

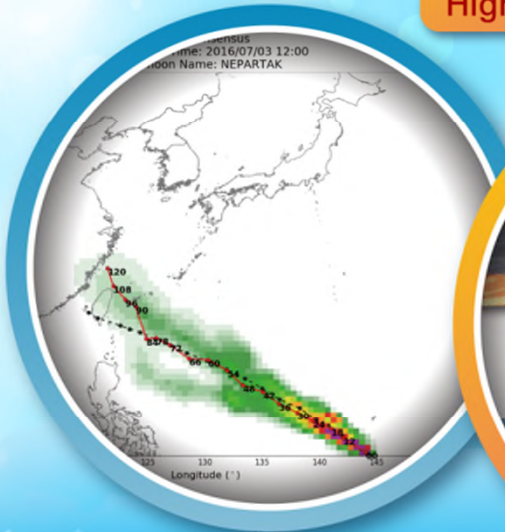
以序率機器學習產生系集成員之技術評估

劉冠倫¹ 陳昱璵¹ 馮智勇¹ 陳佳莉¹ 林涵芳¹ 張惠玲²

多采科技有限公司¹、中央氣象署²

ManySplendid

高效能科學計算模式研發
High-efficiency scientific modeling



都市積淹水與洪泛測預報
Urban water inundation and flood forecasting



大數據人工智慧技術開發
Big data and artificial intelligence

水利水資源工程規劃設計
Hydraulic and water resources engineering planning

氣候服務跨域諮詢與整合
Climate services for interdisciplinary integration

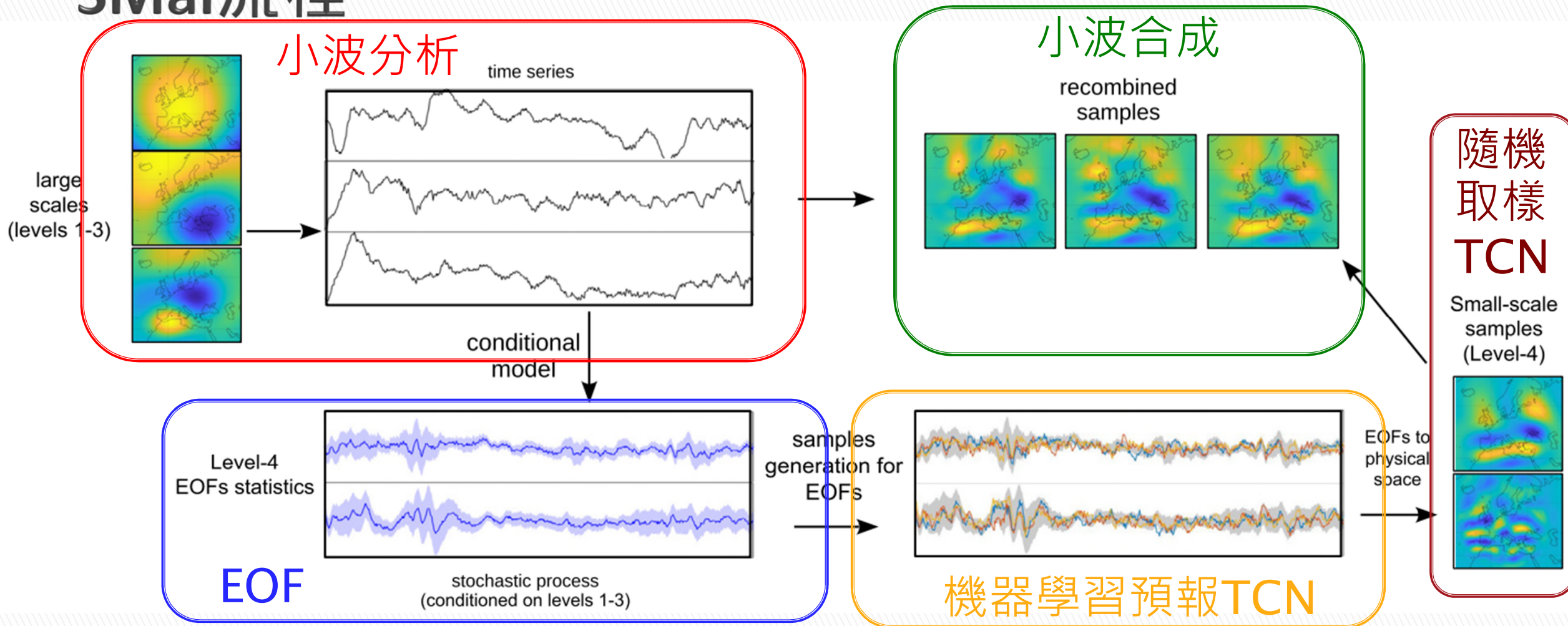
大綱

- ▶ 緣起
- ▶ SMaL流程簡介
- ▶ 模型訓練說明及產出分析
- ▶ 產製系集成果
- ▶ 將SMaL流程套用至系集預報產製

Stochastic Machine-Learning (SMaL) 發展緣起

- ▶ 系集成員的實作上有部分是採用對分析場產生N組擾動後，再以NWP進行模擬得N個動力模式成員的系集預報
 - 產製分析場擾動往往需使用較多計算資源
 - 系集成員數值模擬亦需大量計算資源
- ▶ 連續之分析場時序資料應有助於從氣象/氣候角度掌握當下時間可能產生擾動訊號的空間位置與幅度
 - 可發展統計方式產製N組分析場擾動而大幅減少計算資源使用量
 - 具結合AI氣象模型快速推論系集成員的潛力
- ▶ Wan et. AI (2021) 提出 SMaL 架構
 - 根據大空間尺度特徵對小空間尺度特徵進行隨機實現的機器學習方法
 - 特徵為產製的分析場可以維持大尺度的特徵，只針對小尺度的特徵進行擾動

SMal流程



為什麼會先做球面小波分析

- ▶ 大尺度特徵會與小尺度特徵有關
 - 且小尺度的特徵並非獨立於大尺度訊號之外

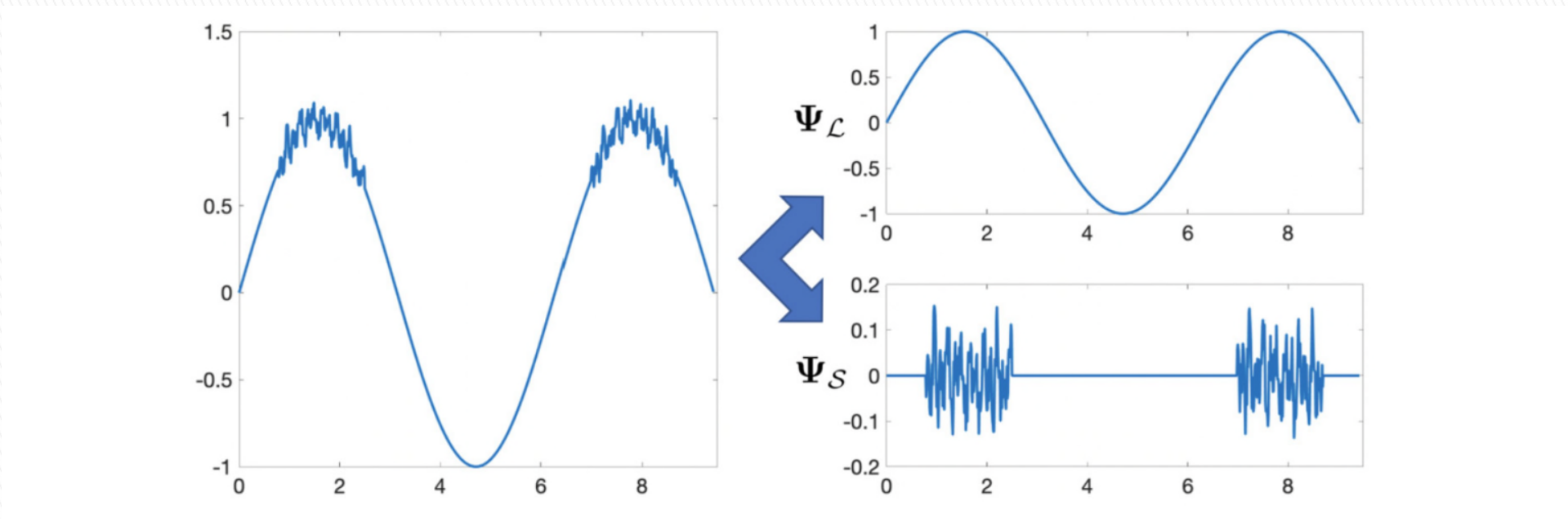
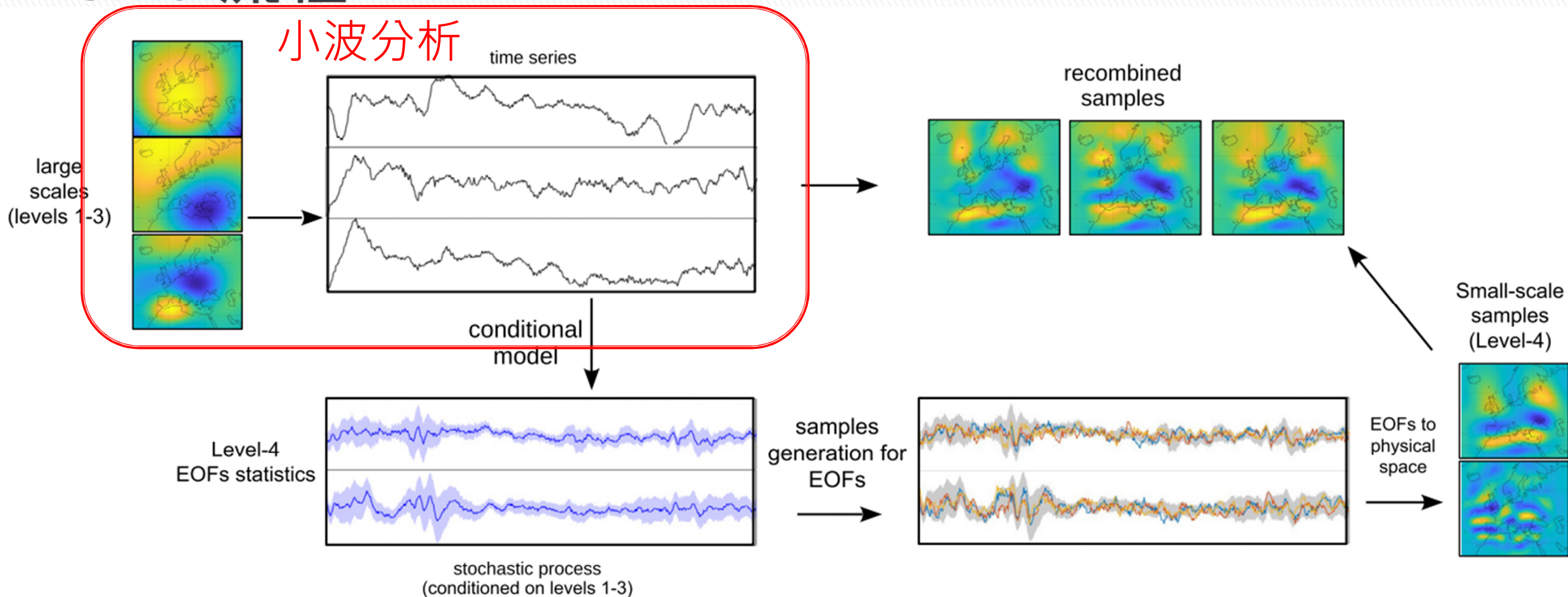


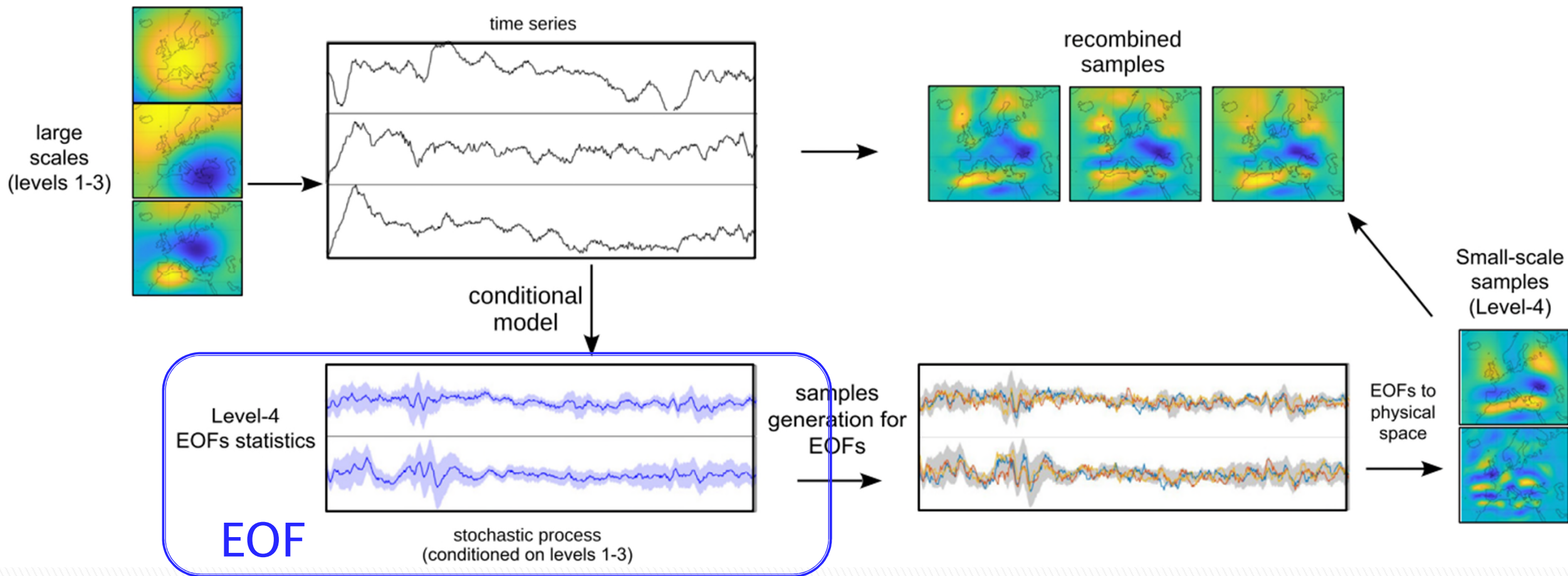
Fig. 2. Decomposition of a spatial field into large and small scales. Note that the small scales may still be characterized by large-scale dependencies.

SMal流程



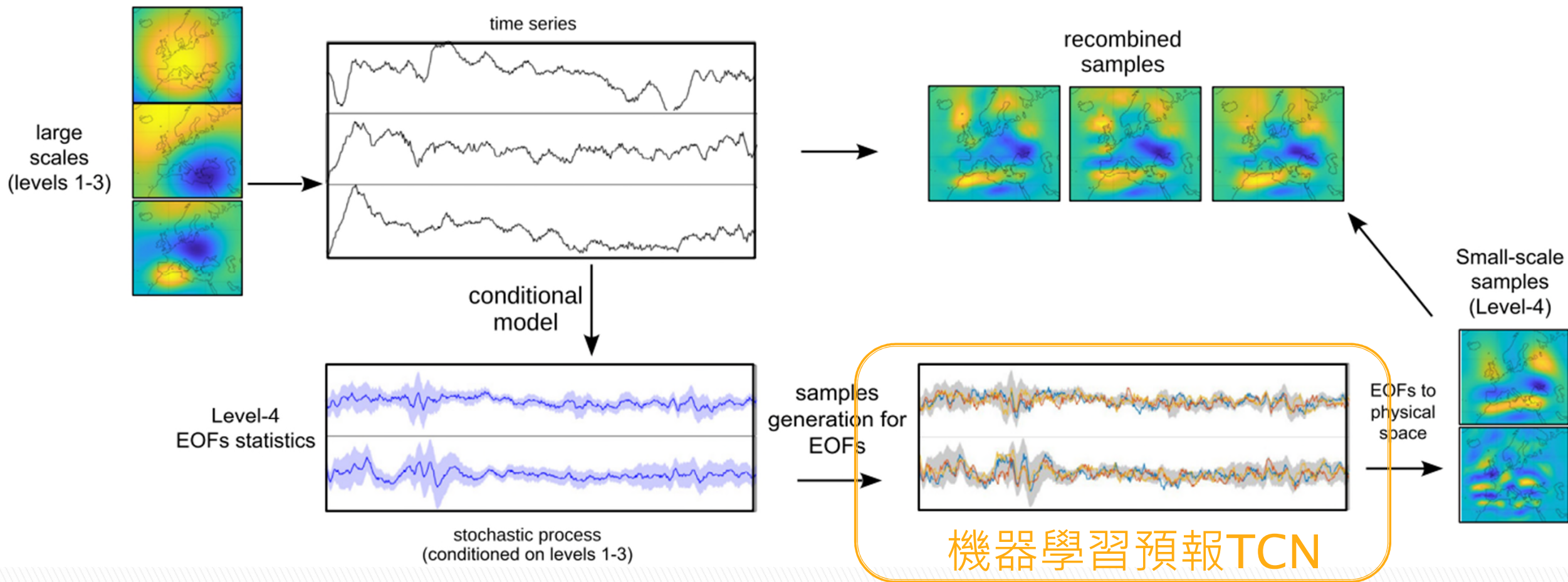
- ▶ 使用小波分析區分出分析場不同空間尺度的特徵

SMal流程



- ▶ 通過經驗正交函數(EOF)分解出小尺度特徵的時間序列

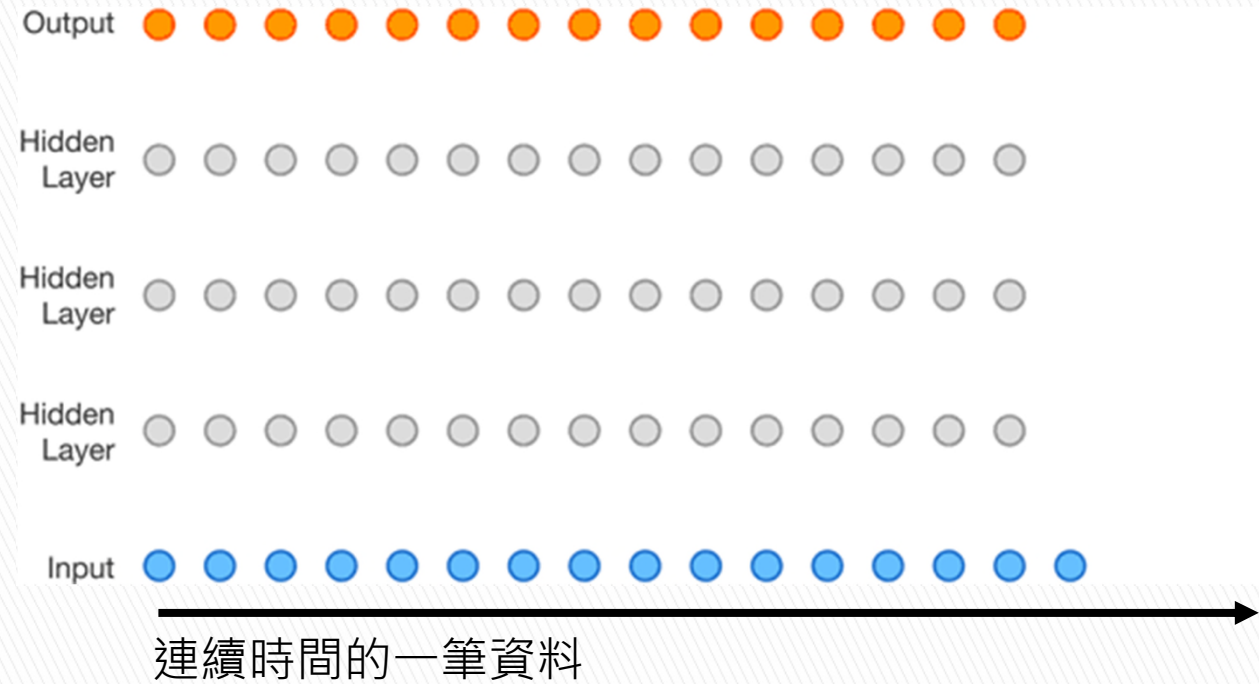
SMal流程



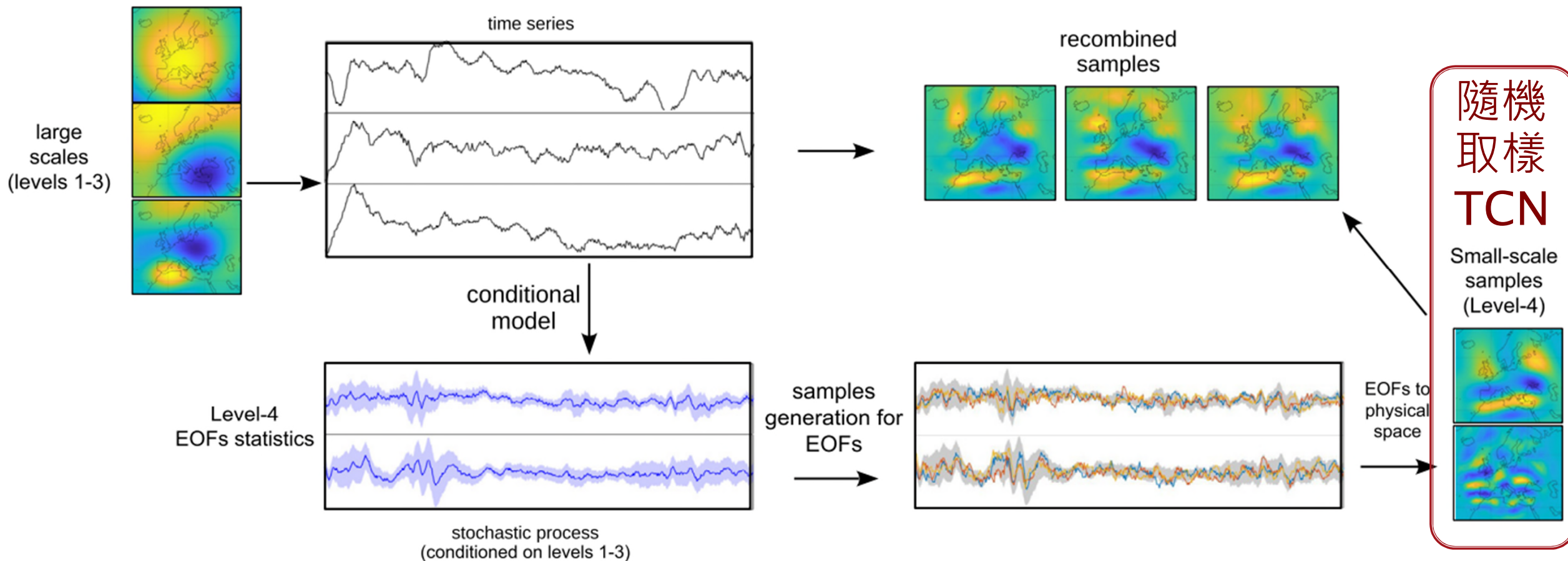
- ▶ 使用時間卷積網絡(TCN)來進行機器學習，利用大尺度特徵來預報EOF出來之小尺度特徵的時間序列

TCN(temporal convolution network)

- ▶ 特色是使用因果卷積和空洞卷積
- ▶ 因果卷積(causal convolution)
 - 特徵是當下的輸出會使用到當下以及之前的資料
 - 不會用到未來的資料
 - 增加層數可以看到越久之前的資料
- ▶ 空洞卷積(dilated convolution)
 - 特徵是會跳過幾筆資料，讓觸及的資料變廣
 - 可以用較少層數但能觸及的資料到更久之前

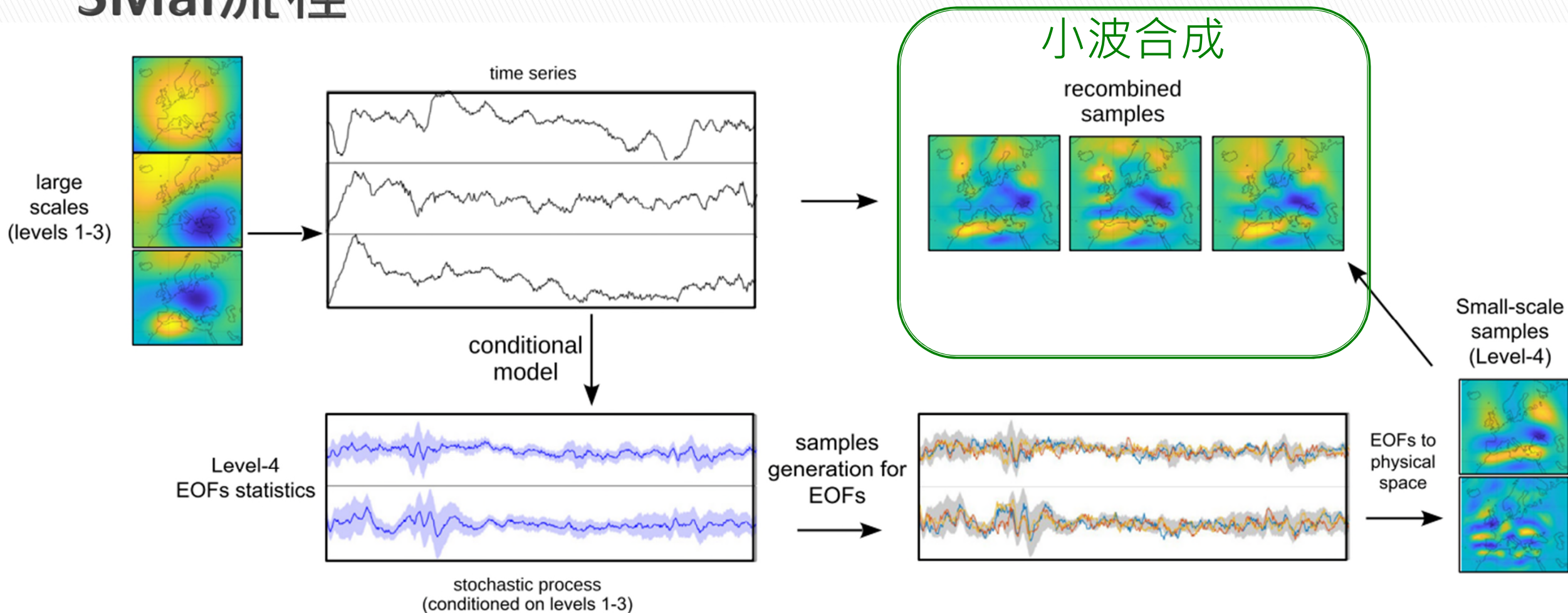


SMal流程



- ▶ 再使用TCN來學習隨機取樣需要的參數，得到取樣後的EOF時間序列

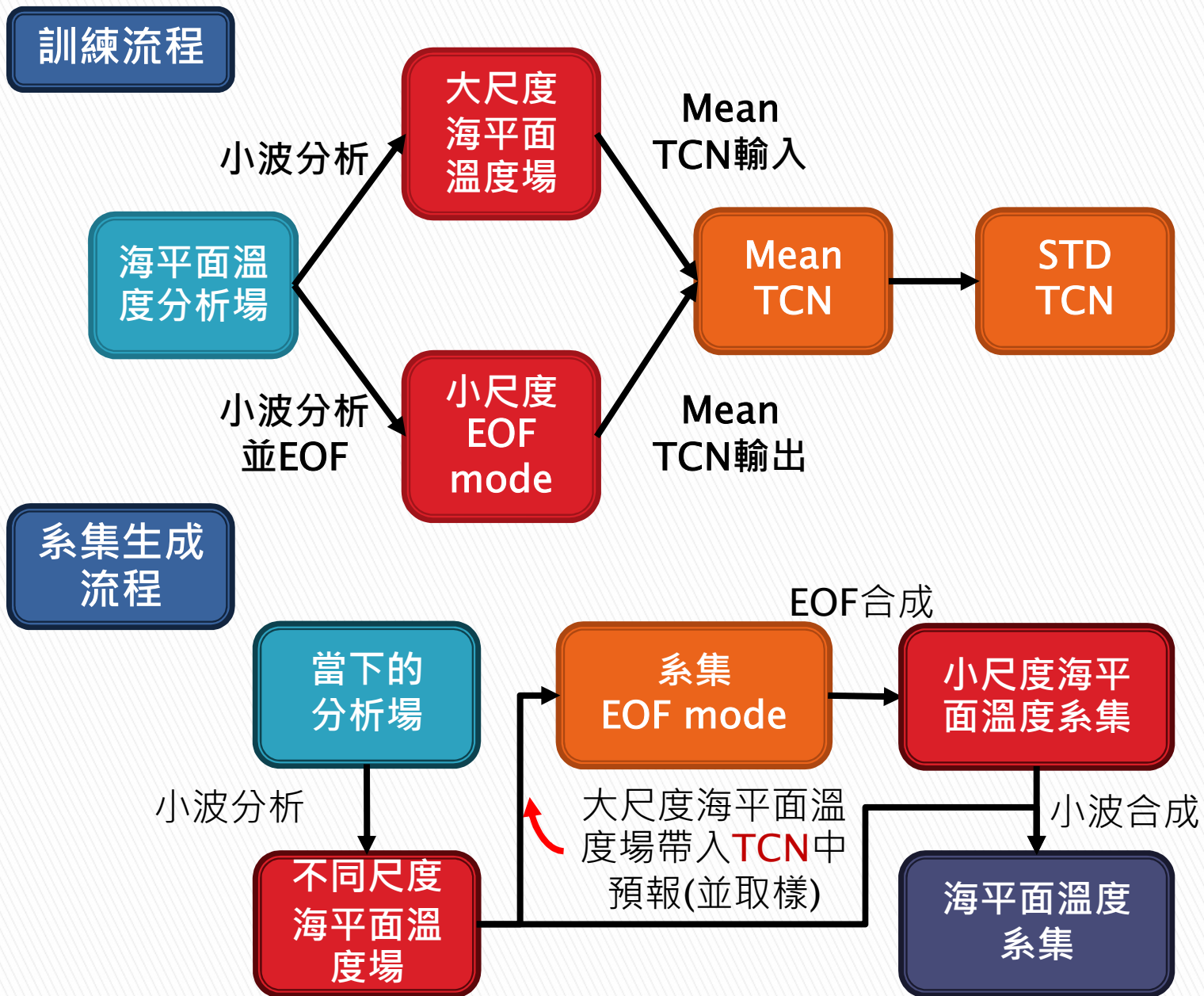
SMal流程



- ▶ 取樣得到的時間序列經過EOF的合成轉換到小波的小尺度特徵，再使用小波合成到原始資料就可以得到多種採樣的原始資料

SMaL流程整理

- ▶ 訓練時使用歷史分析場資料代入TCN
 - 本次測試使用海平面溫度場
- ▶ 實作時則當下或所需時間的分析場資料代入TCN
 - TCN需要連續時間資料



HYCOM海平面溫度資料

- ▶ 空間解析度

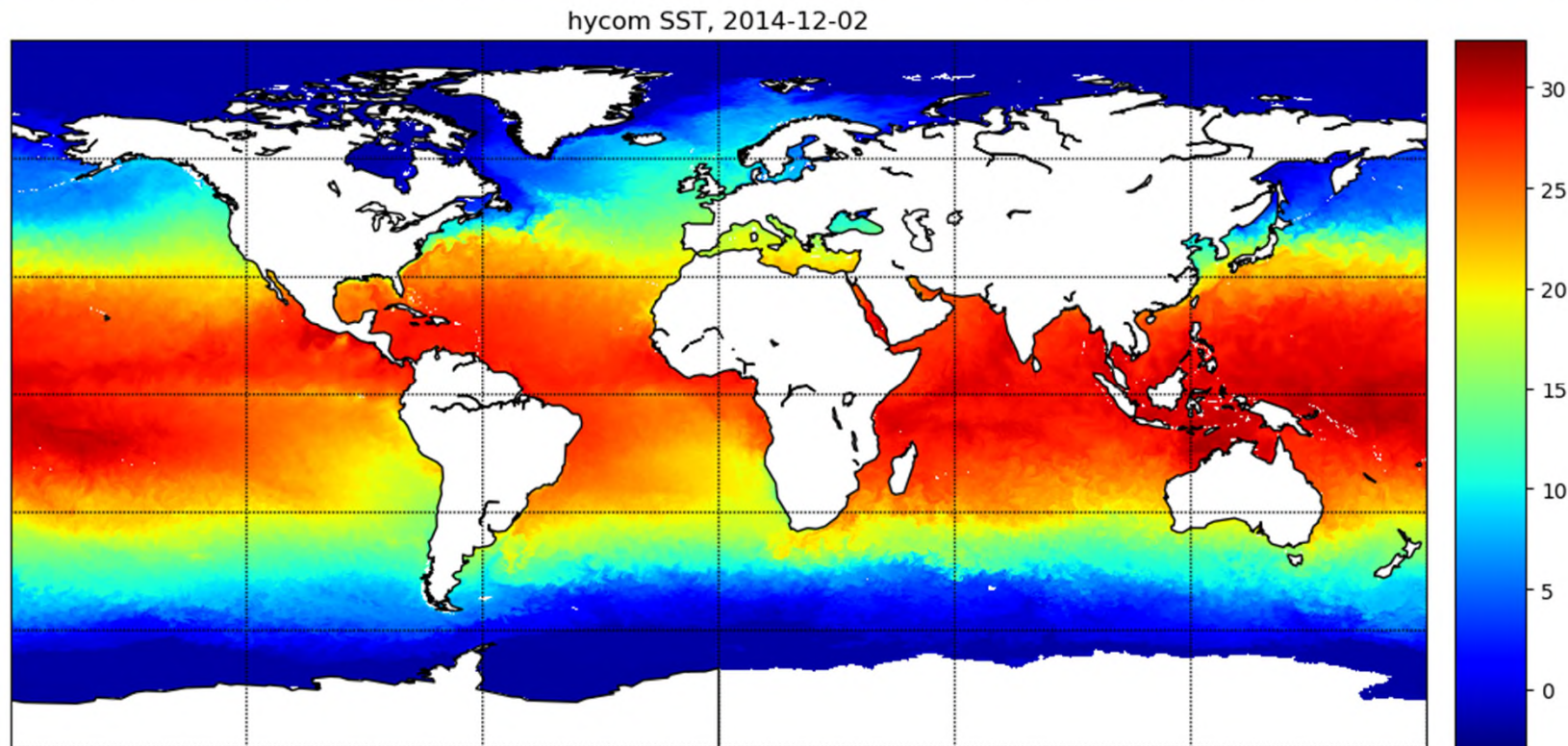
- 約為8公里

- ▶ 經緯度：

- lat(3251,4251)
- lon(4500)
- 使用時會內插至相同網格點

- ▶ 時間：

- 2014年至2024年
- 本次測試使用00Z



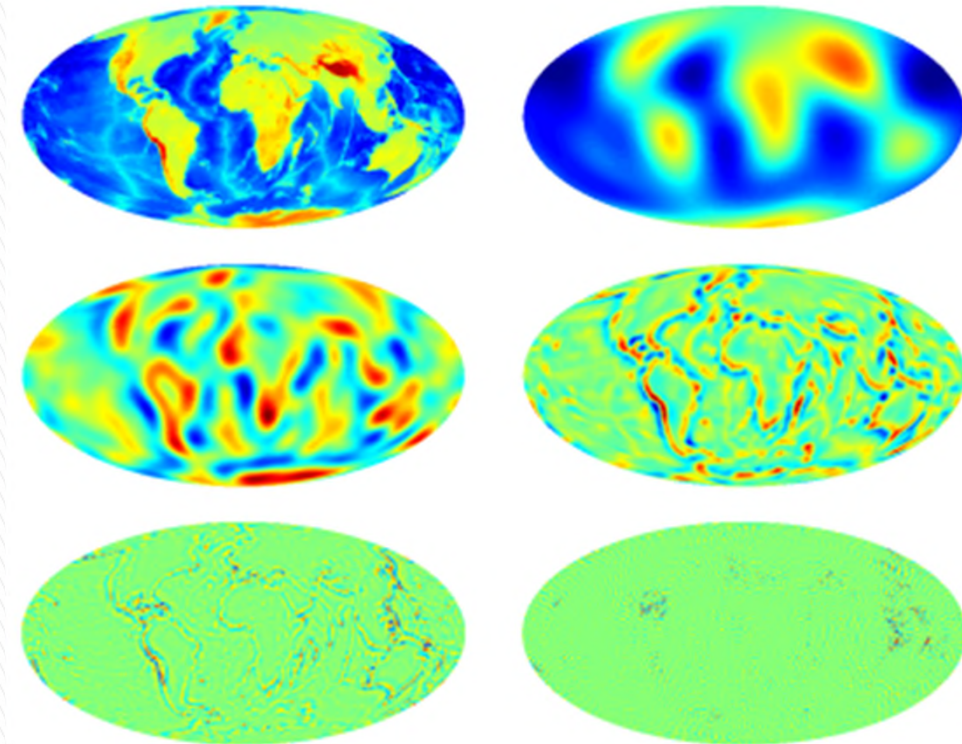
小波分析實作

- ▶ 實作時使用離散球面小波分析
 - 使用pys2let進行實作

pys2let 2.2.4 <https://pypi.org/project/pys2let/>

- ▶ 能將二維空間的資料拆解成多種尺度資料
- ▶ 使用時會是一張二維空間資料產出多張二維空間資料

球面小波分析示意圖

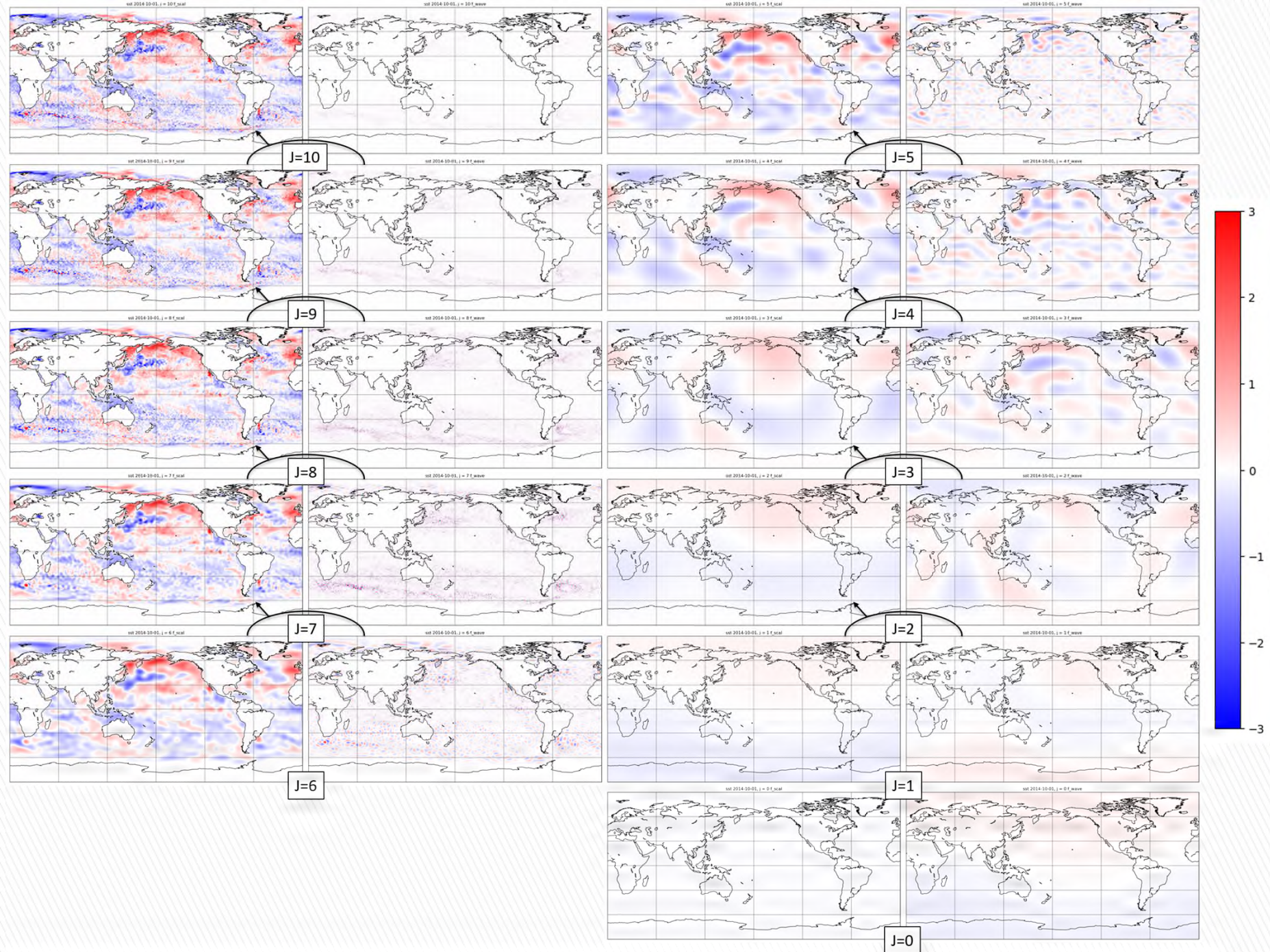


(a) Full-resolution scale-discretised wavelet transform

Y. Wiaux et
al.(2008)

海平面溫度

- ▶ 小波分析
 - 可以分成多種尺度資料



TCN資料排序

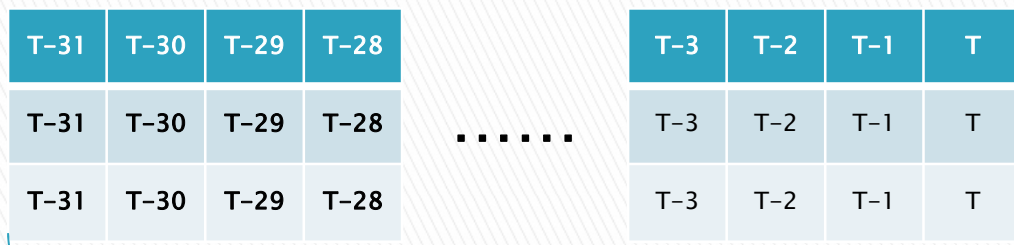
▶ TCN架構

◦ 輸入：

- 32筆(天)，多尺度資料，00z海平面溫度

T：時序中最新的分析場時間點

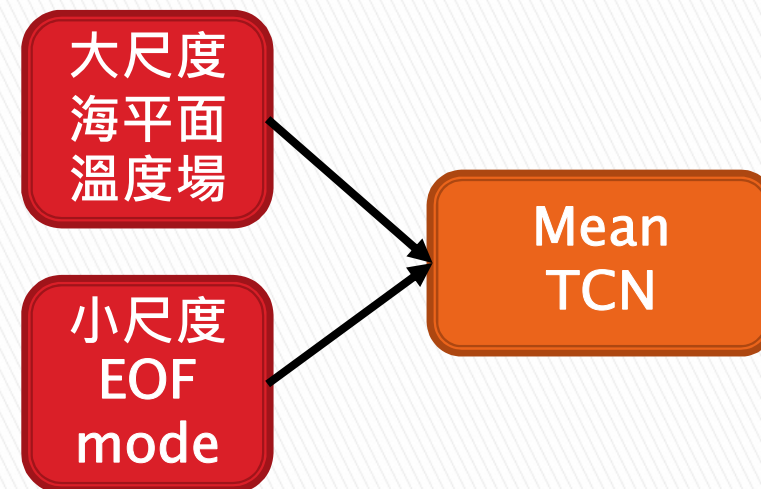
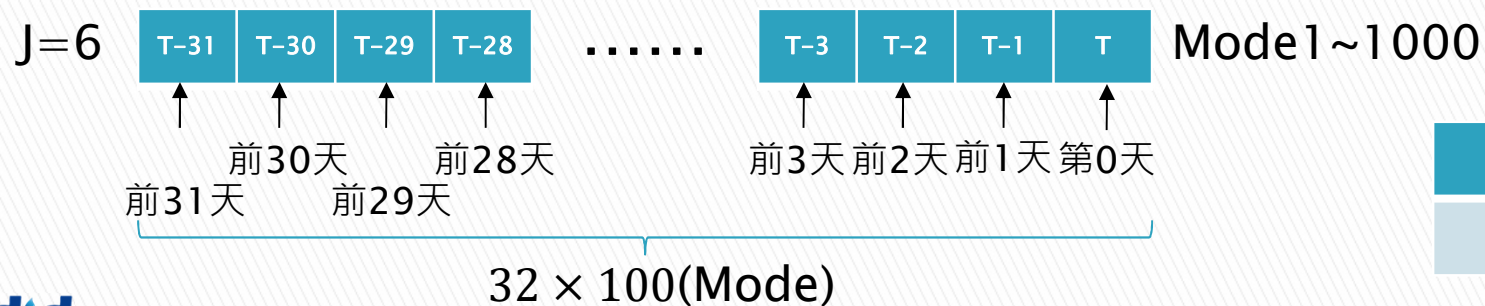
多尺度的資料



◦ 輸出：

32 × 85450(格點)

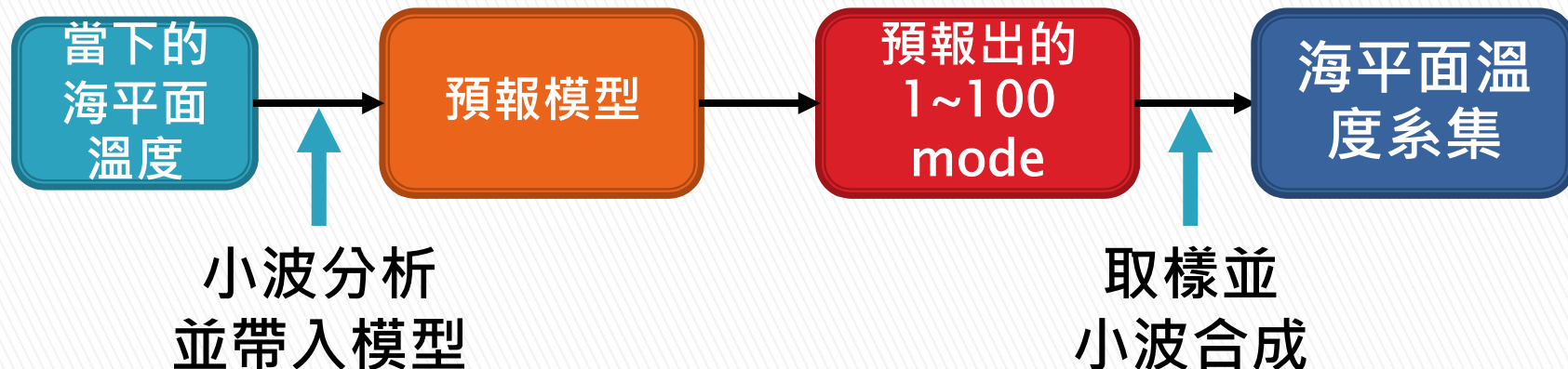
- 32筆(天)，小尺度資料，00z海平面溫度的EOF時間序列



| train | valid | test |
|-------|-------|------|
| 2029 | 677 | 677 |

取樣得到系集的方法

1. 當下的海平面溫度帶入Mean TCN預報模型得到100個mode的預報值(Mean)
 2. 預報出來的Mean帶入STD TCN預報模型得到100個mode的標準差
 3. 假設誤差是常態分布，常態分布的Mean及STD分別用上述2項帶入，再使用常態分布的隨機取樣得到一組100個mode的取樣
 4. 將100個mode的取樣後EOF合成回f_sca1，再小波合成回溫度
- 重複隨機取樣可以得到多組系集

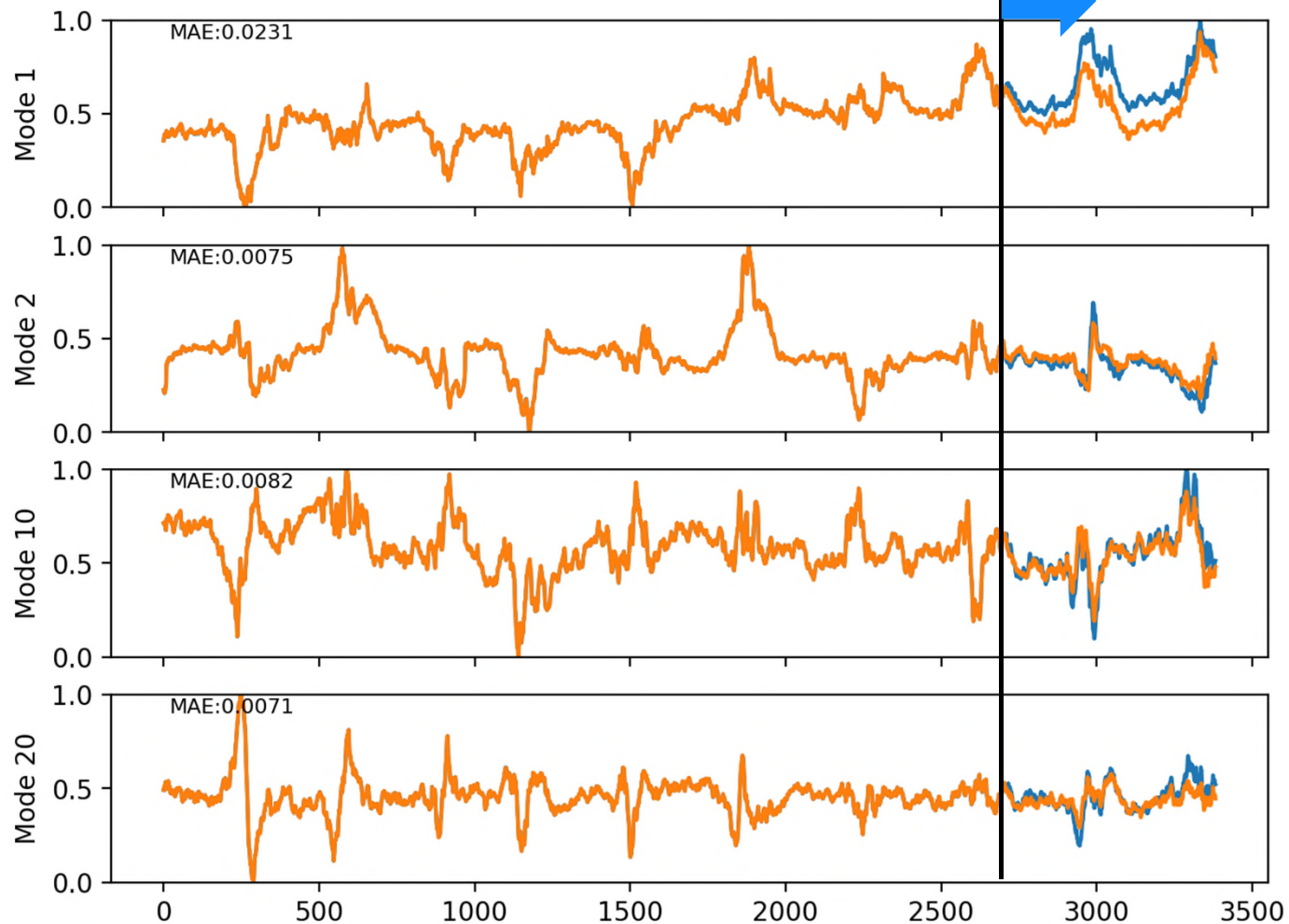


Mean TCN 訓練及測試資料比較

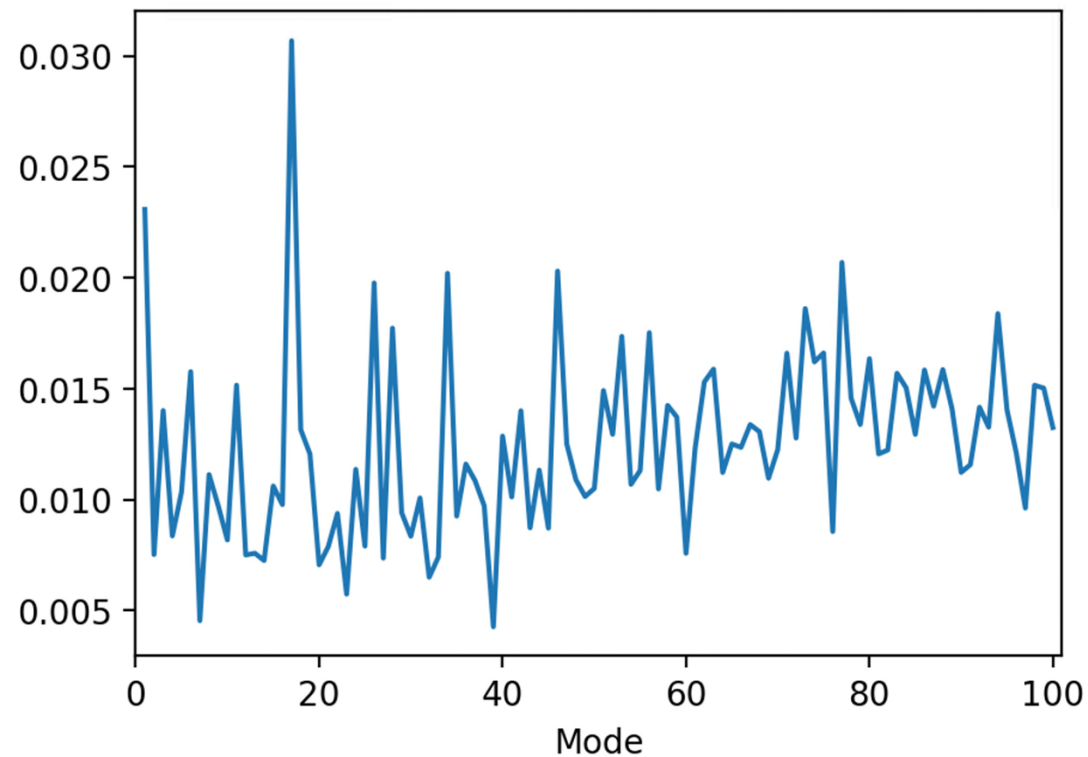
有少數幾個Mode的表現較差一些，但整體TCN表現和label相當貼近

▶ 藍色為label，橘色為TCN輸出

測試資料

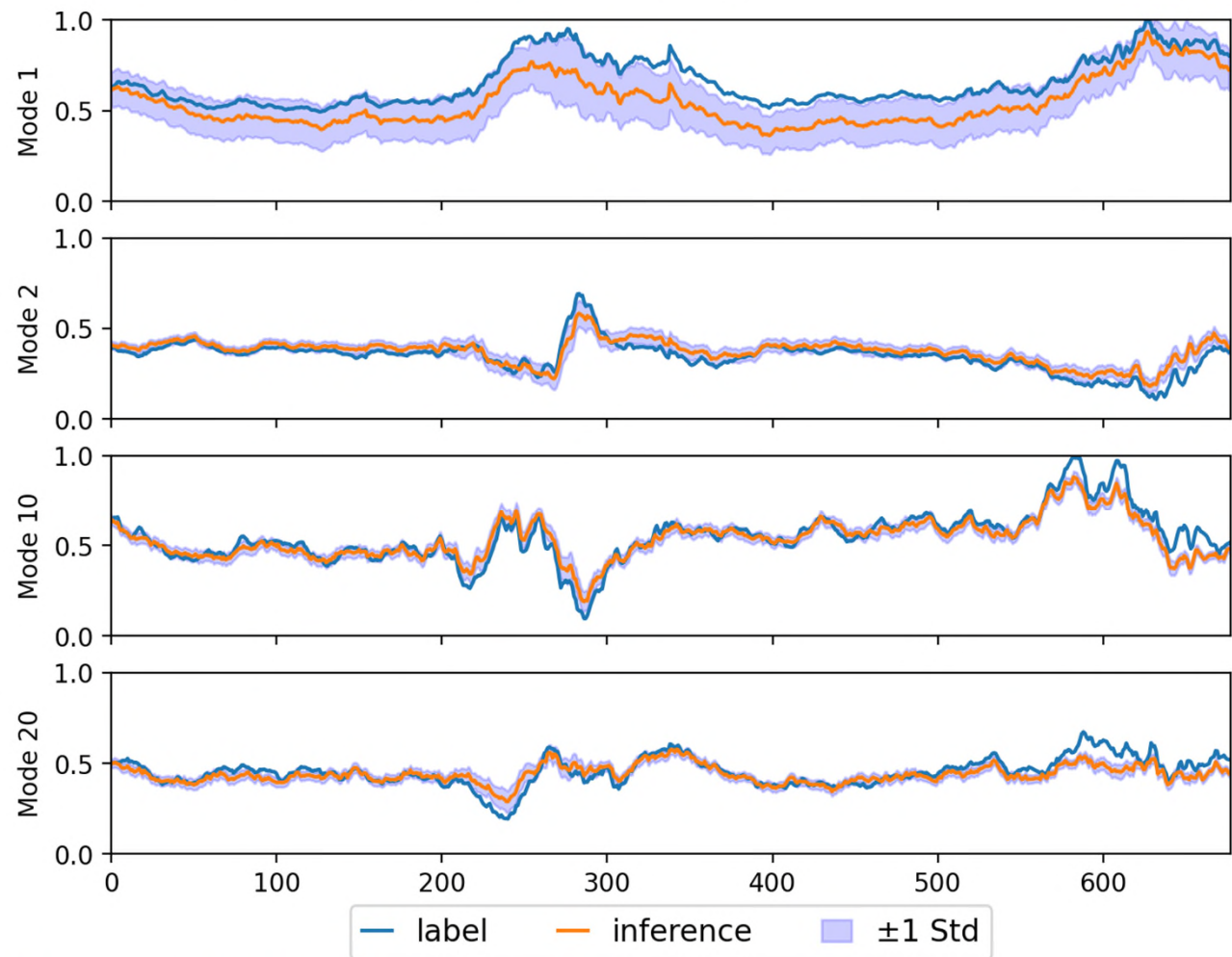


各mode MAE



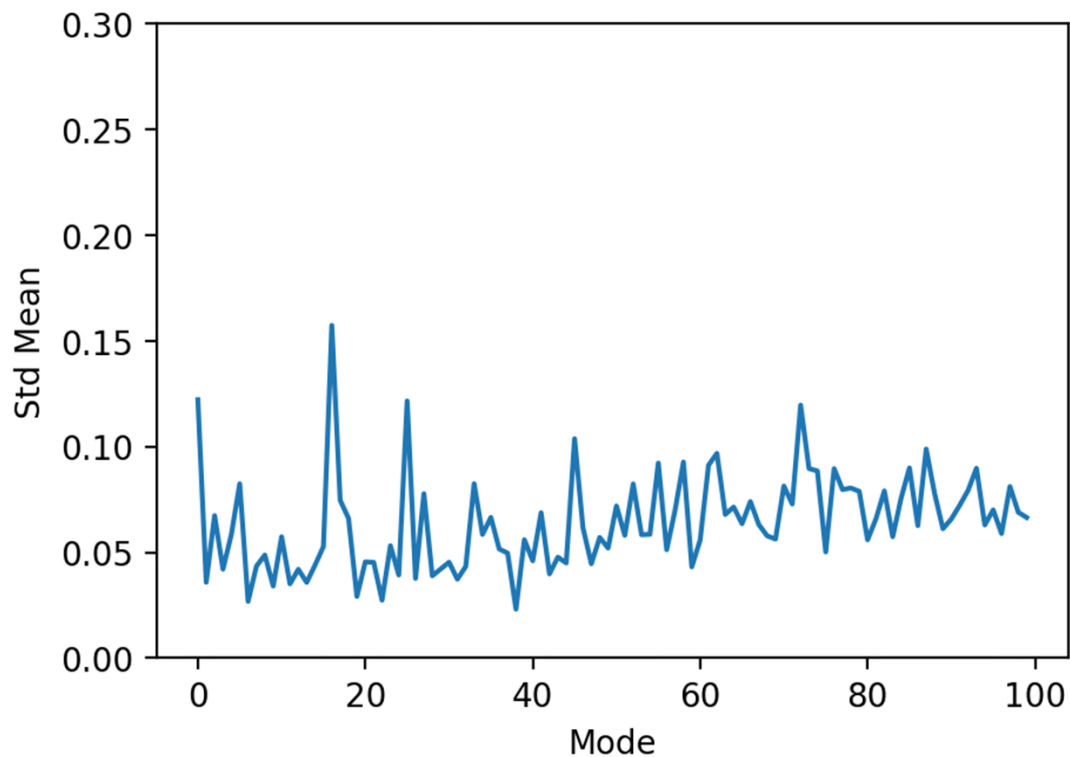
STD TCN 訓練及測試資料比較

▶ 藍色為label，橘色為 Mean TCN輸出

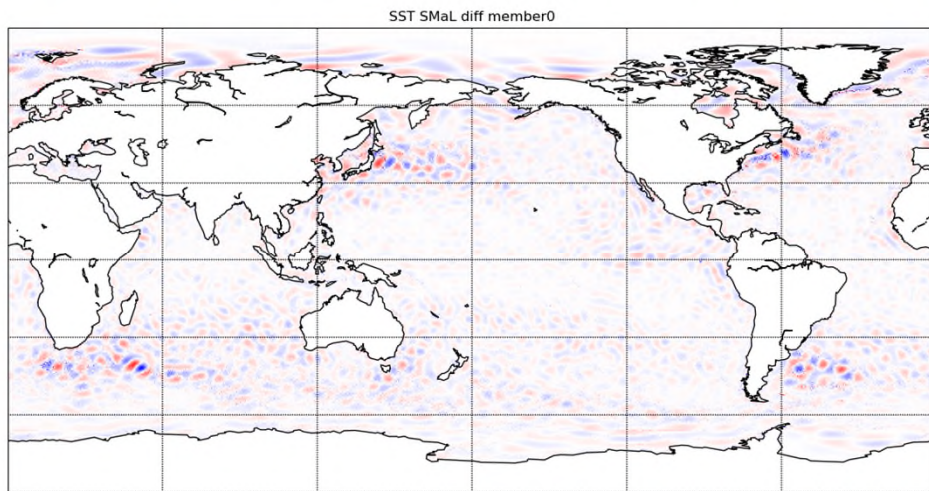


當Mean TCN報比較差的mode，STD TCN報此mode的標準差也會比較大

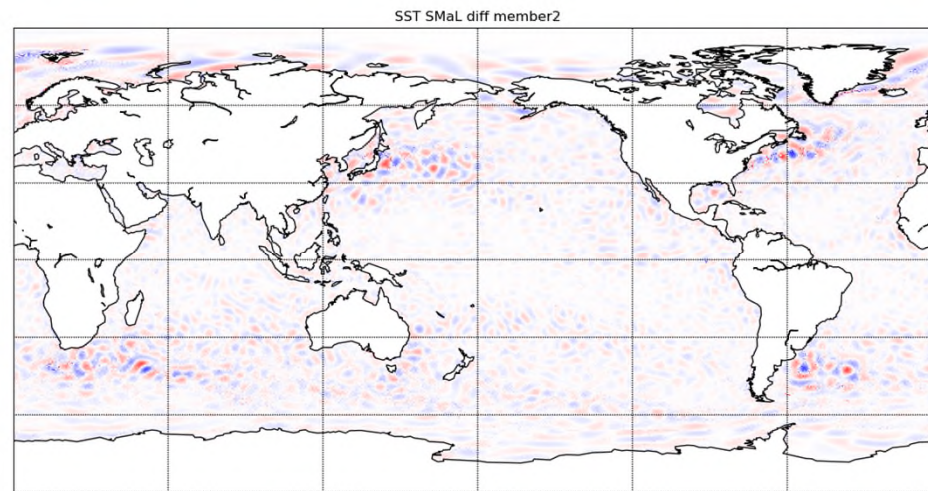
各mode STD的mean



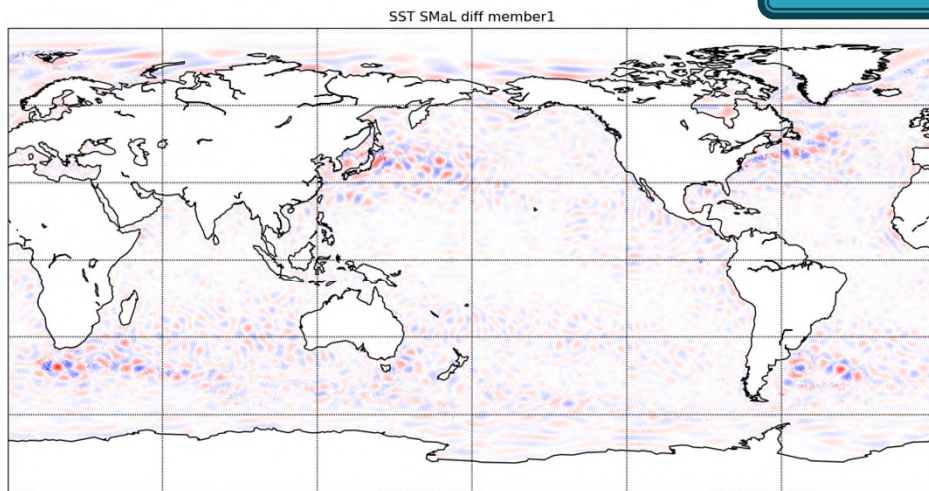
系集成員和觀測的差異



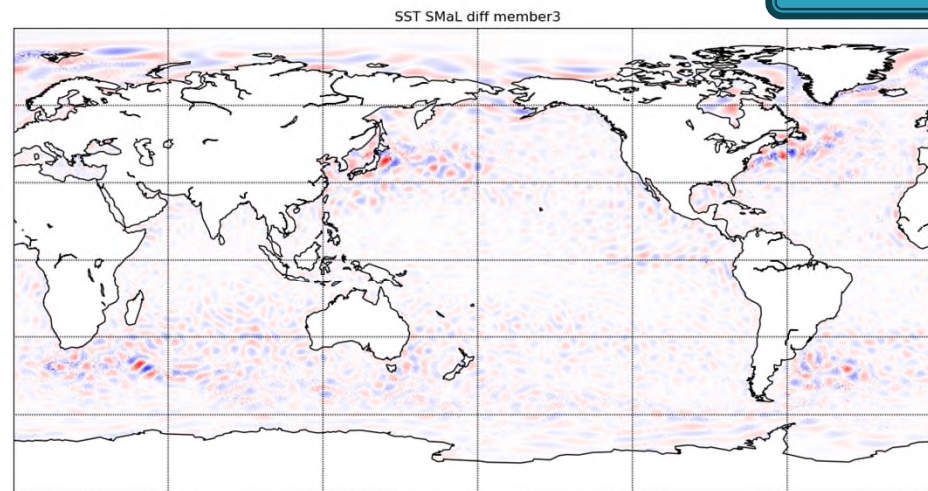
成員 0



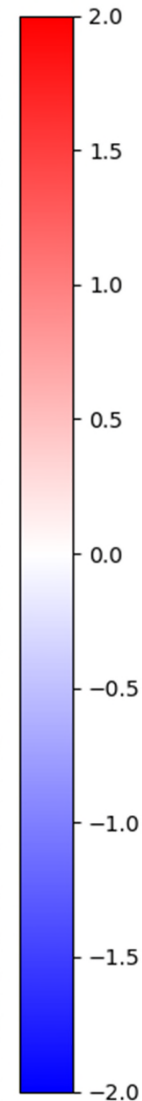
成員 2



成員 1



成員 3



SMaL流程應用到預報資料-GEFS V12的T2M

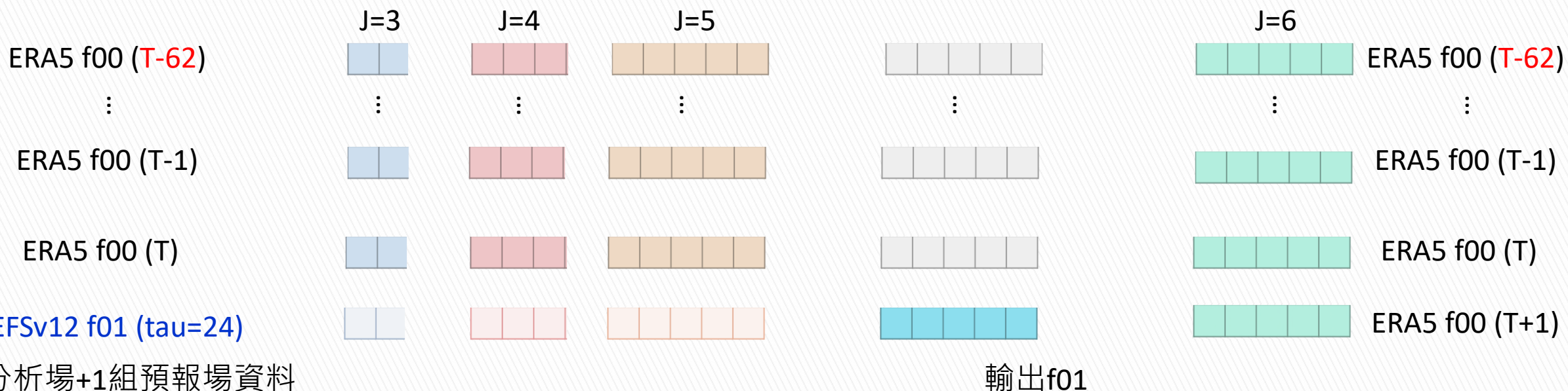
- ▶ 分析場可以成功產製有差異的系集，接下來想試試看是否使用SMaL流程也可以產製系集預報場
 - 文獻中為測試分析場
- ▶ 應用到預報上需要調整不少東西
 - 使用單一的決定性預報來產置系集成員，變數為東亞區域的T2M
 - 訓練資料的放置方式，原先只有使用分析場來建置模型，所以需要調整
 - 像是從連續64天的分析場，改成第1至8天的預報場以及前55天至預報當天的分析場
 - 要使用的EOF mode數也需要調整

調整訓練資料的放置方式- 預報第01日整點溫度為例

Input shape = (None, 64, $8^2+16^2+32^2$) Output (None, 64, 30)

J=3、4、5 小波分析資料 f_scal

Label (None, 64, 30)
J=6 小波分析 f_scal 之
EOF mode 1~30 時間序列

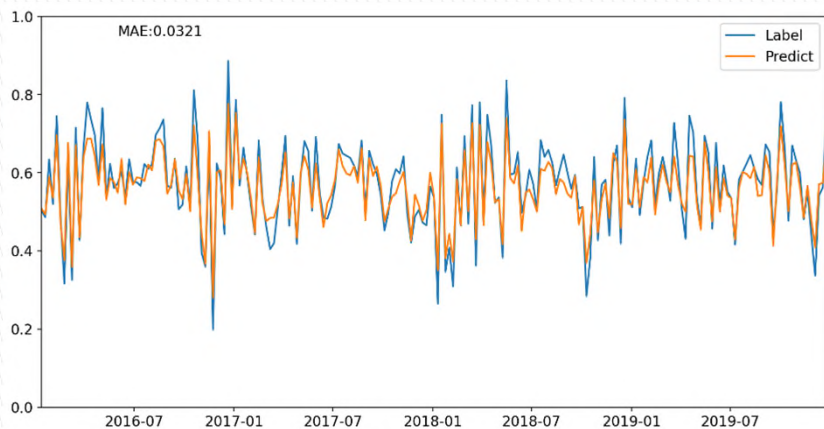


| | train | valid | test |
|---------|-------|-------|------|
| samples | 618 | 206 | 207 |

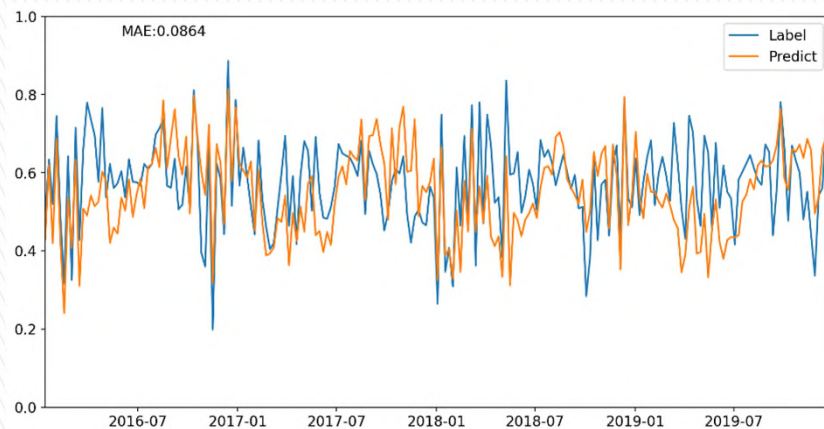
GEFSv12 reforecast 僅每週三預報未來29天

分析場+預報場建模-預報第01/08/15/22/29日

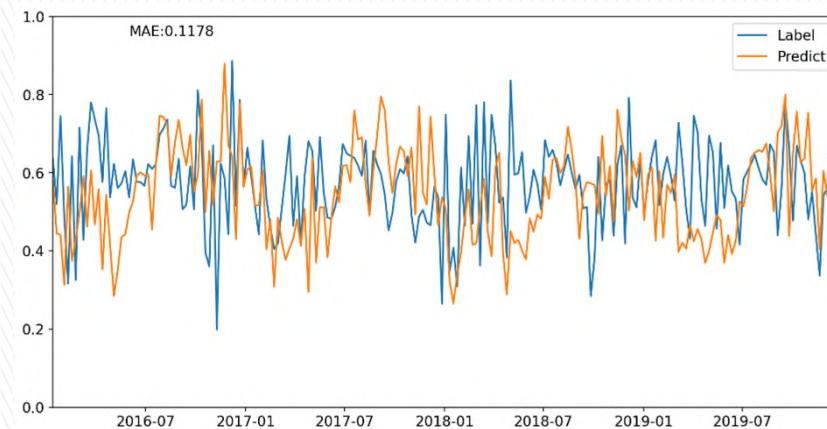
預報第01日



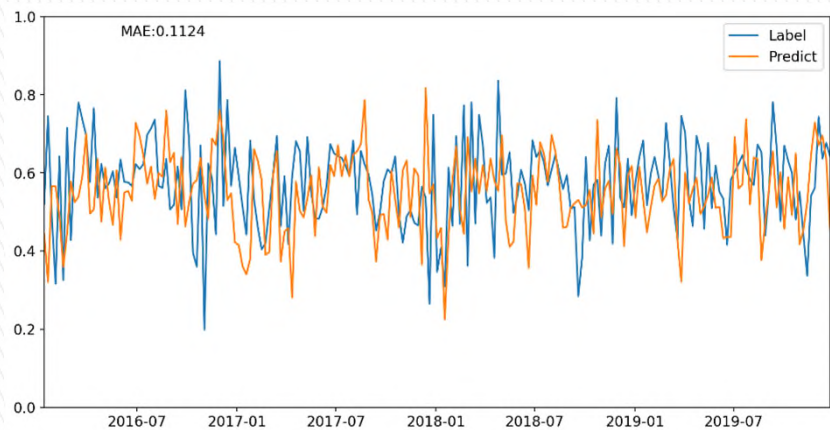
預報第08日



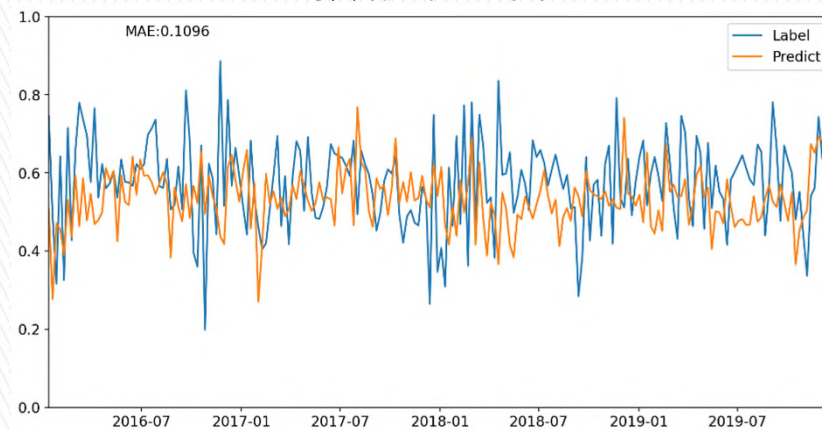
預報第15日



預報第22日

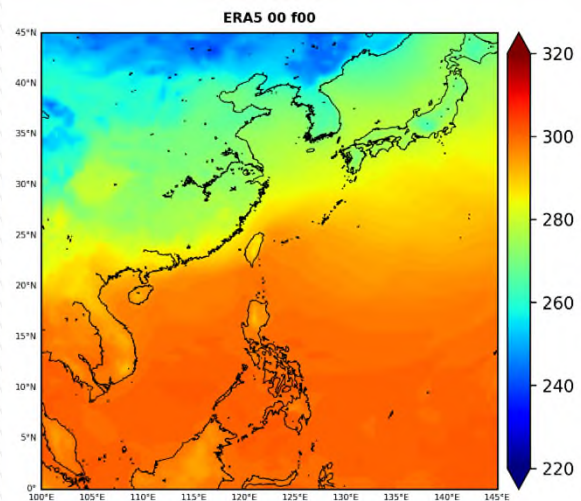


預報第29日



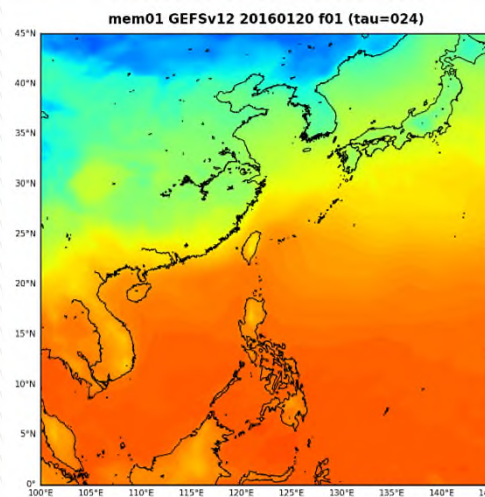
溫度場空間分布- 20160121 00Z 預報第一日

label (ERA5)

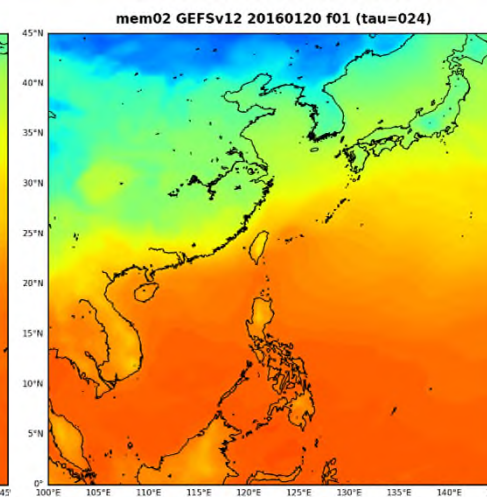


GEFSv12

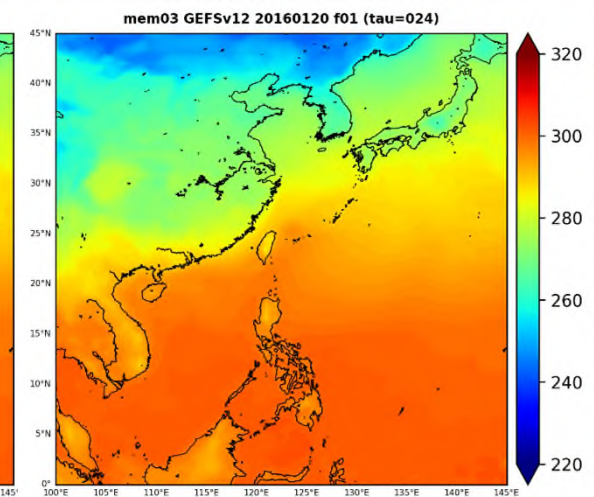
ENS1



ENS2

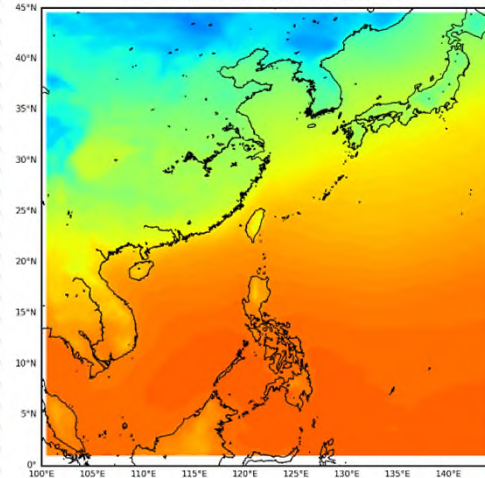


ENS3

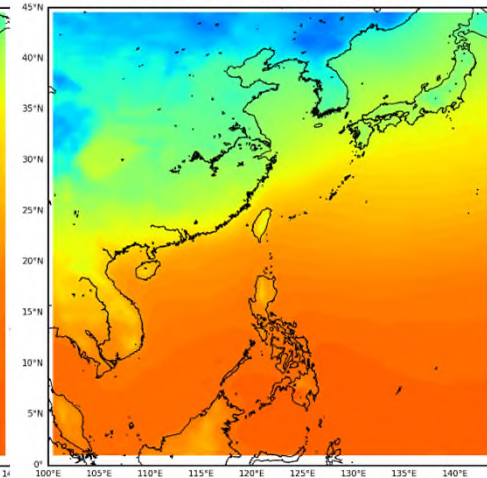


SMaL

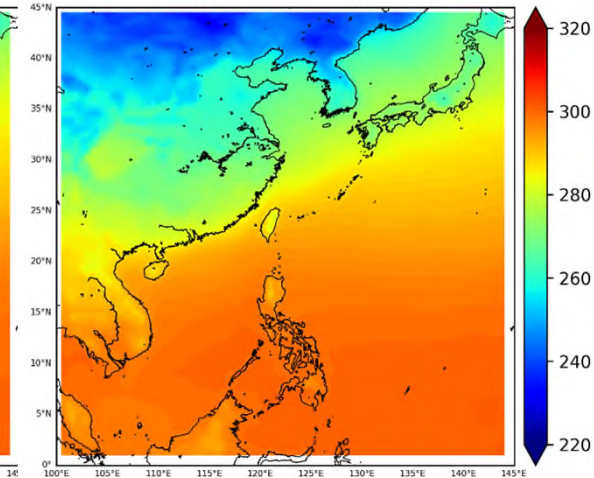
mem01 SMaL 20160120 f01 (tau=024)



mem02 SMaL 20160120 f01 (tau=024)



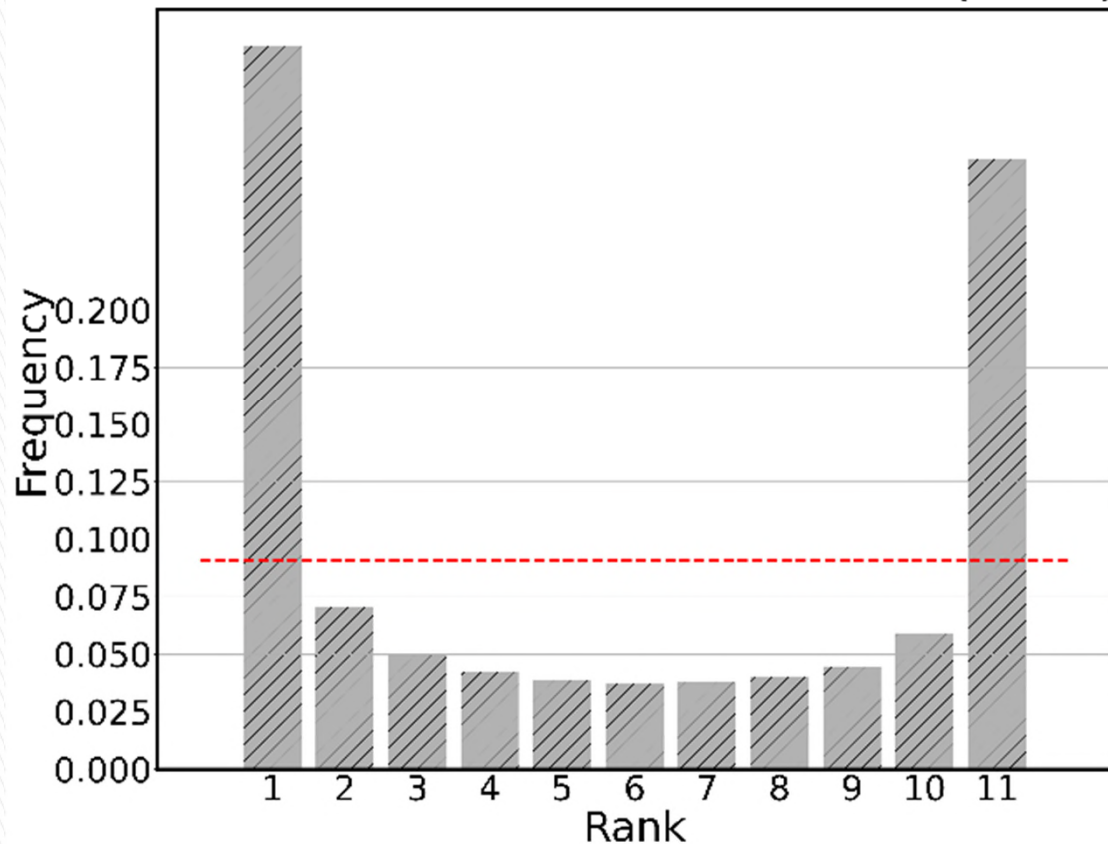
mem03 SMaL 20160120 f01 (tau=024)



系集離散度評估-預報第1日

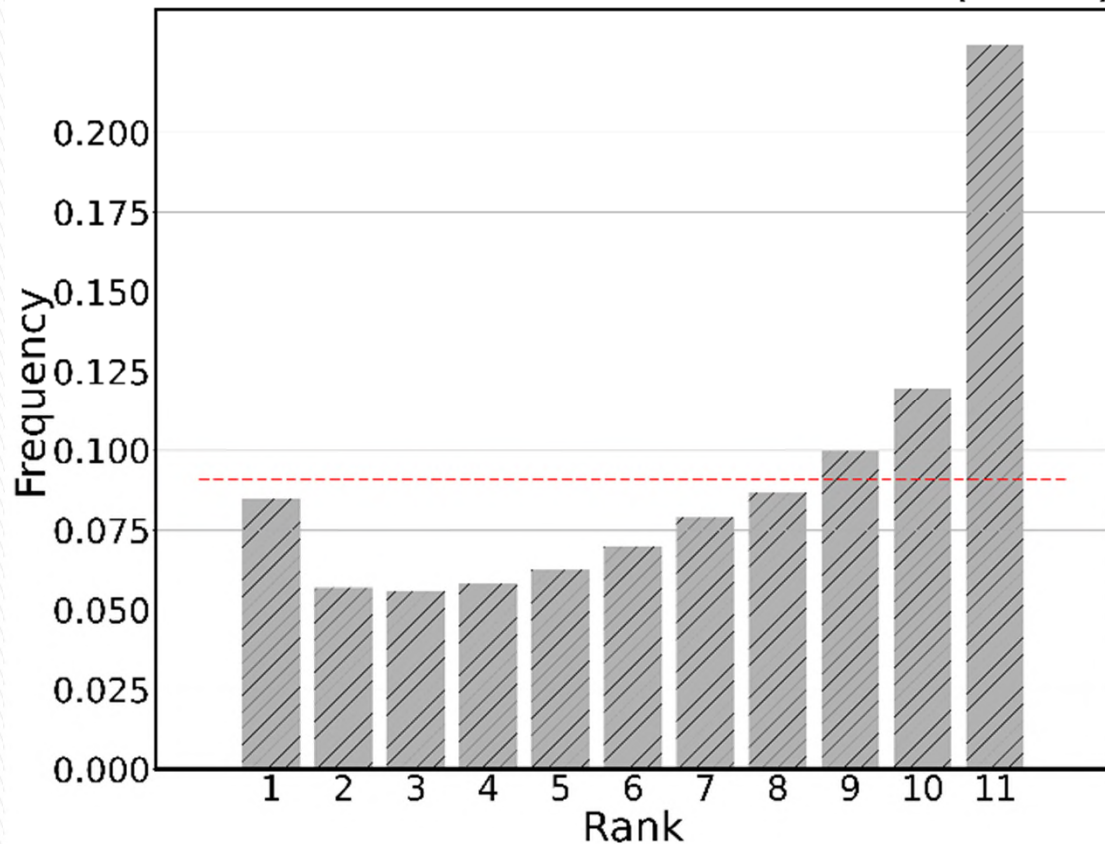
GEFSv12系集

Rank Histogram f01 (tau=024)



混合式TCN模型

Rank Histogram f01 (tau=024)



小結

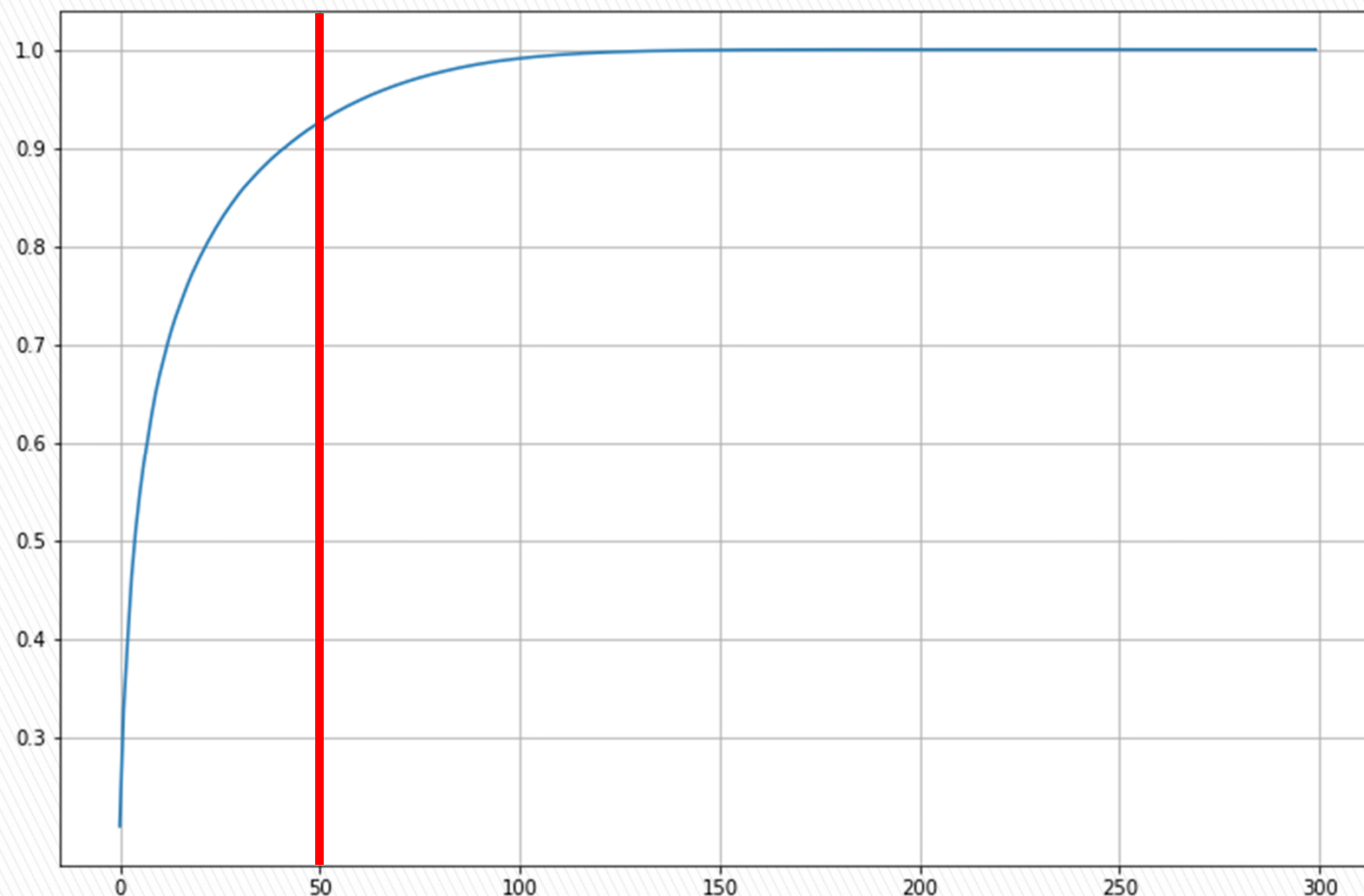
- ▶ SMaL可以產出有差異的系集分析場
 - 只使用單一分析場就能產出有差異的系集成員
 - 其成員和觀測在大尺度特徵上類似，但小尺度特徵有各有差異
 - 以RH圖來看，原始系集的分布略為集中
- ▶ SMaL流程套用至預報場也可以產出有差異的系集預報場
 - 使用決定系預報可以產出有差異的系集成員
 - 不同模式所需要使用的標準差倍率會不同
 - 調整取樣標準差後產製出系集的分布還算均勻

簡報結束
敬請指教

www.manysplendid.com

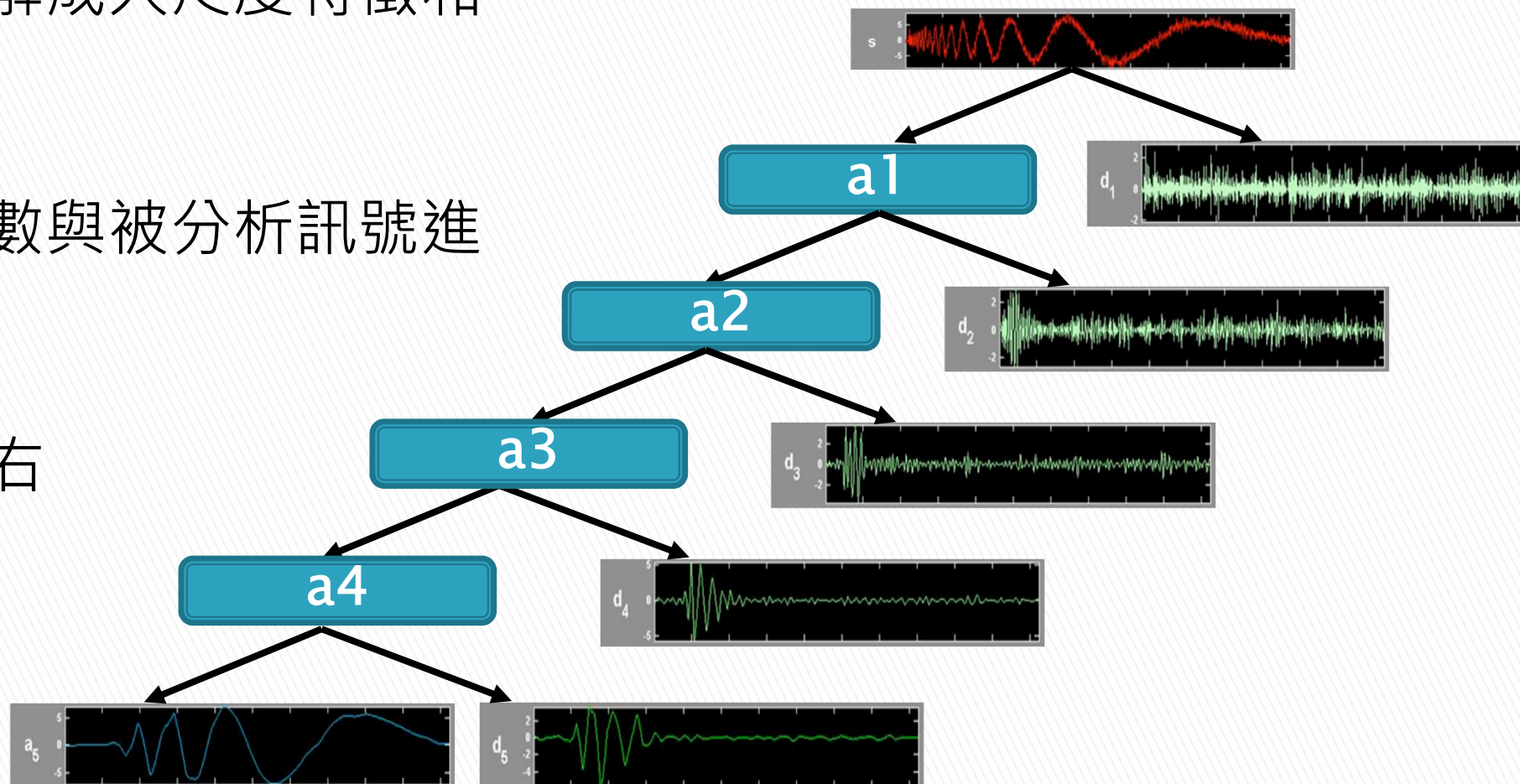
兩米溫度

- ▶ EOF各mode貢獻量
 - 取50mode，約92%



小波分析

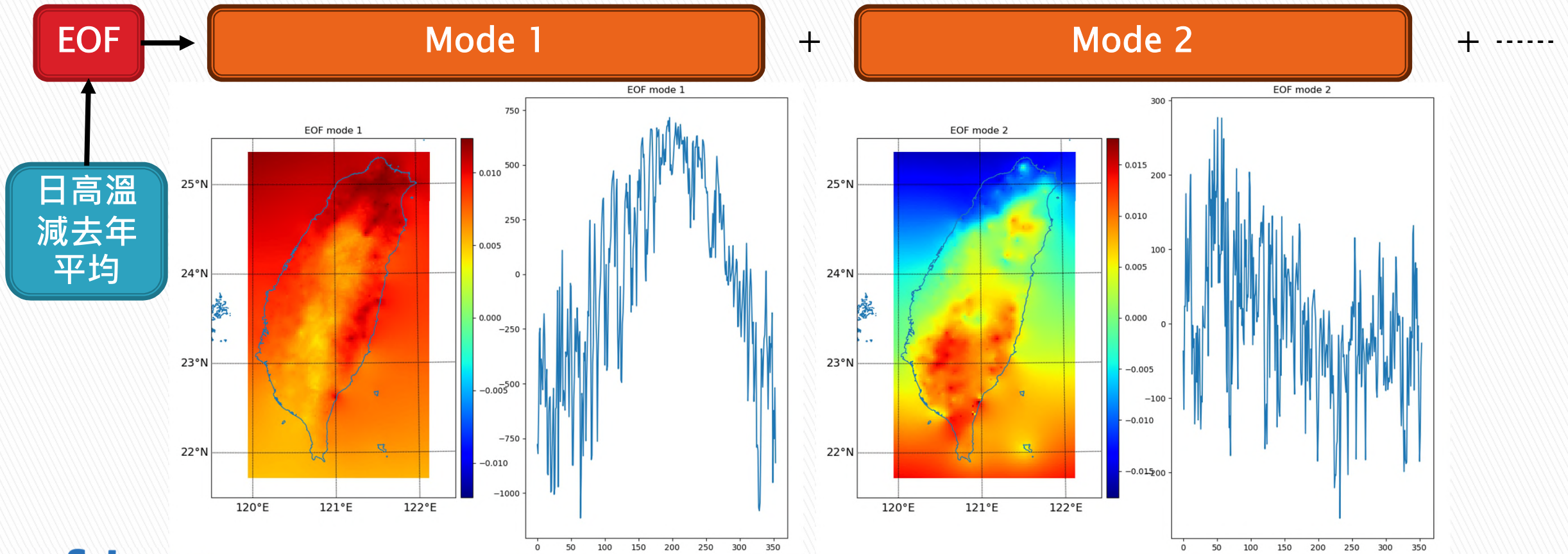
- ▶ 將原始資料拆解成大尺度特徵和小尺度特徵
- ▶ 利用小波母函數與被分析訊號進行積分計算
- ▶ 一維示意圖如右



(ROBI POLIKAR, THE WAVELET TUTORIAL)

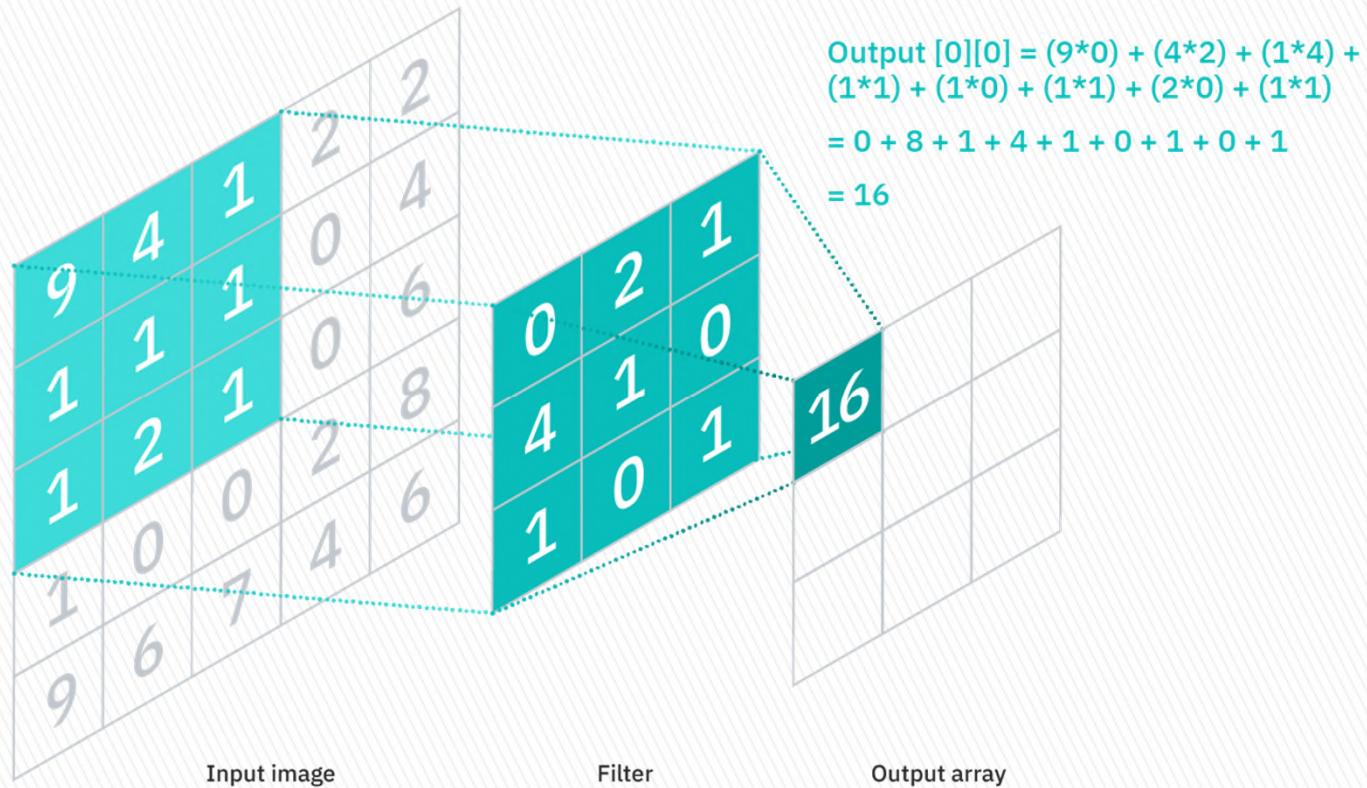
EOF(empirical orthogonal function)

- ▶ 特色是可以將時空間資料拆解成多組mode組成的時空間特徵
 - 每一mode會有一個空間分布和一個時間序列



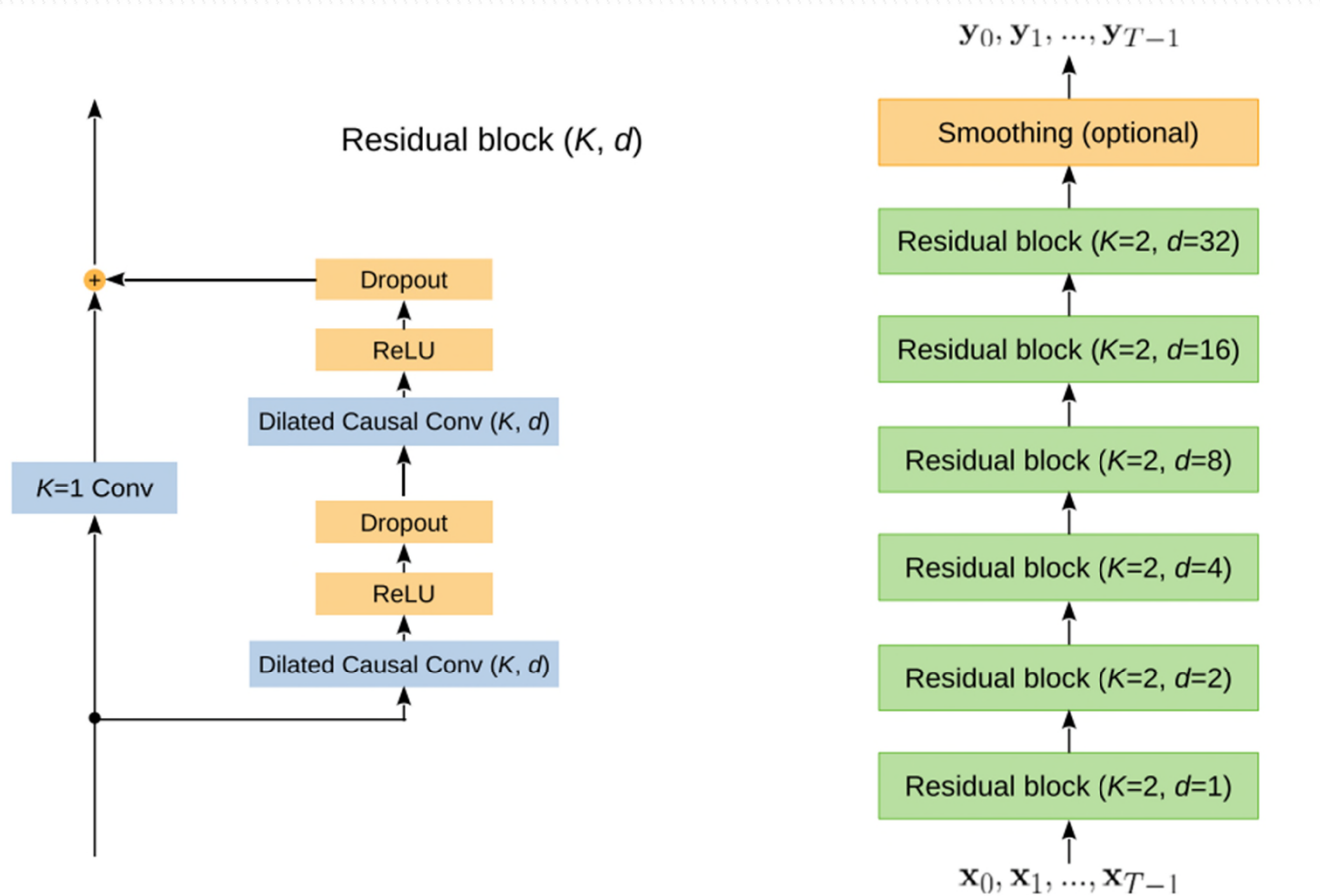
卷積

- ▶ 將輸入資料透過不同Filter做計算後得到不同特徵



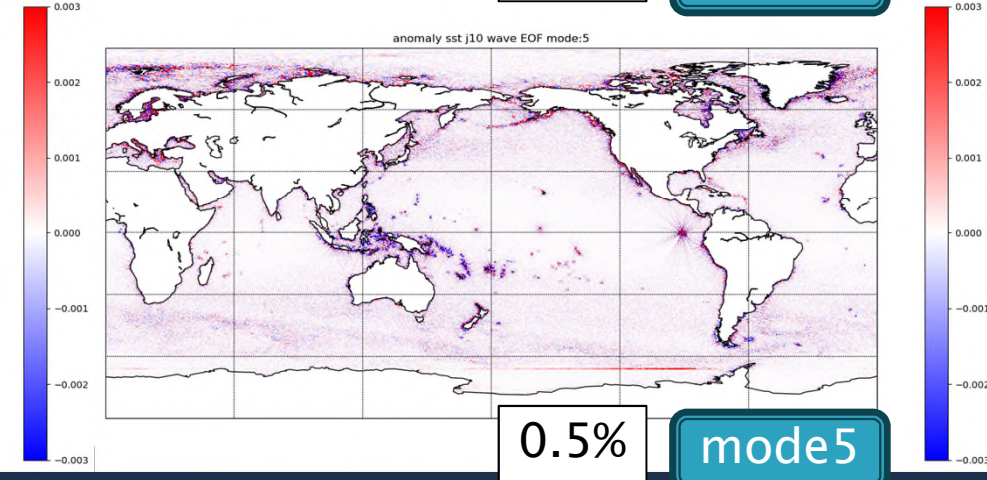
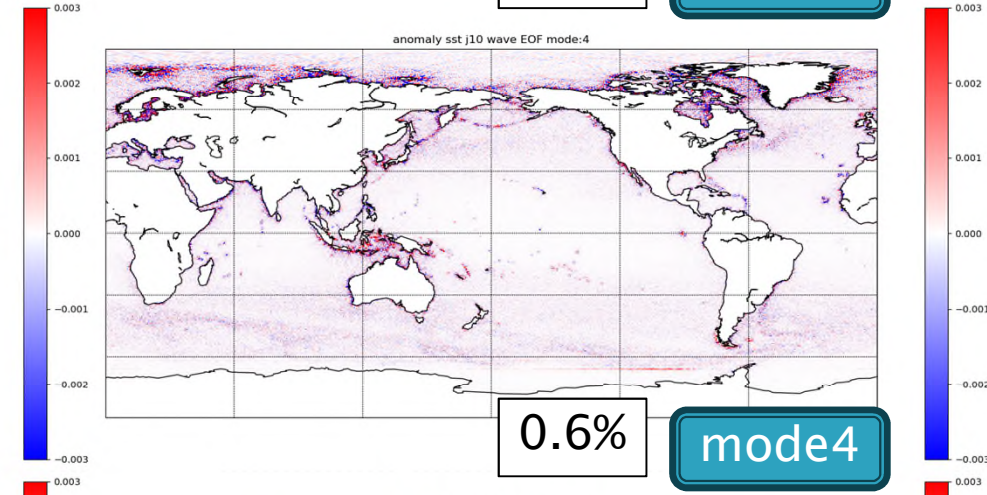
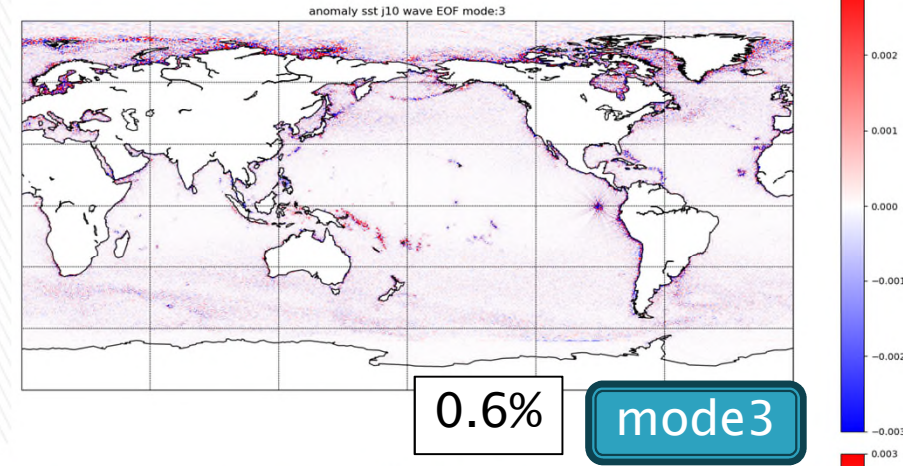
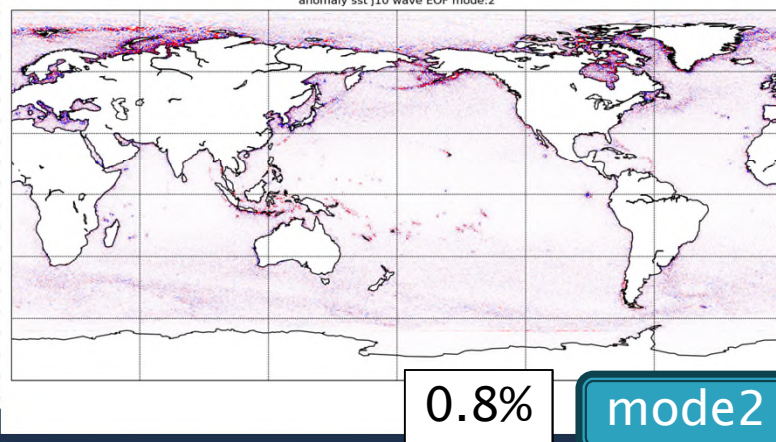
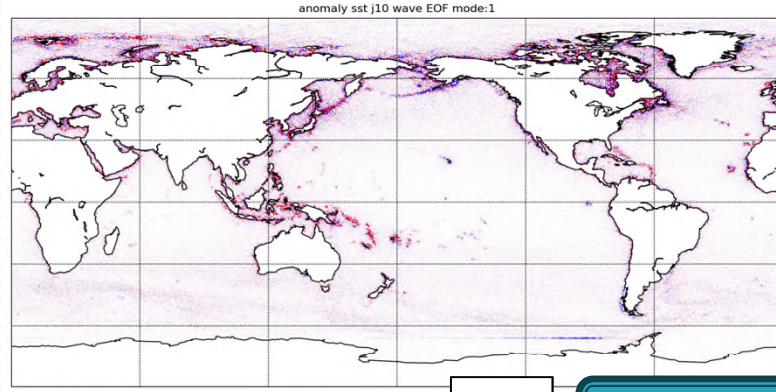
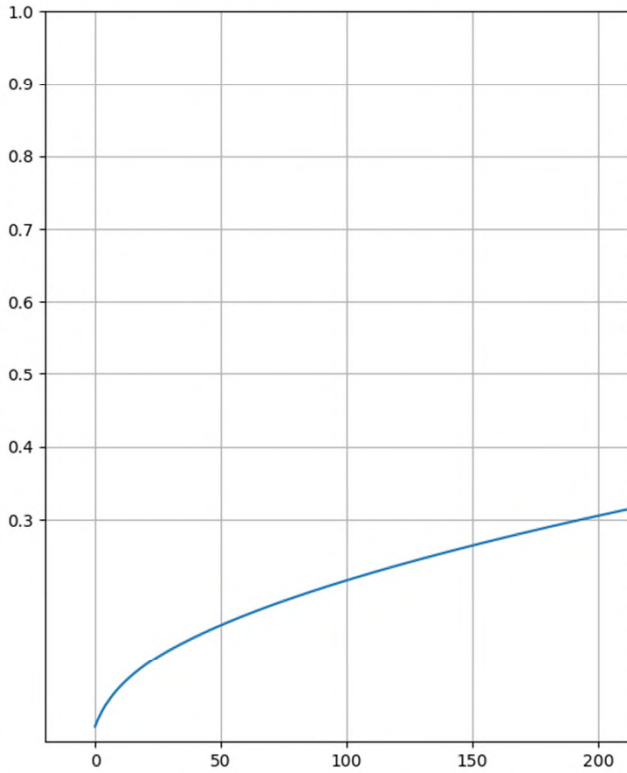
TCN架構

▶ 共六層residual block



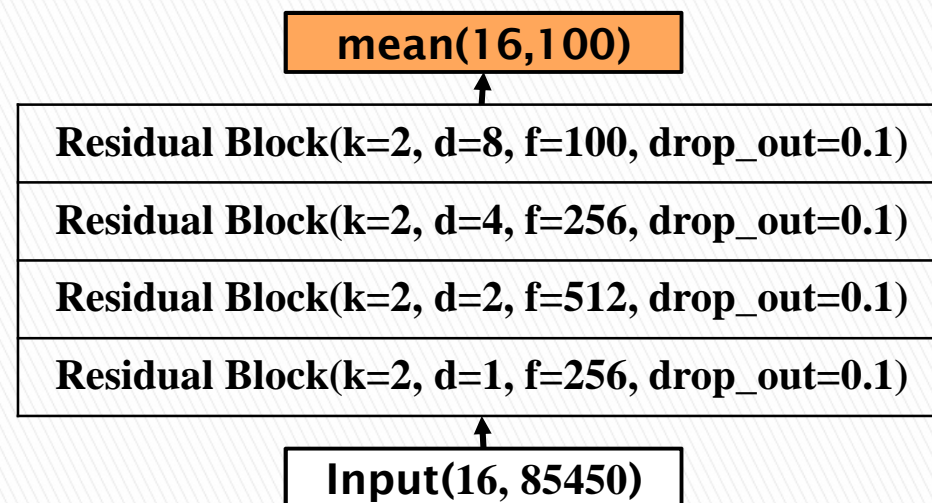
EOF J=10 wave Mode1~5

貢獻度



j=10 scale 建置 j=10 wave Mean TCN- 本次設定

- ▶ 資料窗區為16天
- ▶ 逐層 Residual block 的 dropout rate 都是 0.1
- ▶ Loss: MAE
- ▶ max epoch=5,000
- ▶ batch_size=50
- ▶ Residual block
 - [256,512,256,100]



| train | valid | test |
|-------|-------|------|
| 2029 | 677 | 677 |

輸入為j=10的scale
輸出為j=10的wave
將training和validation資料進行洗牌(shuffle)
後才開始訓練，比例不變
loss function為計算最後1天MAE來訓練