Temperature	t2m	K
Total Precipitable Liquid Water	tql	kg m <sup>-2</sup>
Total Column Odd Oxygen	tox	kg m <sup>-2</sup>
2-Meter Eastward Wind	u2m	m s <sup>-1</sup>
Surface Pressure	ps	Pa
Tropopause Temperature Using Blended TROPP Estimate	tropt	K
Northward Wind at 50 Meters	v50m	m s <sup>-1</sup>
Zero Plane Displacement Height	disph	m
Total Column Ozone	to3	Dobsons
Surface Skin Temperature	ts	K
10-Meter Air Temperature	t10m	K
Tropopause Pressure Based on Thermal Estimate	troppt	Pa
Total Precipitable Ice Water	tqi	kg m <sup>-2</sup>

變數名稱	縮寫	單位
Sea Level Pressure	slp	Pa
Tropopause Pressure Based on Blended Estimate	troppb	Pa
Total Precipitable Water Vapor	tqv	$\text{kg m}^{-2}$
2-Meter Northward Wind	v2m	$\text{m s}^{-1}$
Tropopause Specific Humidity Using Blended TROPP Estimate	tropq	$\text{kg kg}^{-1}$
10-Meter Northward Wind	v10m	$\text{m s}^{-1}$
Eastward Wind at 50 Meters	u50m	$\text{m s}^{-1}$
10-Meter Eastward Wind	u10m	$\text{m s}^{-1}$
2-Meter Specific Humidity	qv2m	$\text{kg kg}^{-1}$
Tropopause Pressure Based on EPV Estimate	troppv	Pa
10-Meter Specific Humidity	qv10m	$\text{kg kg}^{-1}$

## C 2b-8 資料集

2b-8 資料集源自 *KAUST Spatial Statistics Competition (2022)* 所公布之大規模模擬資料，設計目的在於於統一環境下評估各類空間統計方法在預測與參數推估任務中的表現。此系列資料以 ExaGeoStat 高效能統計運算框架生成，透過可重現之高斯過程 (Gaussian process, GP) 模擬，提供標準化且具比較性的實驗條件 (Abdulah, Alamri, Nag, et al., 2022; Abdulah, Alamri, Ltaief, et al., 2022)。

競賽中 Sub-competition 2b 採用非分離且平穩之 GP 模型，其共變異數結構依據 Gneiting (2002) 所提出之非分離型共變異數形式進行設定。對於任兩空間位置

$s \in [0, 1]^2$  與時間  $t \in \mathbb{R}$ ，共變異數函數表示如下：

$$C(\mathbf{h}, u; \boldsymbol{\theta}) = \frac{\sigma^2}{a_t |u|^{2\alpha} + 1} M_\nu \left( \frac{\|\mathbf{h}\|/a_s}{(a_t |u|^{2\alpha} + 1)^{\beta/2}} \right), \quad (37)$$

其中  $\mathbf{h}$  為空間距離、 $u$  為時間差； $\sigma^2 > 0$  為變異數， $\nu > 0$  與  $\alpha \in [0, 1]$  為平滑參數； $a_s, a_t > 0$  為空間與時間尺度； $\beta \in (0, 1]$  描述空間與時間間之交互強度； $M_\nu(\cdot)$  為 Matérn 型相關函數。

依據競賽規格，Sub-competition 2a 與 2b 共產生十八組資料，涵蓋不同尺度配置（弱、中度、強）、不同空間點數（1K 與 10K）、不同時間長度（100 與 1000 時距），並針對預測任務提供三類缺失設定：隨機移除空間位置（RS）、隨機移除空間與時間位置（RST）、以及最後十個時間點系統性缺失（T10）。所有參數組合詳列於 Abdulah, Alamri, Nag, et al. (2022) 所列之 Table 1。而 2b-8 即為其中一項具特定空間點數、時間長度與模型參數組合之資料版本。

