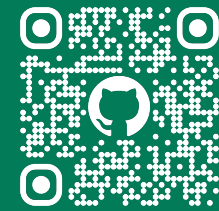


KDD Paper



Source code

# 時間変化する因果関係の抽出に 基づいた高速将来予測

千原直己<sup>†,‡</sup> 松原靖子<sup>†</sup> 藤原廉<sup>†,‡</sup> 櫻井保志<sup>†</sup>

† 大阪大学産業科学研究所

‡ 大阪大学大学院情報科学研究科



# 題目

- 研究背景
- 提案手法
- 評価実験
- まとめ

# 題目

- **研究背景**
- 提案手法
- 評価実験
- まとめ

# 研究背景：時系列データ

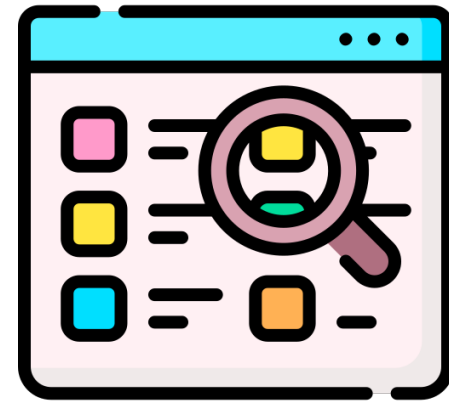
- 時系列データは様々な分野において日々生成されている



**Motion analysis**



**Epidemiology**



**Web activity**

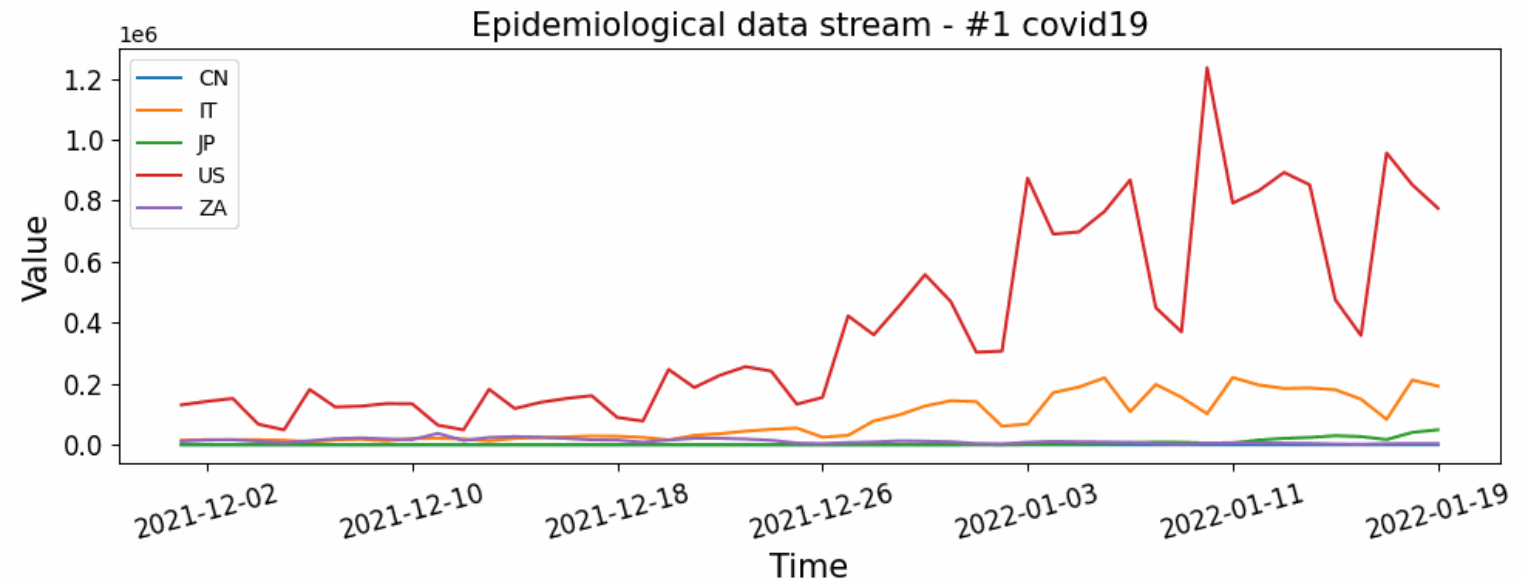


# 研究背景：時系列データ

- 時系列データは様々な分野において日々生成されている
- 現実の環境では，これらのデータは迅速かつ継続的に生成される



## Epidemiology



# 研究背景：依存関係

- 時系列データ中の変数間の関係性は重要な特徴量の1つ
  - ❖ e.g., 相関関係, 因果関係, 独立性など

# 研究背景：依存関係

- 時系列データ中の変数間の関係性は重要な特徴量の1つ
  - ❖ e.g., 相関関係, **因果関係**, 独立性など

- 原因と結果の関係を意味する
- 長年研究の対象となっている

# 研究背景：依存関係

- 時系列データ中の変数間の関係性は重要な特徴量の1つ
  - ❖ e.g., 相関関係, **因果関係**, 独立性など
- 因果探索のための大半の既存手法は, 因果関係がデータ内で変動しないことを仮定している 🥲
  - ❖ 複雑な実世界への応用のためには, 因果構造の時間変化を考慮することが不可欠である

# 研究背景：依存関係

- 時系列データ中の変数間の関係性は重要な特徴量の1つ
  - ❖ e.g., 相関関係, **因果関係**, 独立性など
- 因果探索のための大半の既存手法は, 因果関係がデータ内で変動しないことを仮定している 🙄
  - ❖ 複雑な実世界への応用のためには, 因果構造の時間変化を考慮することが不可欠である
  - ❖ この時間依存性はさまざまな影響による環境の移り変わりに従うと考えられる

# 研究背景：依存関係

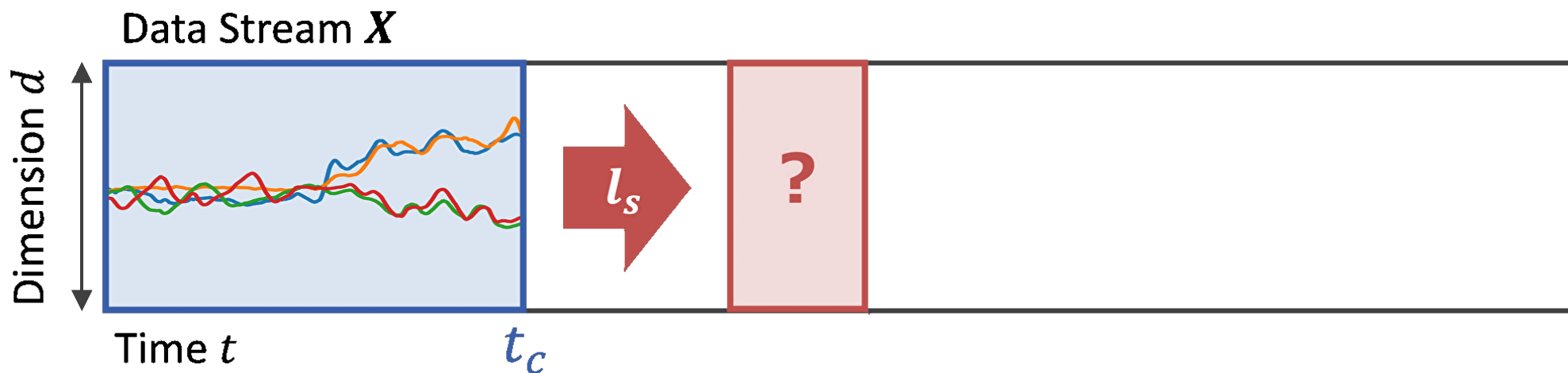
- ▶ 時系列データ中の変数間の関係性は重要な特徴量の1つ
  - ❖ e.g., 相関関係, 因果関係, 独立性など

時間変化する因果関係の抽出および**将来予測**を同時に達成するための最新手法である **ModePlait** を提案

# 研究背景：問題定義

- **Given:** 時系列データストリーム  $X = \{x(1), \dots, x(t_c), \dots\}$  ( $t_c$ : 現在時刻)
- **Goals:** 以下の重要な課題を全て達成する
  - ❖ 特徴的な時系列パターン（レジーム）の発見
  - ❖ 時間変化する因果関係の抽出
  - ❖  $l_s$  ステップ先の値の予測

半無限長な時系列データ



# 題目

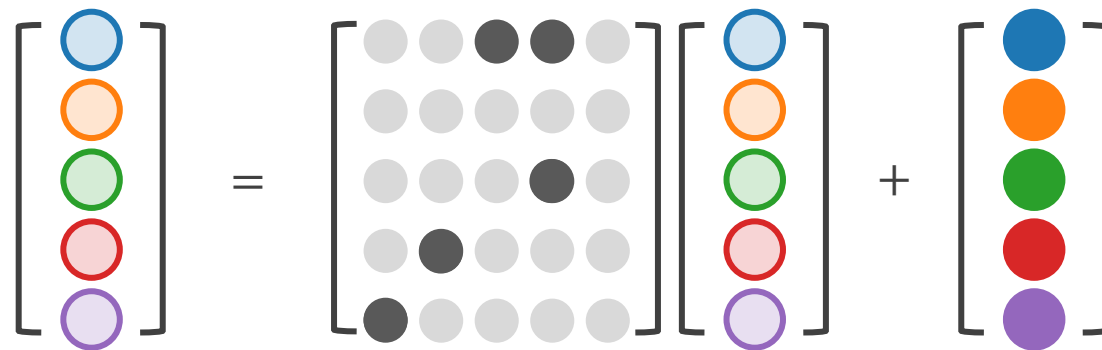
- 研究背景
- **提案手法**
- 評価実験
- まとめ



# 提案手法：概要

- 提案手法は構造方程式モデル [Pearl 2009] に基づいて設計した

$$\begin{array}{c} \underline{X_{\text{sem}}} \\ \text{観測変数} \end{array} = \begin{array}{c} \underline{B_{\text{sem}}} \\ \text{因果隣接行列} \end{array} X_{\text{sem}} + \begin{array}{c} \underline{E_{\text{sem}}} \\ \text{外生変数} \end{array}$$

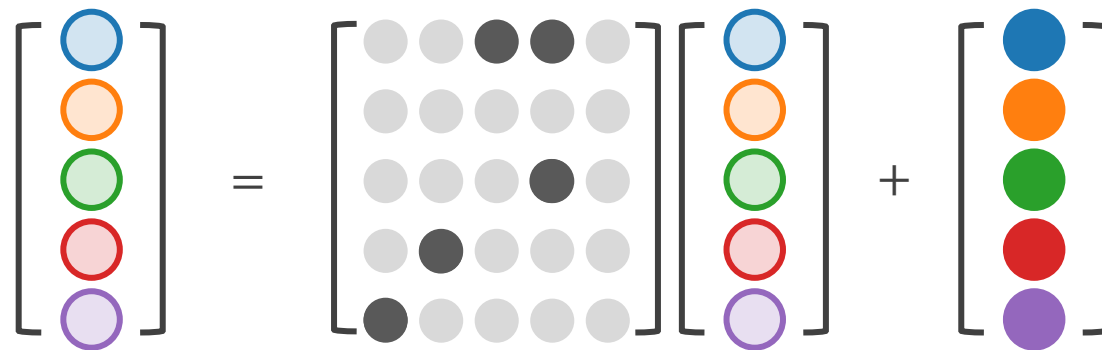


一般的な構造方程式モデル (SEM)

# 提案手法：概要

- 提案手法は構造方程式モデル [Pearl 2009] に基づいて設計した

$$\underbrace{X_{\text{sem}}}_{\text{観測変数}} = \underbrace{B_{\text{sem}}}_{\text{因果隣接行列}} X_{\text{sem}} + \underbrace{E_{\text{sem}}}_{\text{外生変数}}$$



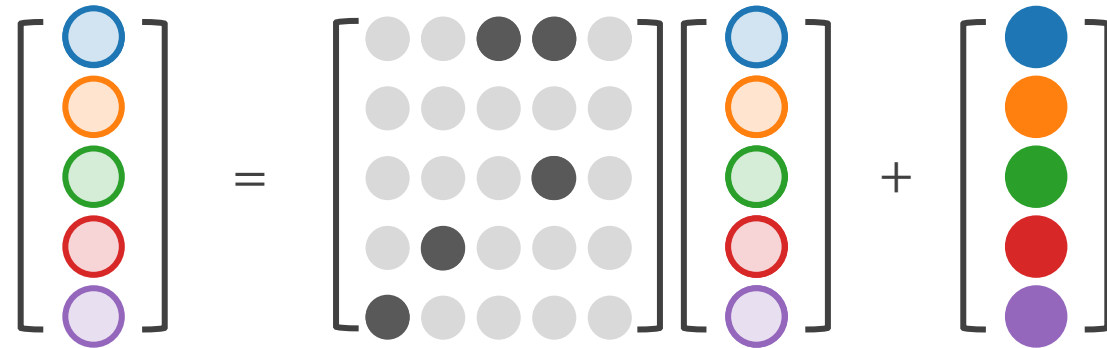
関係があれば ●  
関係が無ければ ○

各変数に固有な成分

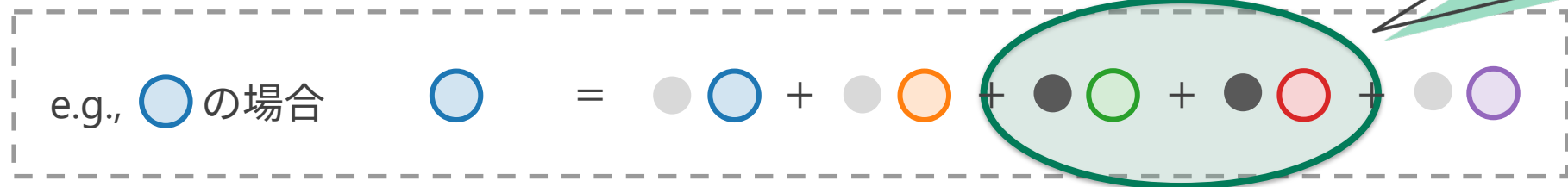
# 提案手法：概要

- 提案手法は構造方程式モデル [Pearl 2009] に基づいて設計した

$$\underbrace{X_{\text{sem}}}_{\text{観測変数}} = \underbrace{B_{\text{sem}}}_{\text{因果隣接行列}} X_{\text{sem}} + \underbrace{E_{\text{sem}}}_{\text{外生変数}}$$



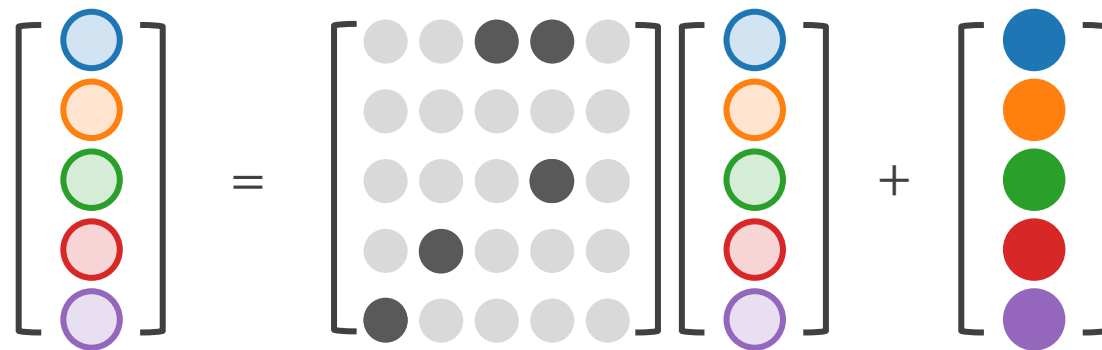
○の原因となる変数



# 提案手法：概要

- 提案手法は構造方程式モデル [Pearl 2009] に基づいて設計した

$$\underbrace{X_{\text{sem}}}_{\text{観測変数}} = \underbrace{B_{\text{sem}}}_{\text{因果隣接行列}} X_{\text{sem}} + \underbrace{E_{\text{sem}}}_{\text{外生変数}}$$



関係があれば ●  
関係が無ければ ○

各変数に固有な成分

# 提案手法：概要

- 達成すべき課題は以下のとおり
  - ❖ 固有信号を数式的に記述する
  - ❖ 因果関係を考慮して時系列データを表現する
  - ❖ 時間変化する因果関係を考慮した時系列データストリームの要約

# 提案手法：概要

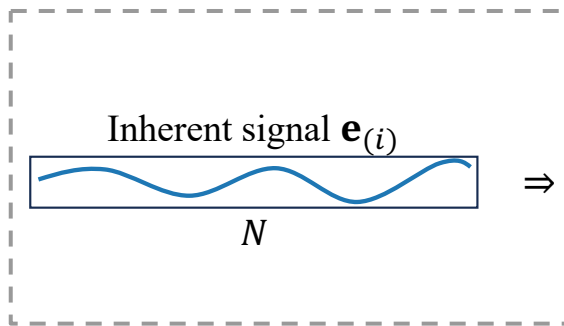
- 達成すべき課題は以下のとおり
  - ❖ 固有信号を数式的に記述する
  - ❖ 因果関係を考慮して時系列データを表現する
  - ❖ 時間変化する因果関係を考慮した時系列データストリームの要約



1. 固有信号の潜在的な時間ダイナミクス
2. 特徴的な時系列パターン
3. レジームの動的な遷移

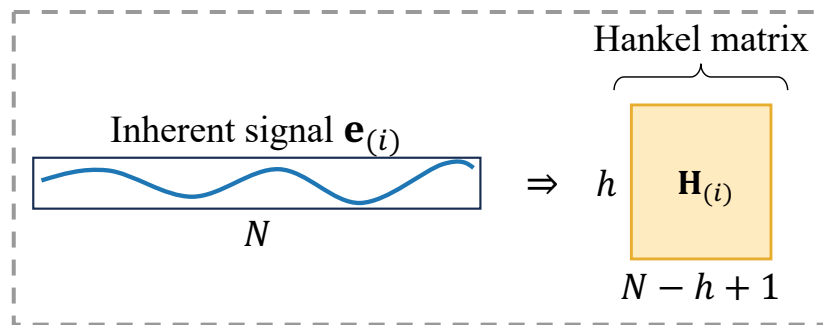
# 固有信号の潜在的な時間ダイナミクス

- 独立性により単変量時系列  $e_{(i)}$  を表現する必要がある
  - ❖ 一般的に単次元データでは情報が不十分 😞



# 固有信号の潜在的な時間ダイナミクス

- 独立性により単変量時系列  $e_{(i)}$  を表現する必要がある
  - ❖ 一般的に単次元データでは情報が不十分 🙄
  - ❖ 時間遅れ埋め込みにより状態空間の次元を拡張する

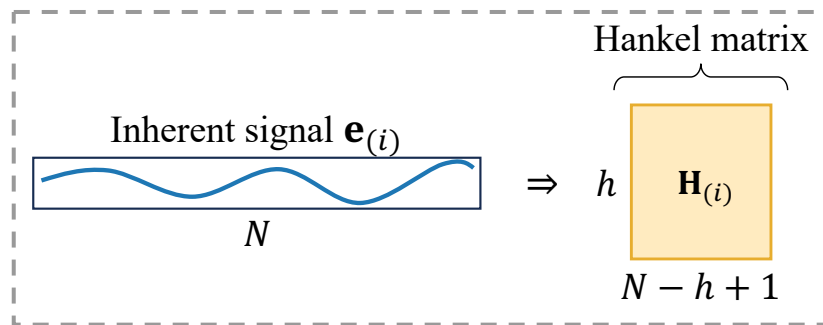


$$\mathbf{H}_{(i)} = \begin{bmatrix} | & | & & | \\ g(e_{(i)}(h)) & g(e_{(i)}(h+1)) & \cdots & g(e_{(i)}(t)) \\ | & | & & | \end{bmatrix}$$



# 固有信号の潜在的な時間ダイナミクス

- 独立性により単変量時系列  $e_{(i)}$  を表現する必要がある
  - ❖ 一般的に単次元データでは情報が不十分 🙄
  - ❖ 時間遅れ埋め込みにより状態空間の次元を拡張する

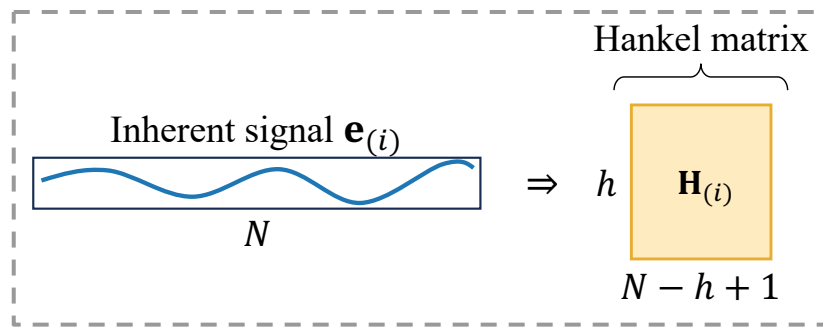


$$\mathbf{H}_{(i)} = \begin{bmatrix} | & | & \cdots & | \\ g(e_{(i)}(h)) & g(e_{(i)}(h+1)) & \cdots & g(e_{(i)}(t)) \\ | & | & & | \end{bmatrix}$$

$$g(e_{(i)}(t)) := (e_{(i)}(t), e_{(i)}(t-1), \dots, e_{(i)}(t-h+1)) \in \mathbb{R}^h$$

# 固有信号の潜在的な時間ダイナミクス

- 独立性により単変量時系列  $e_{(i)}$  を表現する必要がある
  - ❖ 一般的に単次元データでは情報が不十分 😞
  - ❖ 時間遅れ埋め込みにより状態空間の次元を拡張する



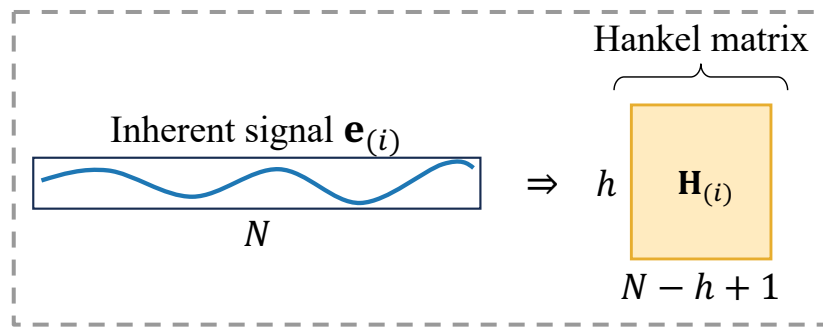
$$\mathbf{H}_{(i)} = \begin{bmatrix} | & | & \cdots & | \\ g(e_{(i)}(h)) & g(e_{(i)}(h+1)) & \cdots & g(e_{(i)}(t)) \\ | & | & & | \end{bmatrix}$$

$$g(e_{(i)}(t)) := (e_{(i)}(t), \underbrace{e_{(i)}(t-1), \dots, e_{(i)}(t-h+1)}_{\text{過去データ}}) \in \mathbb{R}^h$$

過去データ

# 固有信号の潜在的な時間ダイナミクス

- 独立性により単変量時系列  $e_{(i)}$  を表現する必要がある
  - ❖ 一般的に単次元データでは情報が不十分 😞
  - ❖ 時間遅れ埋め込みにより状態空間の次元を拡張する



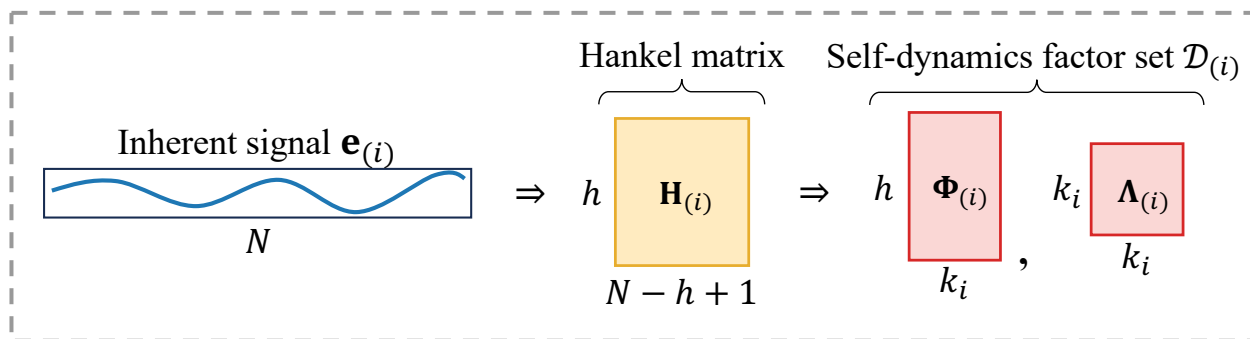
$$\mathbf{H}_{(i)} = \begin{bmatrix} | & | & \dots & | \\ g(e_{(i)}(h)) & g(e_{(i)}(h+1)) & \dots & g(e_{(i)}(t)) \\ | & | & & | \end{bmatrix}$$

Takens' theorem  
に基づいている

$$g(e_{(i)}(t)) := (e_{(i)}(t), \underbrace{e_{(i)}(t-1), \dots, e_{(i)}(t-h+1)}_{\text{過去データ}}) \in \mathbb{R}^h$$

# 固有信号の潜在的な時間ダイナミクス

- $i$  番目の固有信号  $e_{(i)}$  は以下のように表現される



固有ダイナミクス集合

$$\mathcal{D}_{(i)} = \{\Phi_{(i)}, \Lambda_{(i)}\}$$

$$\underline{s_{(i)}(t+1)} = \underline{\Lambda_{(i)}} \underline{s_{(i)}(t)} : k_i \text{次元の潜在空間}$$

潜在状態

固有値行列

$$\underline{e_{(i)}(t)} = \underline{g^{-1}}(\underline{\Phi_{(i)}} \underline{s_{(i)}(t)}) : \text{観測空間への射影 } (\mathbb{C}^{k_i} \rightarrow \mathbb{R})$$

時間遅れ  
埋め込み

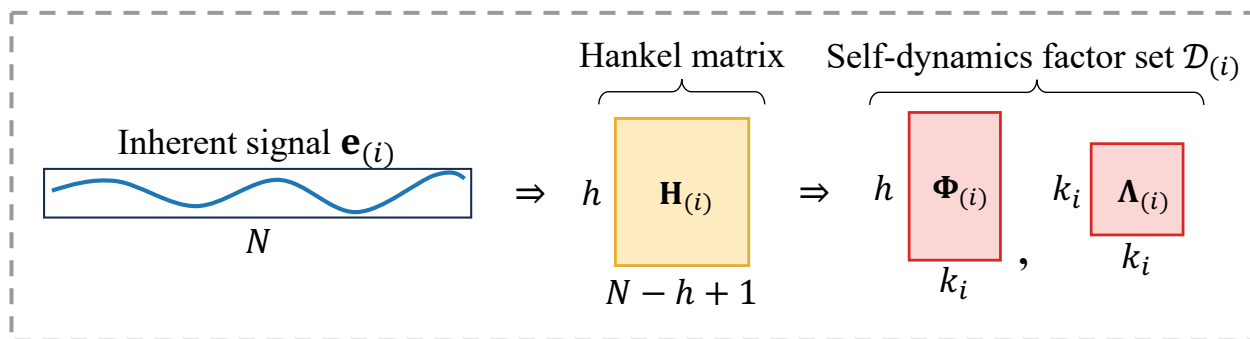
モード

次元拡張

固有信号

# 固有信号の潜在的な時間ダイナミクス

- $i$  番目の固有信号  $e_{(i)}$  は以下のように表現される



固有ダイナミクス集合

$$\mathcal{D}_{(i)} = \{\Phi_{(i)}, \Lambda_{(i)}\}$$

$$\mathbf{s}_{(i)}(t+1) = \Lambda_{(i)} \mathbf{s}_{(i)}(t) : k_i \text{次元の潜在空間}$$

潜在状態

固有値行列

$$e_{(i)}(t) = g^{-1}(\Phi_{(i)} \mathbf{s}_{(i)}(t)) : \text{観測空間への射影 } (\mathbb{C}^{k_i} \rightarrow \mathbb{R})$$

固有信号

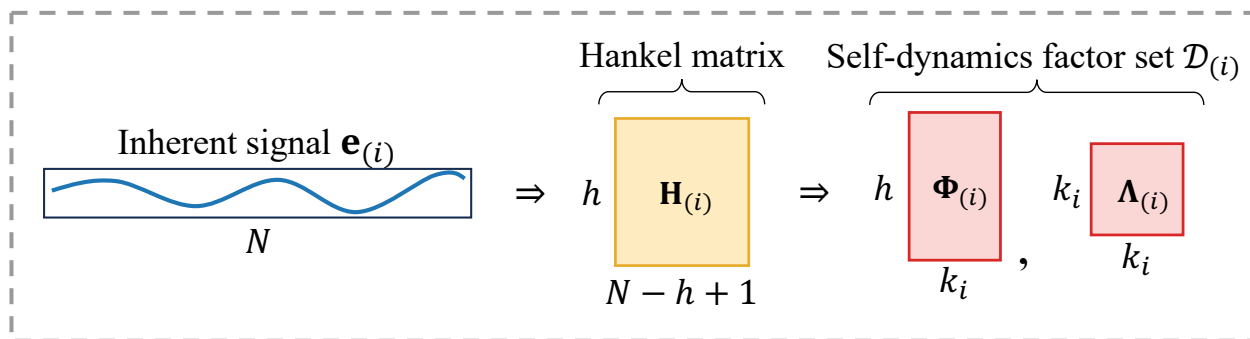
時間遅れ  
埋め込み

モード

次元拡張

# 固有信号の潜在的な時間ダイナミクス

➤  $i$  番目の固有信号  $e_{(i)}$  は以下のように表現される



固有ダイナミクス集合

$$\mathcal{D}_{(i)} = \{\Phi_{(i)}, \Lambda_{(i)}\}$$

$$\underline{s_{(i)}(t+1)} = \underline{\Lambda_{(i)}} \underline{s_{(i)}(t)} : k_i \text{次元の潜在空間}$$

潜在状態

固有値行列

$$\underline{e_{(i)}(t)} = \underline{g^{-1}}(\underline{\Phi_{(i)}} \underline{s_{(i)}(t)}) : \text{観測空間への射影 } (\mathbb{C}^{k_i} \rightarrow \mathbb{R})$$

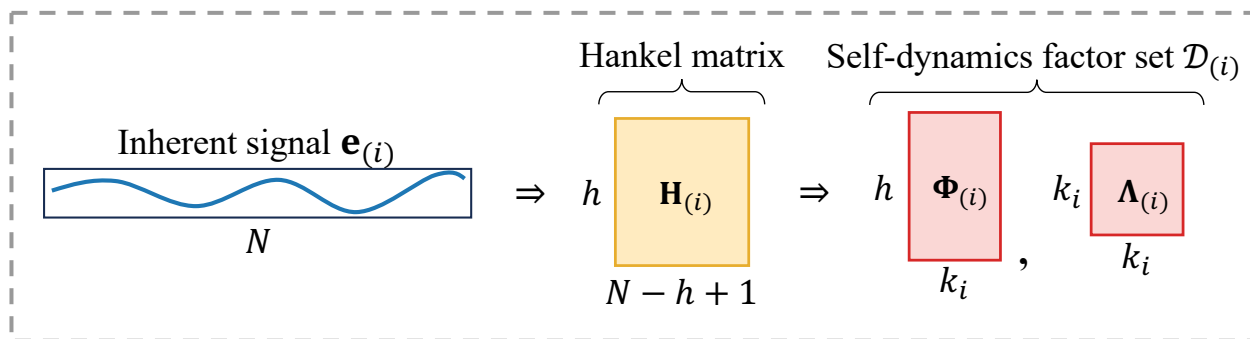
固有信号

時間遅れ  
埋め込み  
モード

次元拡張

# 固有信号の潜在的な時間ダイナミクス

➤  $i$  番目の固有信号  $e_{(i)}$  は以下のように表現される



固有ダイナミクス集合

$$\mathcal{D}_{(i)} = \{\Phi_{(i)}, \Lambda_{(i)}\}$$

$$\mathbf{s}_{(i)}(t+1) = \Lambda_{(i)} \mathbf{s}_{(i)}(t) : k_i \text{次元の潜在空間}$$

潜在状態

固有値行列

$$e_{(i)}(t) = g^{-1}(\Phi_{(i)} \mathbf{s}_{(i)}(t)) : \text{観測空間への射影 } (\mathbb{C}^{k_i} \rightarrow \mathbb{R})$$

固有信号

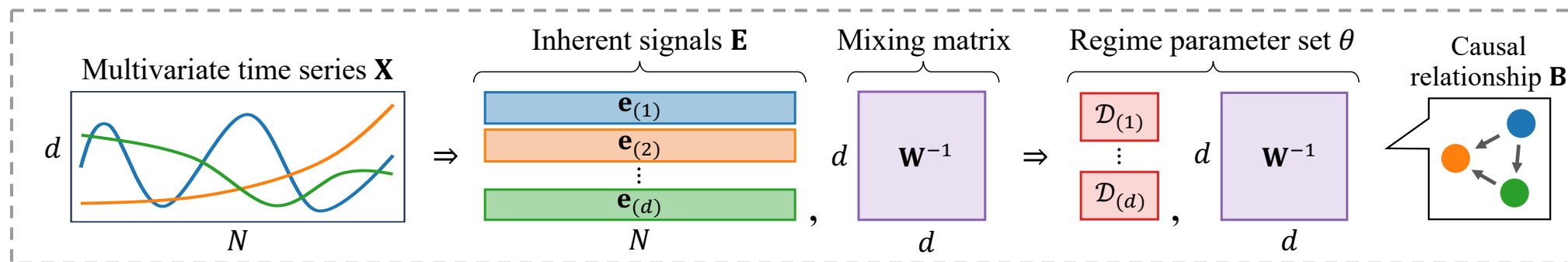
時間遅れ  
埋め込み

モード

次元拡張

# 特徴的な時系列パターン

➤ 単一の時系列パターンは以下のとおり



$$\mathbf{s}_{(i)}(t+1) = \Lambda_{(i)} \mathbf{s}_{(i)}(t) \quad (1 \leq i \leq d)$$

$$e_{(i)}(t) = g^{-1}(\Phi_{(i)} \mathbf{s}_{(i)}(t)) \quad (1 \leq i \leq d)$$

$d$  個の固有ダイナミクス集合

$$\mathbf{v}(t) = \mathbf{W}^{-1} \mathbf{e}(t) \quad (\mathbf{e}(t) = \{e_{(i)}(t)\}_{i=1}^d)$$

推定値

混合行列

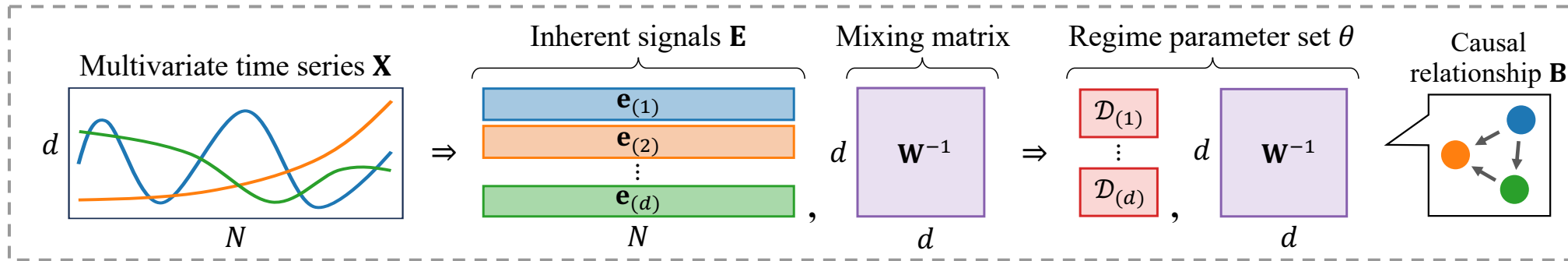
レジーム

$$\theta = \{\mathbf{W}, \mathcal{D}_{(1)}, \dots, \mathcal{D}_{(d)}\}$$



# 特徴的な時系列パターン

➤ 単一の時系列パターンは以下のとおり



$$\mathbf{s}_{(i)}(t+1) = \Lambda_{(i)} \mathbf{s}_{(i)}(t) \quad (1 \leq i \leq d)$$

$$e_{(i)}(t) = g^{-1}(\Phi_{(i)} \mathbf{s}_{(i)}(t)) \quad (1 \leq i \leq d)$$

$d$  個の固有ダイナミクス集合

$$\underline{\mathbf{v}}(t) = \underline{\mathbf{W}^{-1}} \mathbf{e}(t) \quad (\mathbf{e}(t) = \{e_{(i)}(t)\}_{i=1}^d)$$

推定値

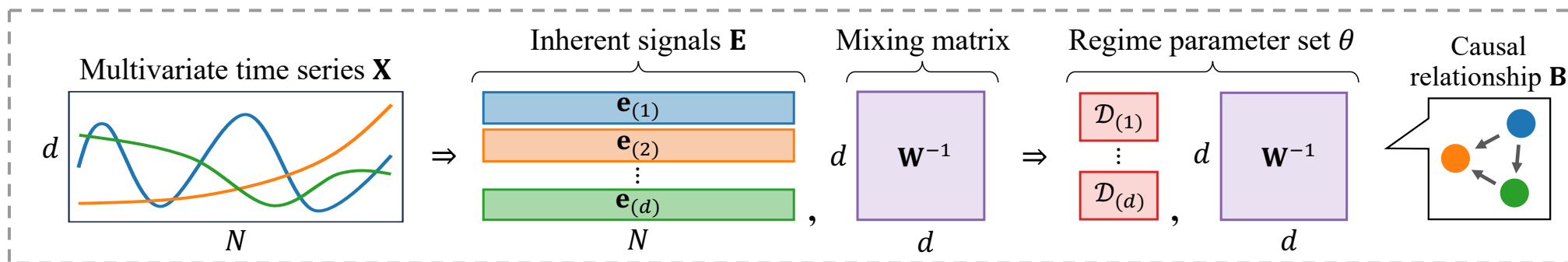
混合行列

レジーム

$$\theta = \{\mathbf{W}, \mathcal{D}_{(1)}, \dots, \mathcal{D}_{(d)}\}$$

# 特徴的な時系列パターン

➤ 単一の時系列パターンは以下のとおり



$$\mathbf{s}_{(i)}(t+1) = \Lambda_{(i)} \mathbf{s}_{(i)}(t) \quad (1 \leq i \leq d)$$

$$e_{(i)}(t) = g^{-1}(\Phi_{(i)} \mathbf{s}_{(i)}(t)) \quad (1 \leq i \leq d)$$

$d$  個の固有ダイナミクス集合

$$\mathbf{v}(t) = \mathbf{W}^{-1} \mathbf{e}(t) \quad (\mathbf{e}(t) = \{e_{(i)}(t)\}_{i=1}^d)$$

推定値

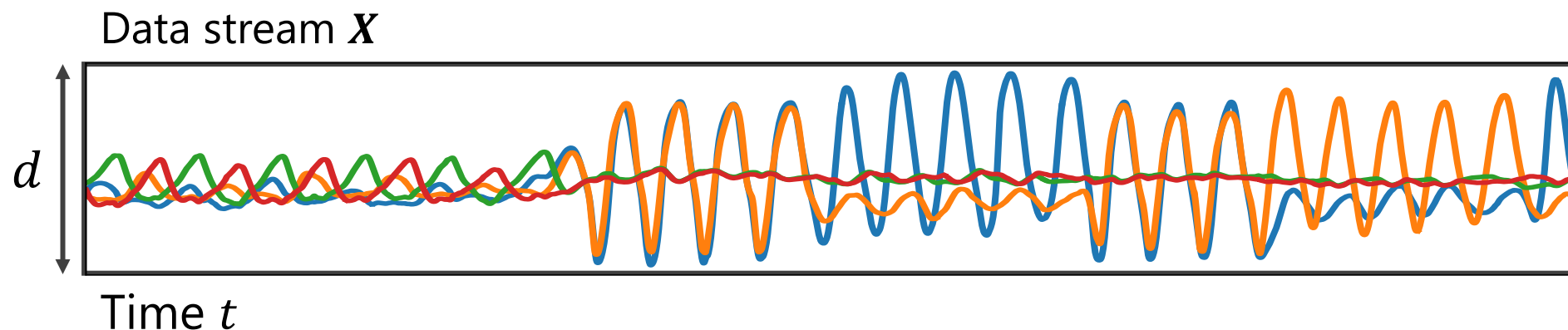
混合行列

レジーム

$$\theta = \{\mathbf{W}, \mathcal{D}_{(1)}, \dots, \mathcal{D}_{(d)}\}$$

# レジームの動的な遷移

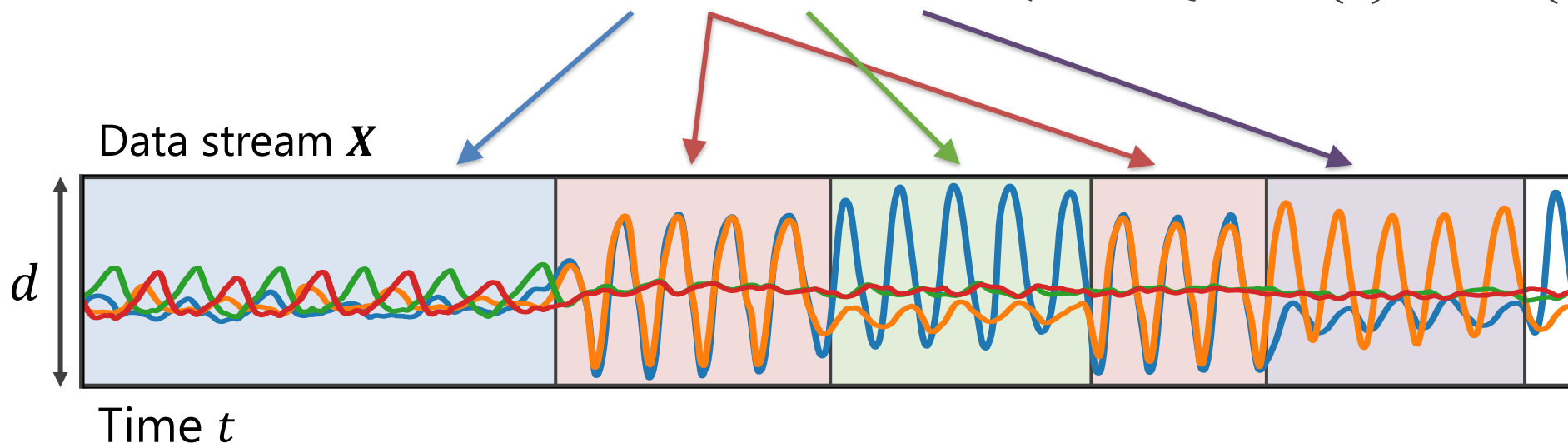
- ▶ 時系列データストリーム中の時系列パターンの動的変化
  - ❖ レジームセット  $\Theta = \{\theta^1, \theta^2, \dots, \theta^R\}$  ( $\theta^i = \{W, \mathcal{D}_{(1)}, \dots, \mathcal{D}_{(d)}\}$ )



# レジームの動的な遷移

➤ 時系列データストリーム中の時系列パターンの動的変化

❖ レジームセット  $\Theta = \{\theta^1, \theta^2, \dots, \theta^R\}$  ( $\theta^i = \{W, \mathcal{D}_{(1)}, \dots, \mathcal{D}_{(d)}\}$ )



# アルゴリズム

提案アルゴリズムは以下の4要素で構成されている

- ModeEstimator
- RegimeCreation
- ModeGenerator
- RegimeUpdater

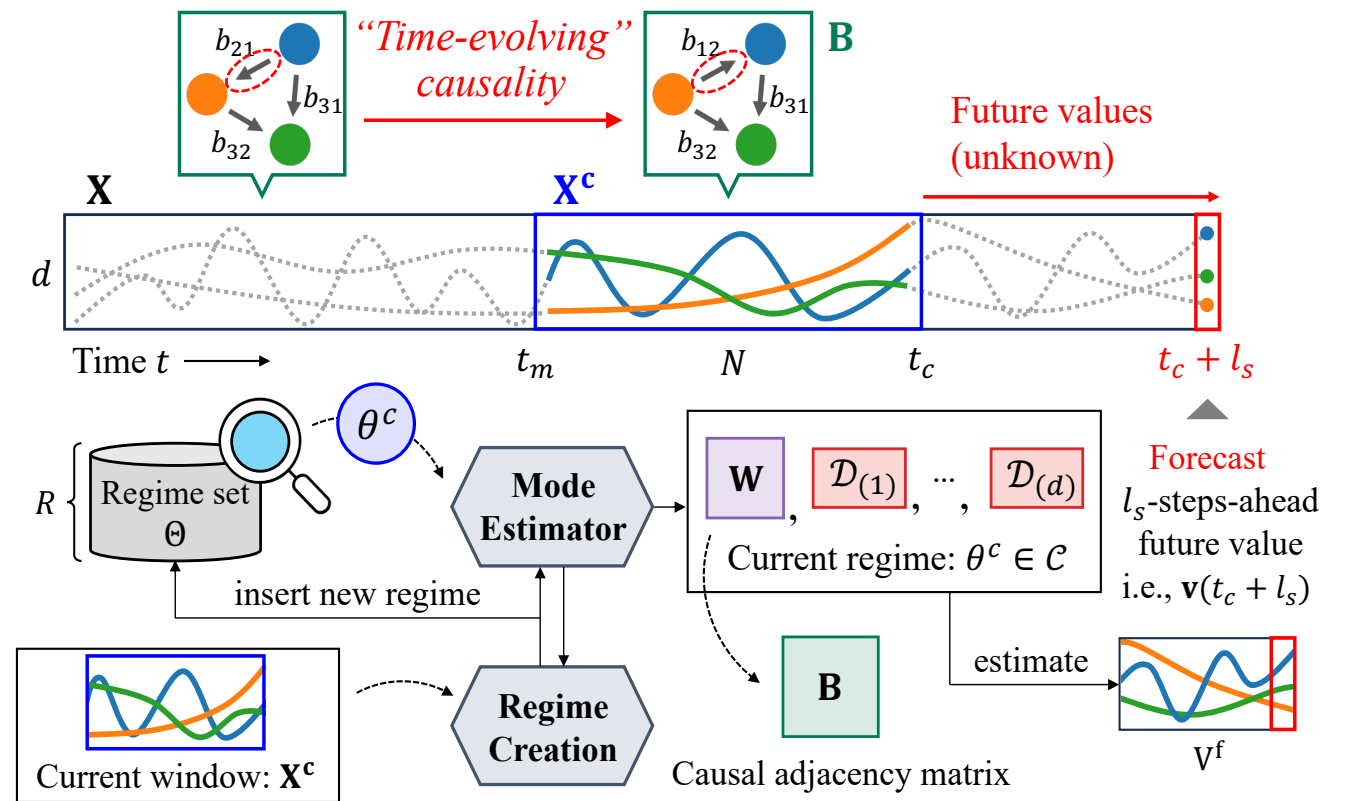
モデル更新用パラメータ:

$$\omega = \{\{P_{(i)}\}_{i=1}^d, \{\epsilon_{(i)}\}_{i=1}^d\}$$

モデルパラメータ集合:      モデル候補:

$$\mathcal{F} = \{\Theta, \Omega\}$$

$$\mathcal{C} = \{\theta^c, \omega^c, S_{en}^c\}$$



# アルゴリズム

提案アルゴリズムは以下の4要素で構成されている

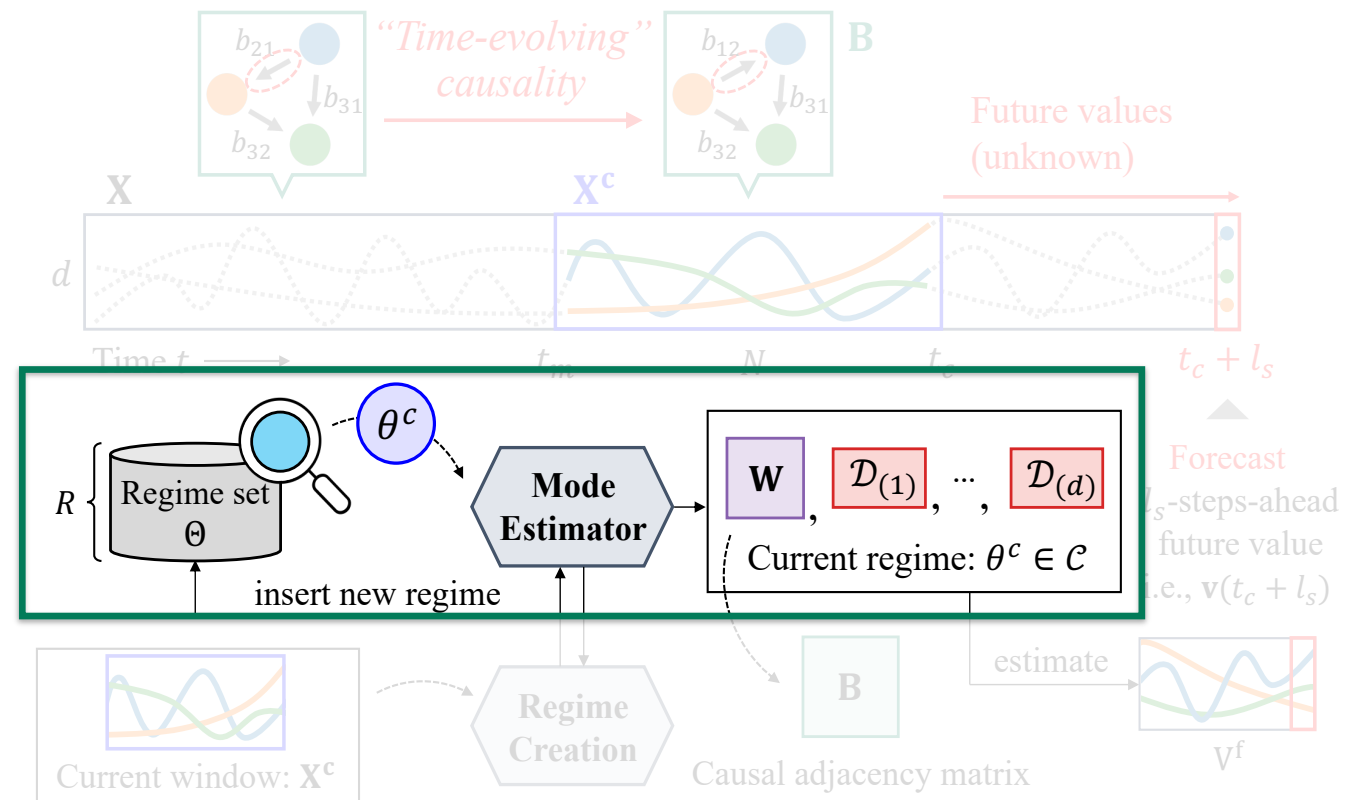
## ➤ ModeEstimator

- ❖ モデルパラメータ集合  $\mathcal{F}$  とモデル候補  $\mathcal{C}$  を逐次的に推定する

➤ RegimeCreation

➤ ModeGenerator

➤ RegimeUpdater



# アルゴリズム

提案アルゴリズムは以下の4要素で構成されている

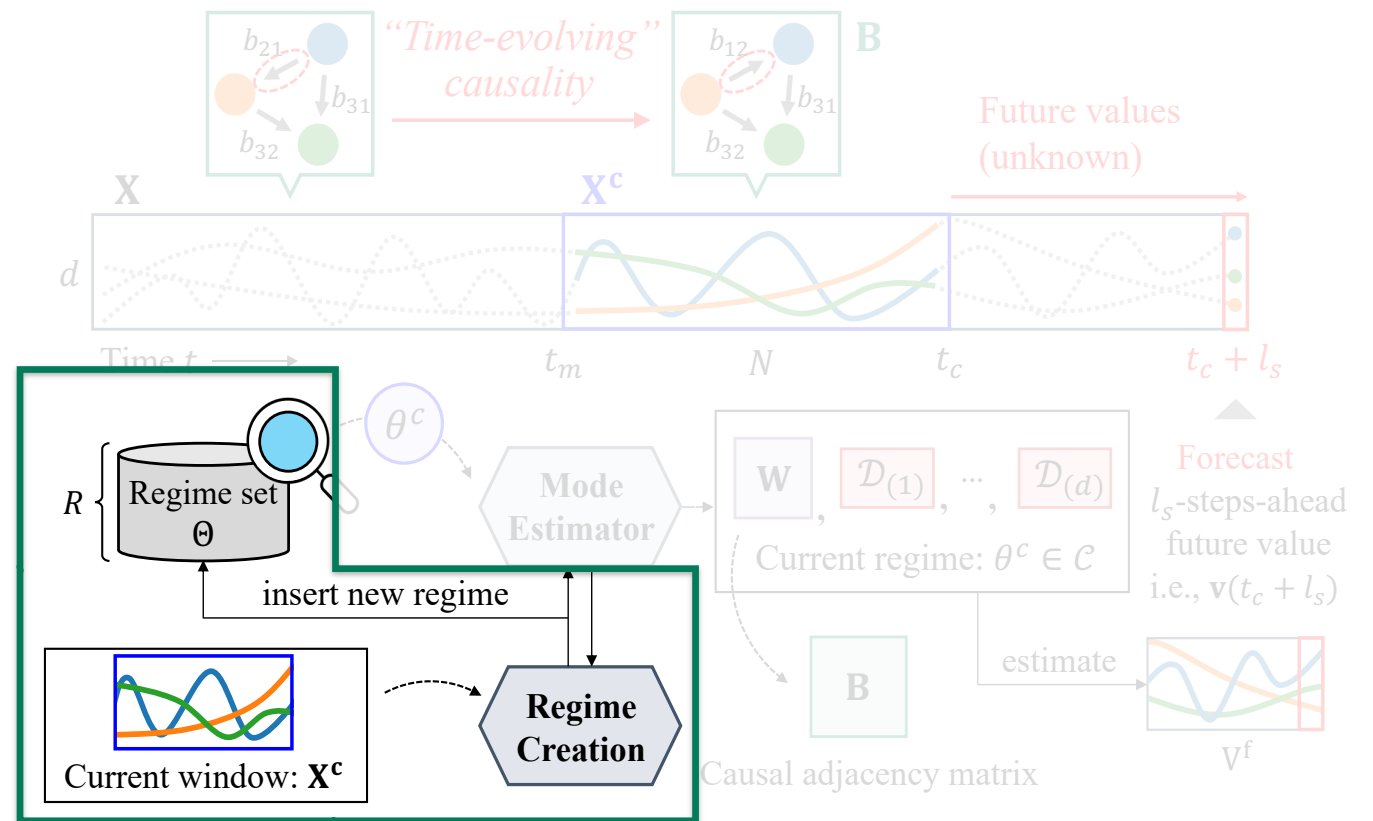
➤ ModeEstimator

➤ **RegimeCreation**

- ❖ 未知のパターンが現れた場合  
カレントウィンドウ  $X^c$  から  
レジーム  $\theta^c$  を推定し  
レジームセット  $\Theta$  に追加する

➤ ModeGenerator

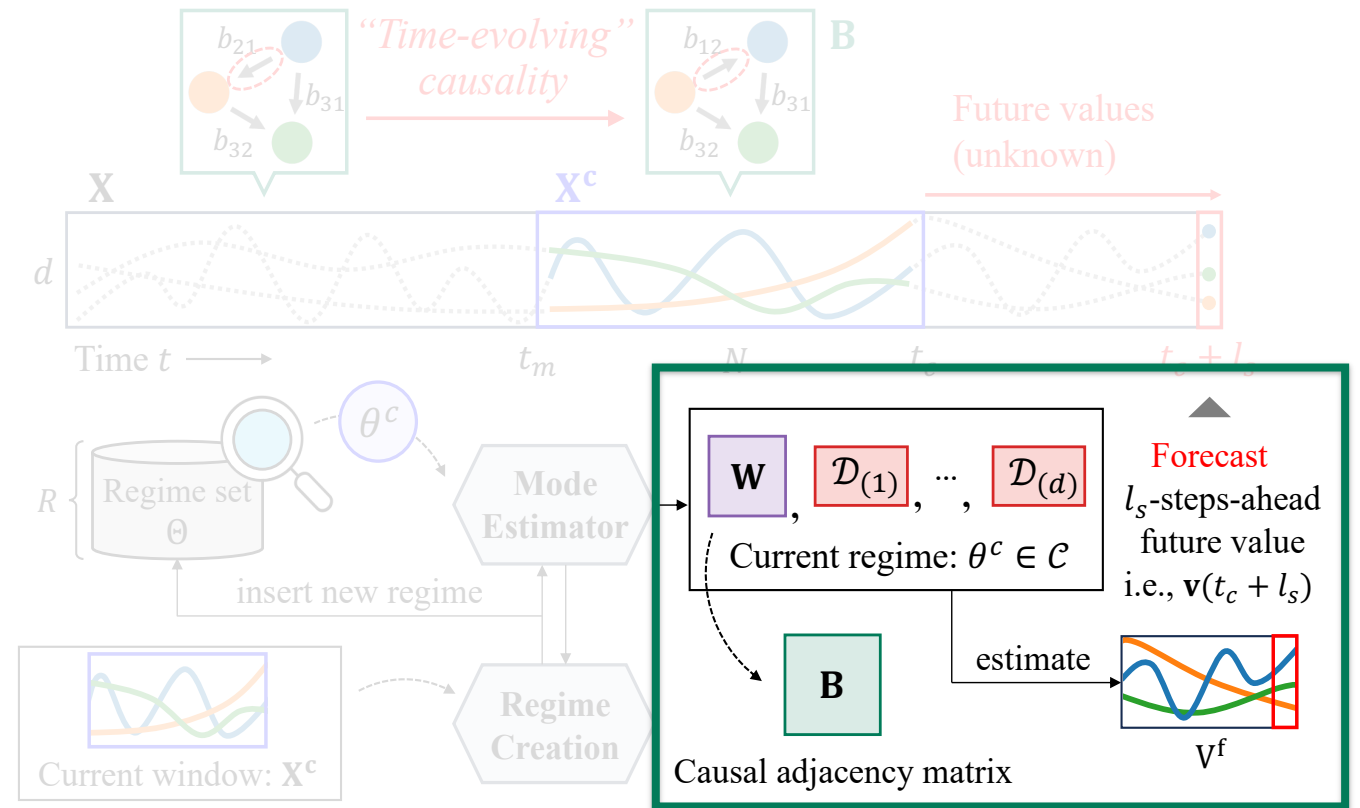
➤ RegimeUpdater



# アルゴリズム

提案アルゴリズムは以下の4要素で構成されている

- ModeEstimator
- RegimeCreation
- **ModeGenerator**
  - ❖ モデル候補  $\mathcal{C}$  から因果隣接行列  $B$  を生成し、 $l_s$  ステップ先の値を予測する
- RegimeUpdater





# アルゴリズム

提案アルゴリズムは以下の4要素で構成されている

- ModeEstimator
- RegimeCreation
- ModeGenerator
- **RegimeUpdater**
- 分離行列  $W$  の更新
  - ❖ 適応フィルタに基づいた更新
  - ❖ 計算時間 / メモリの観点から効率的
- 固有ダイナミクス集合  $\mathcal{D}_{(i)}$  の更新

- ❖ 最新の値  $x(t_c)$  を用いて  
レジーム  $\theta^c = \{W, \mathcal{D}_{(1)}, \dots, \mathcal{D}_{(d)}\}$   
を更新する

Details in paper

$$\mathbf{A}_{(i)}^{new} = \mathbf{A}_{(i)}^{prev} + (g(e_{(i)}(t_c)) - \mathbf{A}_{(i)}^{prev} g(e_{(i)}(t_c - 1)))\gamma_{(i)}$$

$$\gamma_{(i)} = \frac{g(e_{(i)}(t_c - 1))^{\top} \mathbf{P}_{(i)}^{prev}}{\mu + g(e_{(i)}(t_c - 1))^{\top} \mathbf{P}_{(i)}^{prev} g(e_{(i)}(t_c - 1))}$$

$$\mathbf{P}_{(i)}^{new} = \frac{1}{\mu} (\mathbf{P}_{(i)}^{prev} - \mathbf{P}_{(i)}^{prev} g(e_{(i)}(t_c - 1))\gamma_{(i)})$$

# 理論的分析

ModePlait は以下の優れた特性を有する

- **定理 2.** ModePlait における因果探索は，ModeGenerator での因果隣接行列  $B$  の抽出と同値である。
  - ❖ 提案手法は真の因果関係を推定可能である
- **定理 3.** (定理 1 に基づくと) 各プロセスにおける ModePlait の計算時間量は少なくとも  $O(N \sum_i k_i + dh^2)$  であり，高々  $O(RN \sum_i k_i + N(d^2 + h^2) + k^2)$  である。
  - ❖ 計算時間が時系列データストリームの長さに依存しない
  - ❖ 半無限長の長さを持つデータに対して実用的である

Details in paper

# 題目

- 研究背景
- 提案手法
- **評価実験**
- まとめ

# 検証項目

以下の全ての疑問に対して回答する

## ➤ Q1. 有効性

時間変化する因果関係の抽出に対する提案手法の有効性

## ➤ Q2. 正確性

各課題に対する提案手法の精度の検証

## ➤ Q3. 計算時間

時系列データストリームの予測に対する計算時間の検証

# 実験設定

## ➤ 9つのデータセット

### □ 人工データ

- ❖ 因果探索の定量的な評価に使用
- ❖ 5種類の異なるパターン

### □ 実データ

- ❖ 幅広い分野のデータを使用
  - COVID-19の感染者数データ
  - Web検索数データ
  - モーションキャプチャによるセンサーデータ

## ➤ 12種類の比較手法

- ❖ CASPER
- ❖ DARING
- ❖ NoCurl
- ❖ NO-MLP
- ❖ NOTEARS
- ❖ LiNGAM
- ❖ GES
- ❖ TimesNet
- ❖ PatchTST
- ❖ DeepAR
- ❖ OrbitMap
- ❖ ARIMA

因果探索のための  
7つのモデル

時系列予測のための  
5つのモデル

# Q1. 有効性

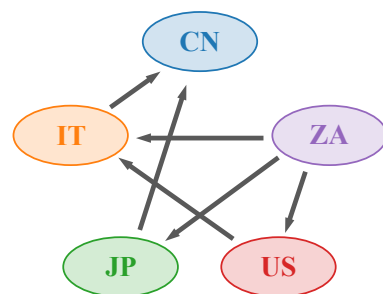
## ➤ 疫病データストリームに対する出力例

- ❖ 主要5カ国のCOVID-19の感染者数によって構成されたデータセット

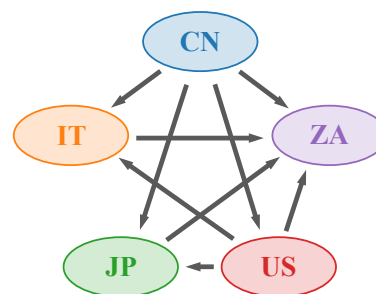


各矢印の基部は原因を  
先端は結果に対応

南アフリカ保健省が最新の  
変異株 501.V2 の発見を発表



(a-i) January 8, 2021

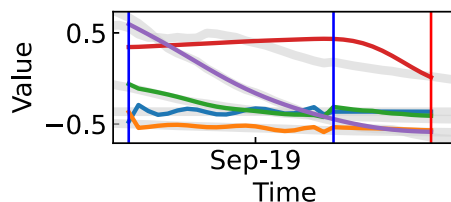


(a-ii) May 19, 2022

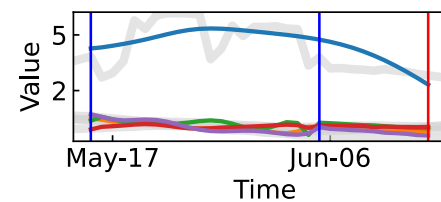
(a) 異なる時刻における因果関係

上海における過去最大の  
ロックダウン

時系列パターンの遷移を  
考慮した高精度な将来予  
測を達成



(b-i) September 27, 2021



(b-ii) June 5, 2022

(b) 10日先の予測のスナップショット

# Q1. 有効性

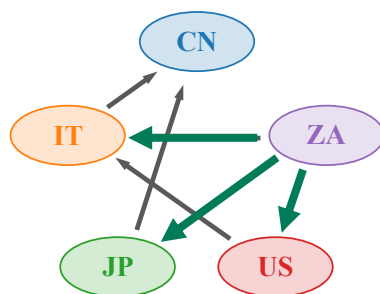
## ➤ 疫病データストリームに対する出力例

- ❖ 主要5カ国のCOVID-19の感染者数によって構成されたデータセット

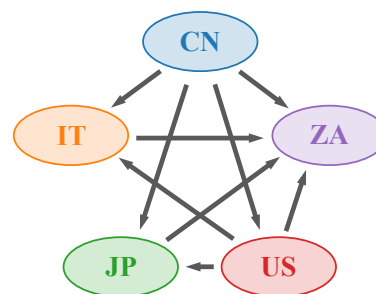


各矢印の基部は原因を  
先端は結果に対応

南アフリカ保健省が最新の  
変異株501.V2の発見を発表



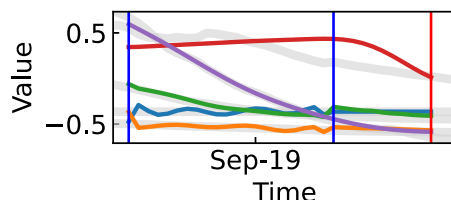
(a-i) January 8, 2021



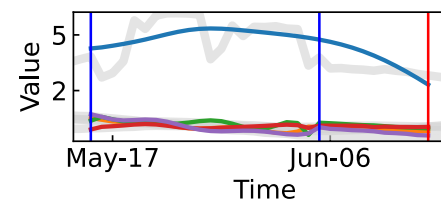
(a-ii) May 19, 2022

(a) 異なる時刻における因果関係

上海における過去最大の  
ロックダウン



(b-i) September 27, 2021



(b-ii) June 5, 2022

(b) 10日先の予測のスナップショット

時系列パターンの遷移を  
考慮した高精度な将来予  
測を達成

# Q1. 有効性

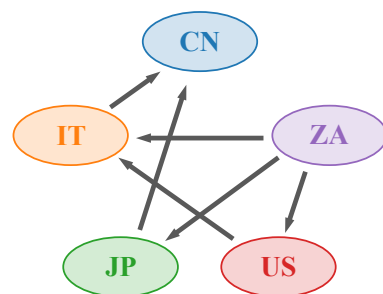
## ➤ 疫病データストリームに対する出力例

- ❖ 主要5カ国のCOVID-19の感染者数によって構成されたデータセット

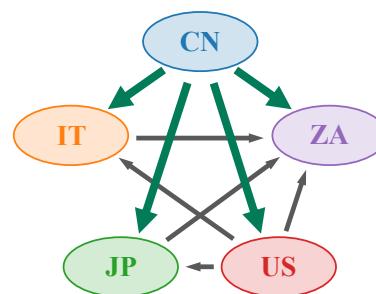


各矢印の基部は原因を  
先端は結果に対応

南アフリカ保健省が最新の  
変異株 501.V2 の発見を発表



(a-i) January 8, 2021

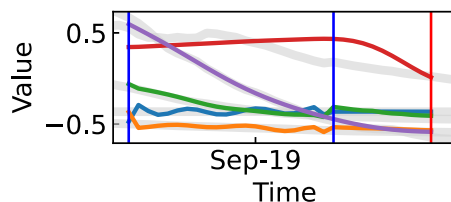


(a-ii) May 19, 2022

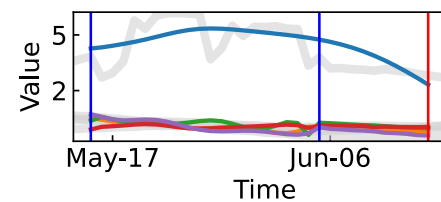
(a) 異なる時刻における因果関係

上海における過去最大の  
ロックダウン

時系列パターンの遷移を  
考慮した高精度な将来予  
測を達成



(b-i) September 27, 2021



(b-ii) June 5, 2022

(b) 10日先の予測のスナップショット



# Q1. 有効性

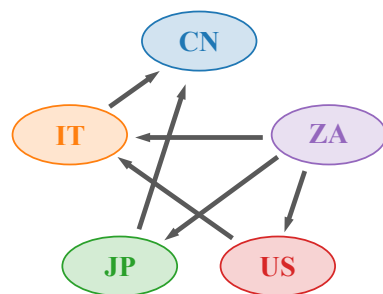
## ➤ 疫病データストリームに対する出力例

- ❖ 主要5カ国のCOVID-19の感染者数によって構成されたデータセット

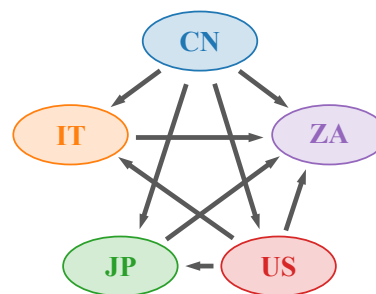


各矢印の基部は原因を  
先端は結果に対応

南アフリカ保健省が最新の  
変異株 501.V2 の発見を発表



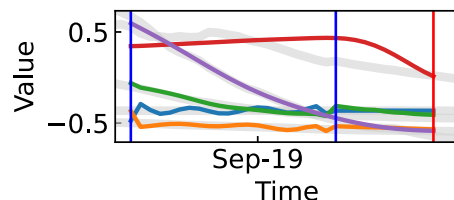
(a-i) January 8, 2021



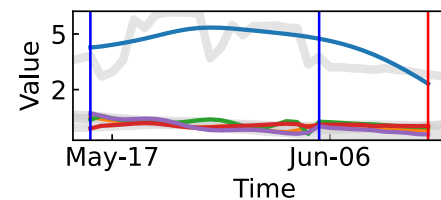
(a-ii) May 19, 2022

(a) 異なる時刻における因果関係

上海における過去最大の  
ロックダウン



(b-i) September 27, 2021



(b-ii) June 5, 2022

(b) 10日先の予測のスナップショット

時系列パターンの遷移を  
考慮した高精度な将来予  
測を達成

## Q2. 正確性：因果探索

- 提案手法の因果探索に関する正確性
  - ❖ 評価指標：SHD, SID (低いほど良)

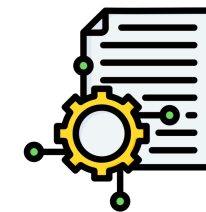


Table 3: Causal discovering results with multiple temporal sequences to encompass various types of real-world scenarios.






Models	MODEPLAIT		CASPER		DARING		NoCurl		NO-MLP		NOTEARS		LiNGAM		GES	
Metrics	SHD	SID	SHD	SID	SHD	SID	SHD	SID	SHD	SID	SHD	SID	SHD	SID	SHD	SID
1, 2, 1	<b>3.82</b>	<b>4.94</b>	5.58	<u>7.25</u>	5.75	8.58	6.31	9.90	6.36	8.74	<u>5.03</u>	9.95	7.13	8.23	7.49	11.7
1, 2, 3	<b>4.48</b>	<b>6.51</b>	5.97	8.44	5.81	9.17	6.13	9.51	6.44	8.77	<u>5.69</u>	9.56	6.79	<u>7.33</u>	7.03	10.1
1, 2, 2, 1	<b>4.32</b>	<b>5.88</b>	5.41	<u>8.41</u>	6.54	9.17	6.69	10.0	6.55	8.72	<u>5.23</u>	9.54	7.12	8.65	7.08	9.77
1, 2, 3, 4	<b>4.21</b>	<b>5.76</b>	6.22	<u>8.33</u>	6.12	9.58	6.10	9.61	6.62	8.87	<u>5.73</u>	10.1	7.10	8.50	7.29	11.3
1, 2, 3, 2, 1	<b>4.50</b>	<b>6.11</b>	6.02	8.28	<u>5.45</u>	<u>7.77</u>	6.20	9.83	6.56	8.83	<u>5.57</u>	9.11	7.46	8.05	7.74	12.1

# Q2. 正確性：時系列予測

## ➤ 提案手法の時系列予測に関する正確性

### ❖ 評価指標：RMSE, MAE (低いほど良)

Table 4: Multivariate forecasting results for both synthetic and real-world datasets. We used forecasting steps  $l_s \in \{5, 10, 15\}$ .

Models		MODEPLAIT		TimesNet		PatchTST		DeepAR		OrbitMap		ARIMA		
Metrics		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	
	#0 synthetic	5	<b>0.722</b>	<b>0.528</b>	0.805	0.578	<u>0.768</u>	0.581	1.043	0.821	0.826	<u>0.567</u>	0.962	0.748
		10	<b>0.829</b>	<b>0.607</b>	<u>0.862</u>	0.655	0.898	0.649	1.073	0.849	0.896	<u>0.646</u>	0.966	0.752
		15	<b>0.923</b>	<b>0.686</b>	<u>0.940</u>	<u>0.699</u>	0.973	0.706	1.137	0.854	0.966	0.710	0.982	0.765
	#1 covid19	5	<b>0.588</b>	<b>0.268</b>	0.659	0.314	<u>0.640</u>	<u>0.299</u>	1.241	0.691	1.117	0.646	1.259	0.675
		10	<b>0.740</b>	<b>0.361</b>	<u>0.841</u>	<u>0.410</u>	1.053	0.523	1.255	0.693	1.353	0.784	1.260	0.687
		15	<b>0.932</b>	<b>0.461</b>	<u>1.026</u>	<u>0.516</u>	1.309	0.686	1.265	0.690	1.351	0.792	1.277	0.718
	#2 web-search	5	<b>0.573</b>	<b>0.442</b>	<u>0.626</u>	<u>0.469</u>	0.719	0.551	1.255	1.024	0.919	0.640	1.038	0.981
		10	<b>0.620</b>	<b>0.481</b>	<u>0.697</u>	<u>0.514</u>	0.789	0.604	1.273	1.044	0.960	0.717	1.247	1.037
		15	<b>0.646</b>	<b>0.505</b>	<u>0.701</u>	<u>0.527</u>	0.742	0.571	1.300	1.069	0.828	0.631	1.038	0.795
	#3 chicken-dance	5	<b>0.353</b>	<b>0.221</b>	0.759	0.490	<u>0.492</u>	<u>0.303</u>	0.890	0.767	0.508	0.316	2.037	1.742
		10	<b>0.511</b>	<b>0.325</b>	0.843	0.564	0.838	0.535	0.886	0.753	<u>0.730</u>	<u>0.476</u>	1.863	1.530
		15	<b>0.653</b>	<b>0.419</b>	0.883	0.592	0.972	0.654	<u>0.862</u>	0.718	0.903	<u>0.565</u>	1.792	1.481
	#4 exercise	5	<b>0.309</b>	<b>0.177</b>	0.471	<u>0.275</u>	0.465	0.304	<u>0.408</u>	0.290	0.424	<u>0.275</u>	1.003	0.748
		10	<b>0.501</b>	<b>0.309</b>	0.630	0.381	0.789	0.518	<u>0.509</u>	0.382	0.616	<u>0.377</u>	1.104	0.814
		15	<u>0.687</u>	<b>0.433</b>	0.786	0.505	1.147	0.758	<b>0.676</b>	0.475	0.691	<u>0.434</u>	1.126	0.901

# Q2. 正確性：アブレーション研究

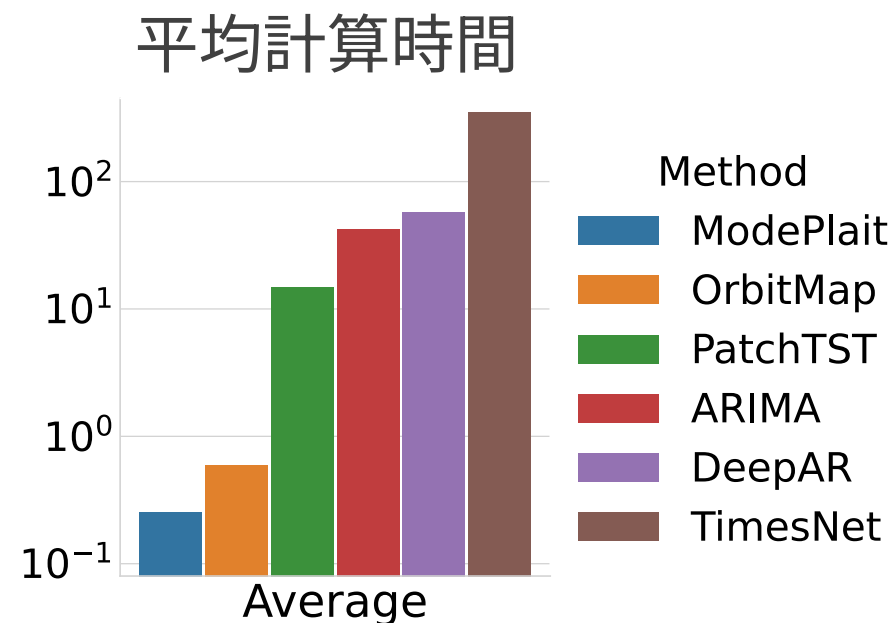
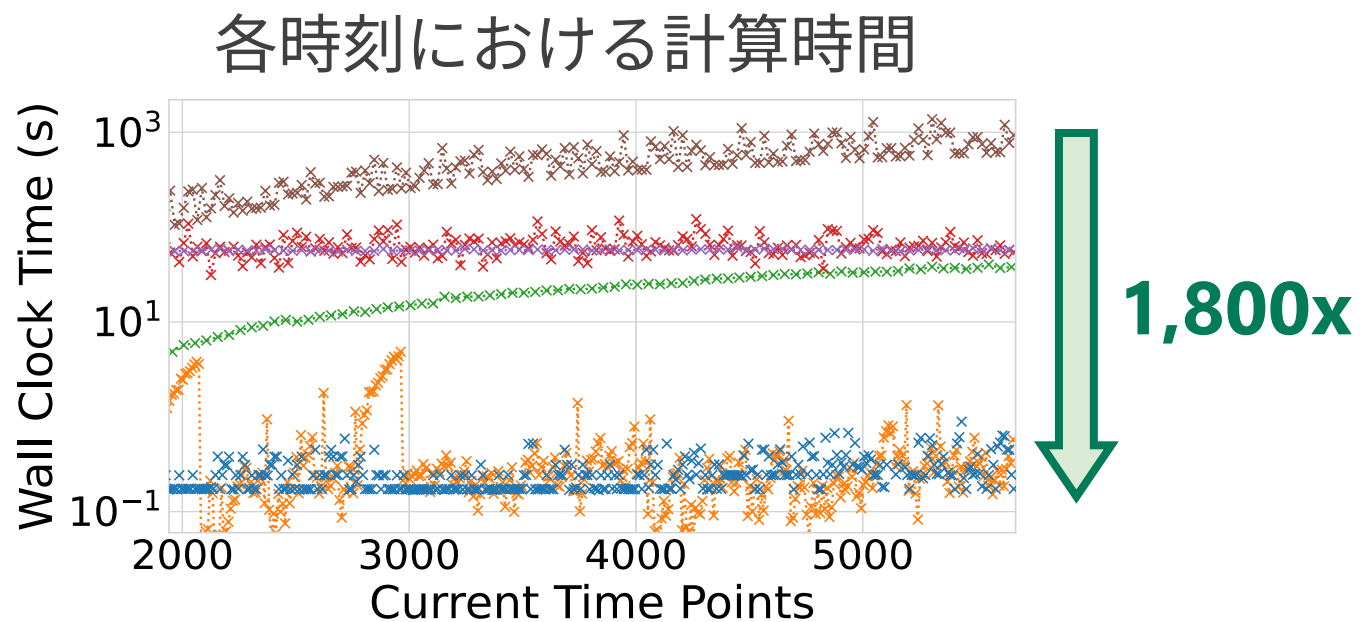
- 提案手法における因果探索の機能が予測性能に与える影響を検証
  - ❖ 評価指標：RMSE, MAE（低いほど良）

Table 5: Ablation study results with forecasting steps  $l_s \in \{5, 10, 15\}$  for both synthetic and real-world datasets.

Datasets		#0 synthetic		#1 covid19		#2 web-search		#3 chicken-dance		#4 exercise	
Metrics		RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
MODEPLAIT (full)	5	0.722	0.528	0.588	0.268	0.573	0.442	0.353	0.221	0.309	0.177
	10	0.829	0.607	0.740	0.361	0.620	0.481	0.511	0.325	0.501	0.309
	15	0.923	0.686	0.932	0.461	0.646	0.505	0.653	0.419	0.687	0.433
w/o causality	5	0.759	0.563	0.758	0.374	0.575	0.437	0.391	0.262	0.375	0.218
	10	0.925	0.696	0.848	0.466	0.666	0.511	0.590	0.398	0.707	0.433
	15	1.001	0.760	1.144	0.583	0.708	0.545	0.821	0.537	0.856	0.533



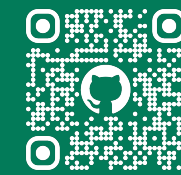
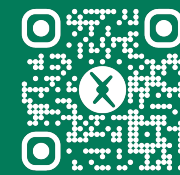
# Q3. 計算時間



データストリーム全体の長さに対して**定数時間**で計算可能

# 題目

- 研究背景
- 提案手法
- 評価実験
- **まとめ**



ModePlait は以下の優れた特性を全て満たす

## ➤ Effective

- ❖ 時系列パターンの遷移に基づいて時間変化する因果関係を抽出可能

## ➤ Accurate

- ❖ 理論的に因果関係を抽出し，正確に将来値を予測する
- ❖ 最新の比較手法を上回る精度を達成

## ➤ Scalable

- ❖ 計算コストが時系列データストリームの長さに依存しない

# Appendix



# 関連研究

- 以下の5つの観点において ModePlait は優れている

	ARIMA/++	TICC	NOTEARS/++	OrbitMap	TimesNet	MODEPLAIT
Stream Processing	-	-	-	✓	-	✓
Forecasting	✓	-	-	✓	✓	✓
Data Compression	-	✓	-	✓	-	✓
Interdependency	-	✓	✓	-	-	✓
Time-evolving Causality	-	-	-	-	-	✓

# 関連研究

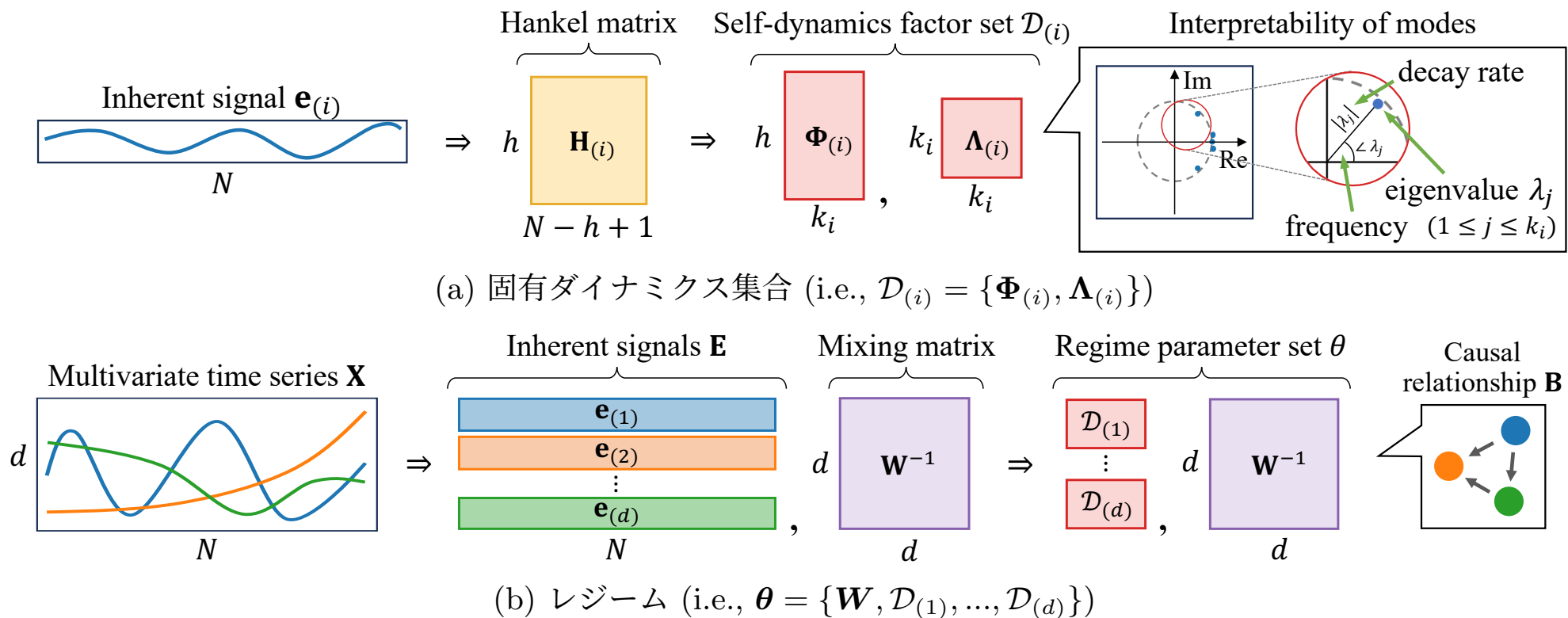
- ARIMA [Box and Jenkins 1976]
  - ❖ 古典的な将来予測手法
  - ❖ 時系列間の非線形な関係性を考慮できない 😓
- OrbitMap [Matsubara and Sakurai 2019]
  - ❖ ストリーム方式に対応した将来予測を実現
  - ❖ 時間変化する因果関係に非対応 😓

# 関連研究

- 一般的な因果探索手法
  - ❖ CASPER [Liu et al. 2023] 等
  - ❖ 時系列データストリームに非対応 😓
- 深層学習ベースの将来予測手法
  - ❖ TimesNet [Wu et al. 2023] 等
  - ❖ モデルの学習に膨大な時間を要するため、最新の情報を考慮した高速かつ連続的な将来予測が困難 😓

# 提案手法：モデル概要図

➤ 提案モデルの概要図は以下のとおり



# 評価実験：評価指標

因果探索の実験では SHD, SID の 2 種類の評価指標を採用した

- structural Hamming distance (SHD)
  - ❖ 因果隣接行列の差異を定量化する指標
  - ❖ 欠落した辺，余分な辺，逆転した辺の数
- structural intervention distance (SID)
  - ❖ 因果探索の精度評価に特に適した評価指標
  - ❖ 推定された因果隣接行列を使用した場合に，介入分布  $p(x_j | \text{do}(X_i = \bar{x}))$  が誤って検出される  $(i, j)$  の組の数

# 評価実験：評価指標

時系列予測の実験では RMSE, MAE の 2 種類の評価指標を採用した

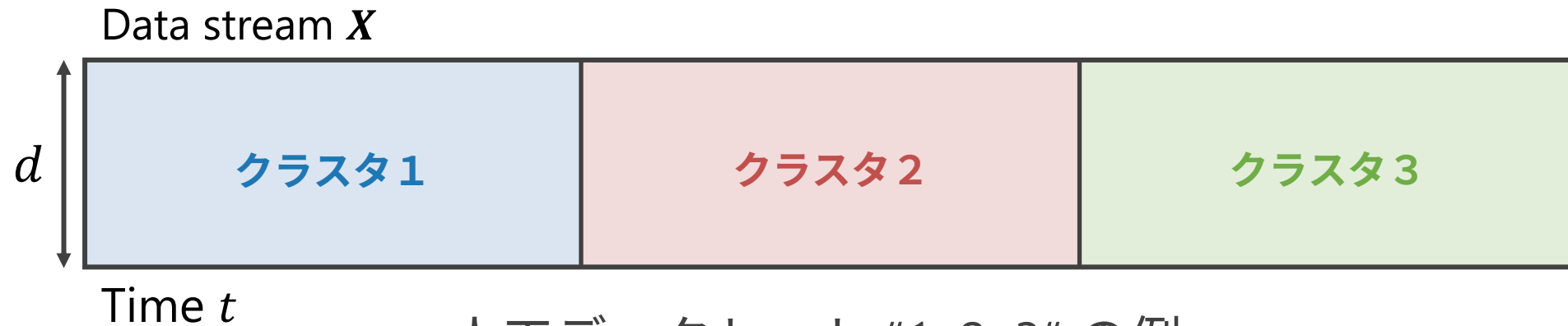
- 二乗平均誤差 (RMSE) ... 外れ値の多さに焦点が当たる

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}$$

- 平均絶対誤差 (MAE) ... 全体的な誤差の大きさを算出

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i|$$

# 評価実験：人工データセット



人工データセット "1, 2, 3" の例

- 人工データセットには複数のクラスタが存在する
  - ❖ 1つのクラスタに1つの因果関係が対応
  - ❖ 因果隣接行列  $B$  は Eröds-Rényi (ER) モデルに基づいて生成
  - ❖ エッジ密度  $p = 0.5$ , 観測変数の数  $d = 5$

実際に存在するエッジの数の割合