

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：

Thesis/Dissertation Title：

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 Information derived from published and unpublished work of others has been acknowledged in the text, and a list of references is given.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

*** 本研究針對大規模時空觀測資料常見之缺失問題，提出一套整合結構化狀態空間擴散模型（Structured State Space Diffusion, SSSD^{S4}）與自適應固定秩克利金法（Adaptive Fixed Rank Kriging, AFRK）之時空建模框架。該方法旨在於缺漏比例高、資料結構複雜以及長序列預測等情境下，兼具時間依賴性刻畫能力與空間相關結構建模能力。研究首先利用 SSSD^{S4} 擷取長期時間序列的潛在動態，再結合 AFRK 以有效描述不同地點間的空間共變異結構，使模型能同時於時間與空間維度上強化資料重建品質。為量化預測準確度，本研究採用平均平方預測誤差（Mean Squared Prediction Error, MSPE）作為核心評估指標。實驗結果顯示，所提出之時空整合方法相較於僅進行時間建模後再以空間方法補全之傳統流程，於多個資料集皆可達成更低之 MSPE，證明其於長序列重建與時空依存結構提取方面具備顯著優勢。此結果亦反映 SSSD^{S4} 與 AFRK 的結合能有效強化模型的預測穩健性與泛化能力，為未觀測位置之未來時空資料預測與插補研究提供一具可延伸性之框架。

關鍵字：時空模型、資料缺失、結構化狀態空間模型、擴散模型、自適應固定秩克利金法



Abstract

Keywords:



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
1 引言	1
2 相關研究	1
2.1 時間序列模型	2
2.1.1 狀態空間模型	2
2.1.2 結構化狀態空間模型	3
2.1.3 擴散模型	4
2.1.4 基於 S4 層的結構化狀態空間擴散模型	6
2.2 空間統計學	7
2.2.1 克利金法	7
2.2.2 固定秩克利金法	8
2.2.3 自適應固定秩克利金法	8
3 研究方法	9
3.1 問題定義	10
3.2 時空建模方法	10
3.2.1 基於 SSSD ^{S4} 之時間建模	10
3.2.2 基於 AFRK 之空間建模	12
3.3 模型訓練與推論演算法	12
4 實驗	13
4.1 資料集	14
4.1.1 Weather2K	14
4.1.2 MERRA-2	14
4.2 實驗環境與計算資源	15

4.3 實驗設計	15
5 實驗結果	19
6 結論	20
6.1 未來工作	20
6.1.1 跨區域泛化能力	20
6.1.2 空間建模擴展	20
6.1.3 缺失機制與模型穩健性	21
6.1.4 時空協同與生成模型整合	21
6.1.5 模型可解釋性與不確定性量化	21
參考文獻	21
附錄	24
A Weather2K 變數表	24
B MERRA-2 變數表	25



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，傳統方法包括自迴歸整合移動平均模型 (Autoregressive Integrated Moving Average, ARIMA) (Box and Jenkins, 1976)、長短期記憶網路 (Long Short-Term Memory, LSTM) (Hochreiter and Schmidhuber, 1997) 等方法，能有效捕捉資料的時間依賴性，但在長期時間序列或非線性結構下仍有限制。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，可兼具資料填補、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 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

$$\begin{aligned}\mathbf{x}'(t) &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t); \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) + \mathbf{D}\mathbf{u}(t),\end{aligned}\tag{1}$$

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。

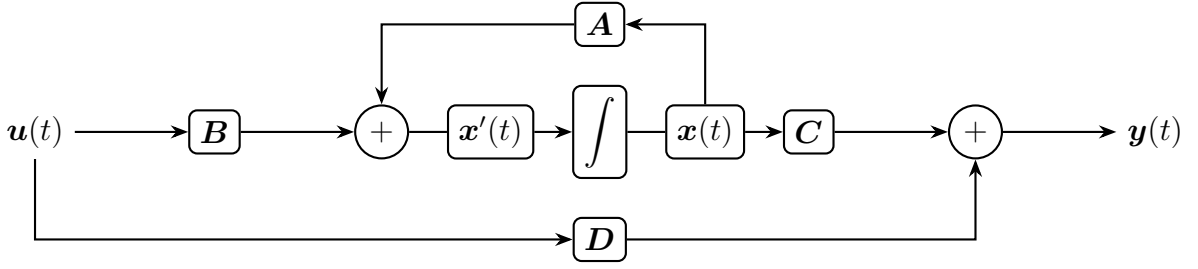


圖 1: 典型的狀態空間模型。

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 (1) 中的隨機矩陣 A 。

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

如 (2) 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)。

2.1.2 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 (1) 離散化，以便將 SSM 應用於離散的輸入序列。定義步長為 Δ ，則離散化的 SSM 如下：

$$\begin{aligned} \mathbf{x}_k &= \bar{A}\mathbf{x}_{k-1} + \bar{B}\mathbf{u}_k; \\ \mathbf{y}_k &= \bar{C}\mathbf{x}_k, \end{aligned} \quad (3)$$

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

$$\begin{aligned}\overline{A} &= (I - \Delta/2 \cdot A)^{-1}(I + \Delta/2 \cdot A); \\ \overline{B} &= (I - \Delta/2 \cdot A)^{-1}\Delta B; \\ \overline{C} &= C.\end{aligned}\tag{4}$$

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。

$$(A, B, C) \sim (V^{-1}AV, V^{-1}B, CV),\tag{5}$$

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：

$$\mathbf{y} = \overline{\mathbf{K}} * \mathbf{u};\tag{6}$$

$$\overline{\mathbf{K}} \in \mathbb{R}^L := (\overline{CB}, \overline{CAB}, \dots, \overline{CA}^{L-1}\overline{B}),\tag{7}$$

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：

$$A = V\Lambda V^* - PQ^\top = V(\Lambda - (V^*P)(V^*Q)^*)V^*,\tag{8}$$

其中 Λ 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)。

2.1.3 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：

$$\begin{cases} q(\mathbf{x}_1|\mathbf{x}_0) = \prod_{t=1}^T q(\mathbf{x}_t|\mathbf{x}_{t-1}); \\ q(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \sqrt{1-\beta_t}\mathbf{x}_{t-1}, \beta_t\mathbf{I}), \end{cases} \quad (9)$$

其中， β_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：

$$\begin{cases} p_\theta(\mathbf{x}_0) = p(\mathbf{x}_T) \prod_{t=1}^T p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t); \\ p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) = \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \Sigma_\theta(\mathbf{x}_t, t)), \end{cases} \quad (10)$$

其中， $p(\mathbf{x}_T) = \mathcal{N}(\mathbf{x}_T; 0, \mathbf{I})$ 為標準常態分佈； μ_θ 與 Σ_θ 為經參數 θ 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 μ_θ 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

$$\mu_\theta(\mathbf{x}_t, t) = \frac{1}{\sqrt{\alpha_t}} \left(\mathbf{x}_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}} \boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t) \right), \quad (11)$$

$$\Sigma_\theta(\mathbf{x}_t, t) = \sigma_t^2 \mathbf{I}, \quad \sigma_t^2 = \beta_t \text{ or } \sigma_t^2 = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t} \beta_t, \quad (12)$$

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

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t} \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I}), \quad (13)$$

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

$$\min_{\theta} = \mathbb{E}_{t, \mathbf{x}_0, \epsilon} \left[\left\| \epsilon - \epsilon_{\theta} \left(\sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, t \right) \right\|^2 \right], \quad (14)$$

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 Strodtthoff (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。

2.1.4 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 Strodtthoff (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 S4 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 Strodtthoff (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^{S4}），作為針對時間序列任務的改進變體。SSSD^{S4} 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 Strodtthoff, 2023)。

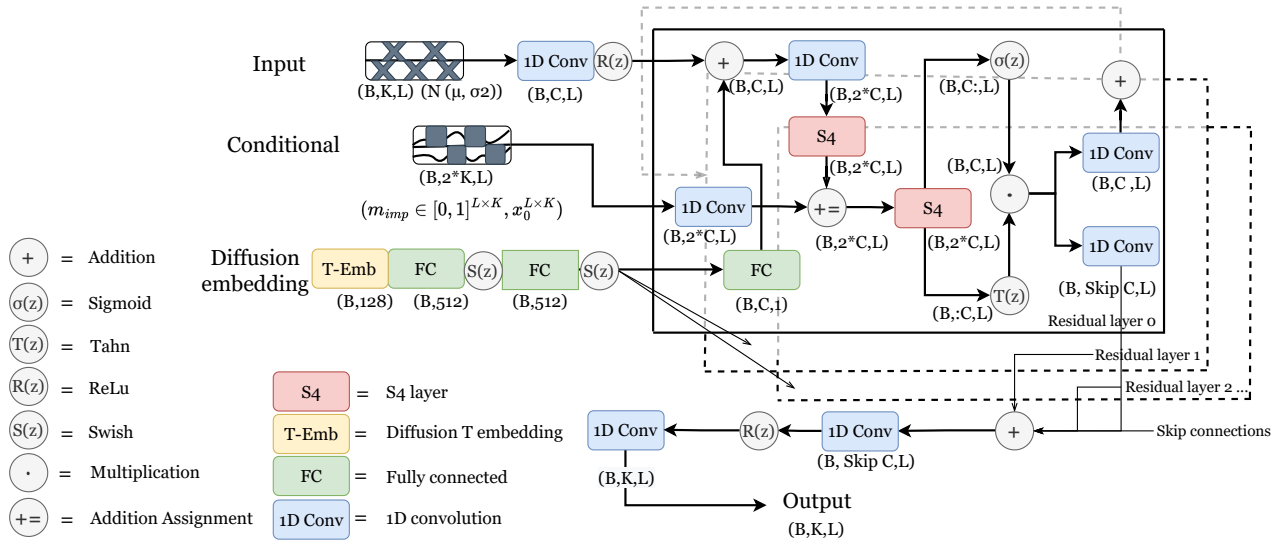


圖 2: SSSD^{S4} 模型架構。

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：

$$Z(\mathbf{s}) = Y(\mathbf{s}) + \varepsilon(\mathbf{s}), \quad \mathbf{s} \in D \subset \mathbb{R}^D, \quad (15)$$

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：

$$Y(\mathbf{s}) = \mu(\mathbf{s}) + \mathbf{f}(\mathbf{s})^\top \mathbf{w} + \xi(\mathbf{s}), \quad (16)$$

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：

$$\text{Cov}(Y(\mathbf{s}), Y(\mathbf{s}')) = \mathbf{f}(\mathbf{s})^\top \mathbf{M} \mathbf{f}(\mathbf{s}') + \sigma_\xi^2 \mathbf{I}(\mathbf{s} = \mathbf{s}'). \quad (17)$$

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：

$$f_k(\mathbf{s}) = \begin{cases} 1, & \text{if } k = 1; \\ x_{k-1}, & \text{if } k = 2, \dots, d+1; \\ \lambda_{k-d-1}^{-1} \times \{\phi(\mathbf{s}) - \Phi \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}\}' \mathbf{v}_{k-d-1}, & \text{if } k = d+2, \dots, n, \end{cases} \quad (18)$$

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 Φ 定義為

$$J(\mathbf{f}) = \boldsymbol{\alpha}' \Phi \boldsymbol{\alpha}, \quad (19)$$

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

$$\phi_i(\mathbf{s}) = \begin{cases} \frac{1}{12} \|\mathbf{s} - \mathbf{s}_i\|^3, & \text{if } d = 1; \\ \frac{1}{8\pi} \|\mathbf{s} - \mathbf{s}_i\|^2 \log(\|\mathbf{s} - \mathbf{s}_i\|), & \text{if } d = 2; \\ -\frac{1}{8} \|\mathbf{s} - \mathbf{s}_i\|, & \text{if } d = 3; \end{cases} \quad (20)$$

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。

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^{S4} 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^{S4} 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 $f(\cdot)$

$$\hat{y}_t(\mathbf{s}^*) = f(\mathbf{Y}_{1:T}(\mathcal{S}_{\text{observed}}), \mathbf{s}^*), \quad \mathbf{s}^* \in \mathcal{S}_{\text{unobserved}}, t > T, \quad (21)$$

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^{S4} 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^{S4} 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^{S4} 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

$$\tilde{y}_t(\mathbf{s}) = \frac{y_t(\mathbf{s}) - \mu(\mathbf{s})}{\sigma(\mathbf{s})}, \quad \mathbf{s} \in \mathcal{S}_{\text{observed}}, \quad (22)$$

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

$$\mu(\mathbf{s}) = \frac{1}{T} \sum_{t=1}^T y_t(\mathbf{s}), \quad (23)$$

$$\sigma(\mathbf{s}) = \sqrt{\frac{1}{T} \sum_{t=1}^T (y_t(\mathbf{s}) - \mu(\mathbf{s}))^2}. \quad (24)$$

隨後，將標準化後的輸入序列 $\tilde{y}_t(\mathbf{s})$ 送入 SSSD^{S4} 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 ϵ_θ 近似正向擴散過程中的真實高斯雜訊 ϵ ，並透過最小化下式更新模型參數

$$\mathcal{L} = \|\epsilon - \epsilon_\theta(\sqrt{\bar{\alpha}_t} \mathbf{x}_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, t)\|^2. \quad (25)$$

本研究於反向擴散過程中，進一步將 ϵ_θ 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 ϵ_θ 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。

$$\hat{\epsilon}_\theta = \text{AFRK}(\epsilon_\theta) = \mu_{\epsilon_\theta} + \mathbf{f}_{\epsilon_\theta}^\top \hat{\mathbf{w}}_{\epsilon_\theta} + \hat{\xi}_{\epsilon_\theta}. \quad (26)$$

經 AFRK 調整之 SSSD^{S4} 模型預測輸出可表示為

$$\hat{\tilde{y}}_t(\mathbf{s}) = g_{\theta_{\text{S4+AFRK}}}(\tilde{y}_{1:t}(\mathbf{s})), \quad \mathbf{s} \in \mathcal{S}_{\text{observed}}, \quad (27)$$

其中， $g_{\theta_{\text{S4+AFRK}}}(\cdot)$ 為採用 S4 結構並經 AFRK 調整之 SSSD 模型，而 $\hat{\tilde{y}}_t(\mathbf{s})$ 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

$$\hat{y}_t(\mathbf{s}) = \sigma(\mathbf{s}) \hat{\tilde{y}}_t(\mathbf{s}) + \mu(\mathbf{s}), \quad (28)$$

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^{S4} 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^{S4} 所產生之已觀測位置的時間預測值 $\hat{y}_t(\mathbf{s})$, $\mathbf{s} \in \mathcal{S}_{\text{observed}}$ 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

$$\tilde{y}_t(\mathbf{s}) = \frac{\hat{y}_t(\mathbf{s}) - \mu_t}{\sigma_t}, \quad \mathbf{s} \in \mathcal{S}_{\text{observed}}, \quad (29)$$

其中， μ_t 與 σ_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

$$\hat{y}_t(\mathbf{s}^*) = \mu(\mathbf{s}^*) + \mathbf{f}(\mathbf{s}^*)^\top \hat{\mathbf{w}}_t + \hat{\xi}_t(\mathbf{s}^*), \quad (30)$$

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

$$\hat{y}_t(\mathbf{s}^*) = \sigma_t \hat{y}_t(\mathbf{s}^*) + \mu_t. \quad (31)$$

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^{S4} 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 η

- 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 θ

- 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: $\hat{\tilde{y}}_t(\mathbf{s}^*) = \text{AFRK}(\tilde{y}_t(\mathbf{s}) | \mathbf{s}^*), \quad \forall \mathbf{s}^* \in \mathcal{S}_{\text{unobserved}}$
 - 13: $\hat{y}_t(\mathbf{s}^*) = \sigma_t \hat{\tilde{y}}_t(\mathbf{s}^*) + \mu_t$
 - 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 03 月 01 日 00 時至 2021 年 08 月 31 日 21 時之觀測序列。實驗選取氣溫 (Air Temperature)、最高溫 (Maximum Temperature)、最低溫 (Minimum Temperature) 與相對濕度 (Relative Humidity) 4 個變數作為資料來源。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 07 月 01 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 43.0° 至 55.0° 與東經 0.0° 至 30.0° 之矩形空間範圍內，並選取該區域內 1,000 個觀測點作為資料來源。

4.2 實驗環境與計算資源

為確保所提出之時空整合模型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^{S4} 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
Training Configuration		
	Batch size	100
	Learning rate	0.0005
	Only generate missing	true
	Masking	forecast
	Missing k	—
SSSD^{S4}		
WaveNet	Input channels	—
	Output channels	—
	Residual layers	20
	Residual channels	20
	Skip channels	20
Diffusion step embedding	Input dimension	64
	Hidden dimension	128
	Output dimension	128
S4	Max sequence length	—
	State dimension	128
	Dropout	0.2
	Bidirectional	true
	Layer normalization	true
Diffusion	Diffusion steps (T)	100
	β_0	0.0001
	β_T	0.01
autoFRK		
	Method	fast
	Thin-plate spline method	rectangular

表 2 詳列了各實驗之參數設定，旨在比較模型於不同訓練條件下之表現。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，以及針對輸入變數之時間重塑 (Temporal Reshaping) 機制。透過對比基準模型 SSSD^{S4} 與訓練階段整合 AFRK 之 SSSD^{S4+AFRK} 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。而時間重塑機制將原始形狀為 (N, T, C) 的序列資料轉換為 $(N, T/p, C \times p)$ ，亦即將連續 p 個時間步長之觀測值併入特徵通道中，用於驗證將時序資訊轉化為通道維度，是否能增強模型對長短期時空依賴關係 (spatio-temporal dependency modeling) 之建模能力。

表 2: 實驗參數組合設定 (Weather2K / MERRA-2)。

控制變因	實驗 1	實驗 2	實驗 3	實驗 4
迭代次數	4,000	4,000	4,000	4,000
訓練策略	SSSD ^{S4+AFRK}	SSSD ^{S4}	SSSD ^{S4+AFRK}	SSSD ^{S4}
已觀測位置	1,492 / 960	1,492 / 960	1,492 / 960	1,492 / 960
未觀測位置	374 / 240	374 / 240	374 / 240	374 / 240
時間重塑	false	false	true	true
併入時間步 p	—	—	8 / 24	8 / 24
Input Channels	4 / 1	4 / 1	32 / 24	32 / 24
S4 Max Sequence Length	1,328 / 3,984	1,328 / 3,984	166 / 166	166 / 166
Missing k	144 / 432	144 / 432	18 / 18	18 / 18

資料集的已觀測與未觀測測站站點分布如圖 3 與圖 4 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。

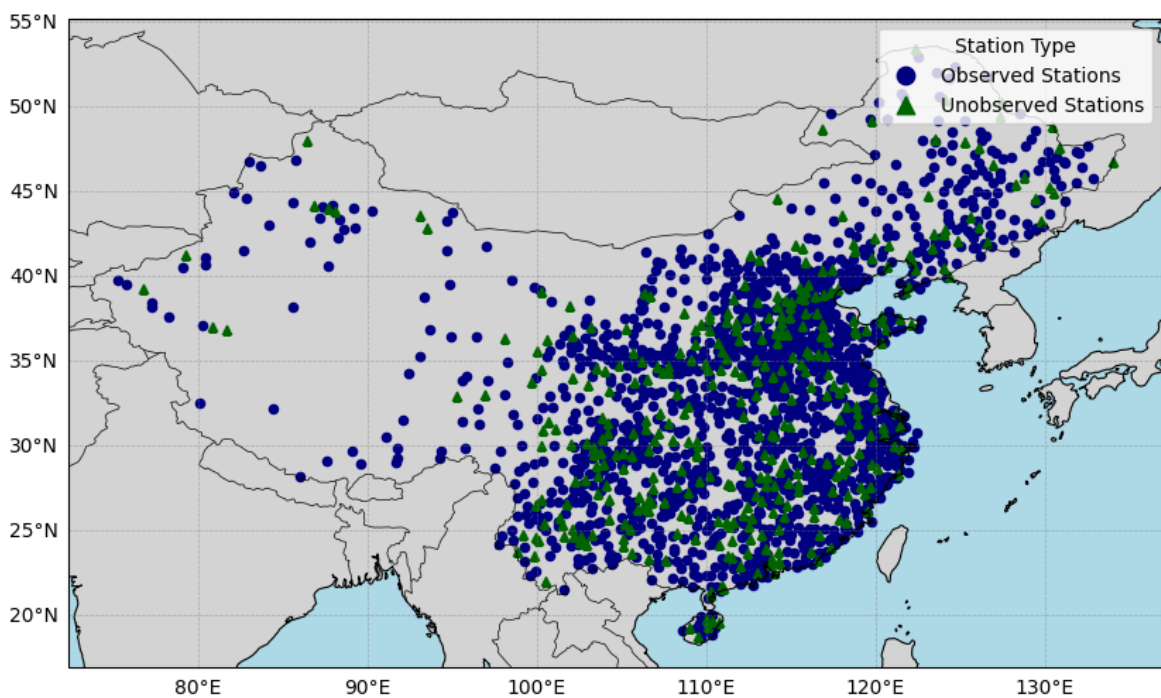


圖 3: Weather2K 3-Hourly 資料站點分布圖。

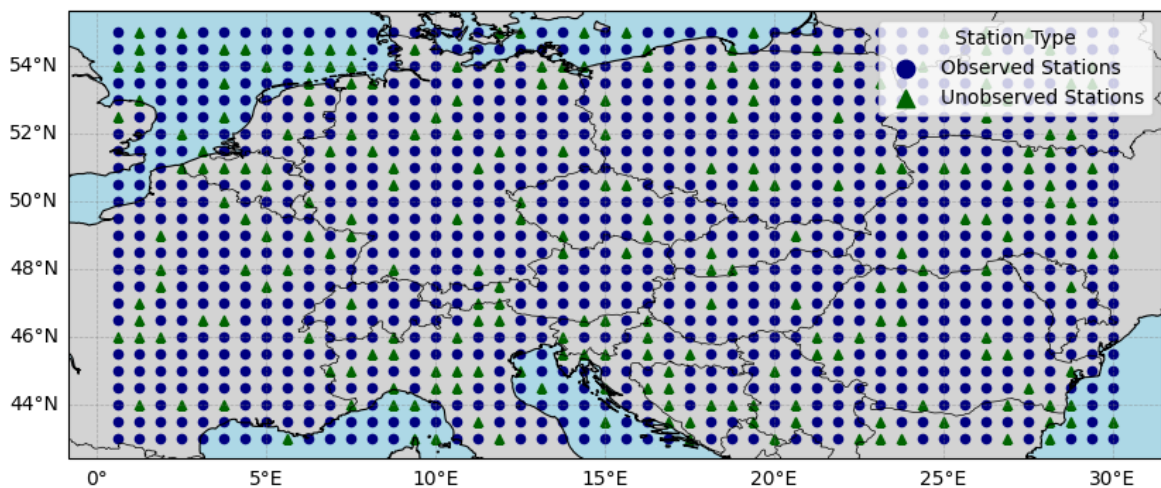


圖 4: MERRA-2 Hourly 資料站點分布圖。

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，用以衡量模型在未觀測位置未來時間點的預測精準度。表 3 呈現 SSSD^{S4}+AFRK 的整合時空框架，與僅使用 SSSD^{S4} 再以 AFRK 填補空間欄位的策略，在兩個資料集上的 MSPE 表現，以及 AFRK 在推論階段選取的 MRTS 基底數量。

表 3: 各資料集於不同時序預測之 MSPE 與推論階段 AFRK 選取之 MRTS 基底數量。

評估項目	實驗 1	實驗 2	實驗 3	實驗 4
Weather2K				
未觀測地點未來	73.0718	75.0860		100.2486
未觀測地點過去	23.4334	19.1816		21.2375
已觀測地點未來	78.6897	85.0839		274.7339
# of MRTS Basis	382	382		178
MERRA-2				
未觀測地點未來	1616.0968	1555.5420		1373.6186
未觀測地點過去	1.5091	1.1911		3.4276
已觀測地點未來	1616.0968	1566.5607		1987.3689
# of MRTS Basis	301	301		111

***表 3 所示，整合的 SSSD^{S4}+AFRK 時空框架，相較於單獨使用 SSSD^{S4} 再以 AFRK 填補空間欄位的策略，於兩資料集上皆獲得更低的 MSPE。此結果顯示，本研究提出的整合方法能有效捕捉複雜的時空依賴關係，展現高度的可靠性與穩健性。

***表 3 的結果表明，單獨使用 SSSD^{S4} 或 AFRK 在特定任務上表現尚可，但在未觀測位置的未來預測上，結合兩者的時空整合框架普遍獲得最低的 MSPE。這反映出整合方法能同時捕捉時間序列中的長程依賴性與空間結構的相關性，顯著提升未觀測位置的預測精準度。不同實驗間的比較亦顯示，整合框架在兩個資料集上皆保持穩定優勢，證明其在大規模時空氣象資料建模中的實務可行性與泛化能力。

6 結論

*** 本研究提出一套結合時間序列生成模型與空間統計插值方法的整合式時空預測框架，由 SSSD^{S4} 負責時間結構建模，並以 AFRK 建立低秩空間基底與跨區域插值機制，XXXX。透過在 Weather2K 與 MERRA-2 兩個異質性高、空間分佈差異顯著的資料集上驗證，本研究顯示該整合架構能在未觀測地點的未來時序預測任務中取得穩定且優於基線模型的表現，特別是在稀疏觀測與跨區域泛化的情境下展現良好適應性。此外，本研究亦指出 SSSD^{S4} 能有效捕捉長期時間依賴，而 AFRK 所引入的空間低秩結構可提供額外的空間一致性補償，使整合模型在高不確定性或觀測缺失的狀況下仍維持合理預測。本研究驗證了結構化時間模型與低秩空間模型結合的可行性，並為未來建構可泛化、可解釋且跨場景適用的時空預測系統奠定基礎。

6.1 未來工作

儘管本研究提出的時空整合框架已在未觀測地點的未來時序預測任務中展現優異性能，仍存在多個潛在改進方向，以進一步提升模型的廣泛適用性與理論價值。

6.1.1 跨區域泛化能力

本研究主要驗證了模型於中國地區地面觀測資料及全球再分析資料的表現。未來可擴展至更多區域性資料集，包括歐洲、北美與東南亞等，藉以檢驗模型在不同氣候型態、地形環境以及資料品質條件下的穩定性與泛化能力，並分析不同地理與氣候條件對 SSSD^{S4} 時間建模與 AFRK 空間插值效果的影響。

6.1.2 空間建模擴展

AFRK 目前所構建的空間結構具有一定彈性，但仍為固定形式。未來研究可考慮引入動態核函數或可調節的空間權重機制，使其能隨時間變化捕捉局部或區域性氣象結構的轉變。此外，非平穩空間模型或區域化建模方法亦可進一步改善 AFRK 對高度異質化或稀疏觀測網格的適應性，從而增強整合框架在極端氣候事件或非典型時空模式下的表現。

6.1.3 缺失機制與模型穩健性

除了既有的未來時序預測 (Time Forecasting, TF) 外，後續可延伸探討 SSSD^{S4} 與 AFRK 在多種缺失機制下的表現差異，包括隨機缺失 (Random Missing, RM)、隨機區塊缺失 (Random Block Missing, RBM) 以及大範圍遮蔽型缺失 (Blackout Missing, BM)。在 RM 與 RBM 中，缺失具有高隨機性與局部破碎性，使時空相關結構受到更強破壞；AFRK 的低秩空間結構可望提供額外的空間一致性約束，以補償 SSSD^{S4} 在時序生成中所面臨的高不確定性。相較之下，BM 屬於大範圍且多時間步的連續遮蔽，對純時間模型與純空間模型皆構成挑戰，因此未來可進一步設計結合 AFRK 空間基底與 SSSD^{S4} 擴散機制的協同重建策略，以提升模型於不同缺失模式下的穩健性。

6.1.4 時空協同與生成模型整合

針對 SSSD^{S4} 與 AFRK 的結合，未來可探索更緊密的時空協同策略。例如，可在 SSSD^{S4} 的時間序列輸出中引入基於 AFRK 空間結構的正則化項，以強化時間預測與空間插值的相互約束，或透過可微分空間插值模塊實現端到端訓練，使模型在未觀測位置的預測精度進一步提升。此外，隨著大規模生成式氣象模型的發展，整合擴散模型或其他大型 AI 氣象模型的生成能力，亦可在高缺失率資料補全、物理一致性預測及極端事件建模上提供額外增益。

6.1.5 模型可解釋性與不確定性量化

模型的可解釋性與不確定性量化仍為重要研究方向。未來可進一步開發時空特徵重要度分析方法，以評估 SSSD^{S4} 的時間依賴表達與 AFRK 的空間插值貢獻，同時估計未觀測位置預測的不確定性與罕見事件的敏感度，從而提升模型在科學研究及實務氣象預報系統中的可靠性與可操作性。

參考文獻

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>.
- 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>.
- 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.
- 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>.
- 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. 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>.
- 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>.
- 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.
- 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.
- 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.

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 變數表

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	%

變數名稱	縮寫	單位
Precipitation in 3h	p3	mm
Wind Direction	wd	degrees
Wind Speed	ws	m s^{-1}
Maximum Wind Direction	mwd	degrees
Maximum Wind Speed	mws	m s^{-1}

B MERRA-2 變數表

MERRA-2 資料集所包含之變數如下。本研究使用其大氣柱總奇氧量（total column odd oxygen）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^{-2}
Total Column Odd Oxygen	tox	kg m^{-2}
2-Meter Eastward Wind	u2m	m s^{-1}
Surface Pressure	ps	Pa

變數名稱	縮寫	單位
Tropopause Temperature Using Blended TROPP Estimate	tropt	K
Northward Wind at 50 Meters	v50m	m s^{-1}
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^{-2}
Sea Level Pressure	slp	Pa
Tropopause Pressure Based on Blended Estimate	troppb	Pa
Total Precipitable Water Vapor	tqv	kg m^{-2}
2-Meter Northward Wind	v2m	m s^{-1}
Tropopause Specific Humidity Using Blended TROPP Estimate	tropq	kg kg^{-1}
10-Meter Northward Wind	v10m	m s^{-1}
Eastward Wind at 50 Meters	u50m	m s^{-1}
10-Meter Eastward Wind	u10m	m s^{-1}
2-Meter Specific Humidity	qv2m	kg kg^{-1}
Tropopause Pressure Based on EPV Estimate	troppv	Pa
10-Meter Specific Humidity	qv10m	kg kg^{-1}