

# データ同化入門

データ解析とシミュレーションの融合

@termoshtt

PyData.Osaka #3 2017/1/21

@大阪イノベーションハブ

# 従来のシミュレーション

## 実験

- 風速計
- 温度計

⇒

## 理論

- N.S 方程式
- 熱力学関係式

⇒

## Simulation

- NICAM
- OpenFOAM

## Simulation の利点

- 実験では観測できない状態も見える
- 実験よりも低コスト
  - 最適化等も可能

# データ同化の必要性

## 実験

- 風速計
- 温度計

⇒

↑  
理想化

## 理論

- N.S 方程式
- 熱力学関係式

⇒

↑  
近似

## Simulation

- NICAM
- OpenFOAM

## Simulation の限界

- 理想化・近似で入った誤差が蓄積する
- 誤差が拡大する (カオス性)

# データ同化

Simulation  $\oplus$  観測データ



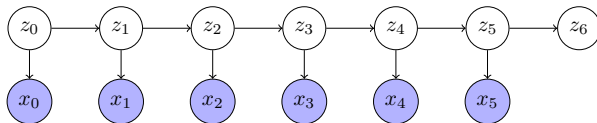
データ同化

- 観測データから状態を推定し高精度化・安定化する
- 不完全でも Simulation model がある場合に有効

→ Big Data Assimilation

# データ同化の設定

Simulation



Observation

## 逐次

- 観測  $x_t$  毎に状態  $z_t$  を推定しなおす
- 予測・制御向き

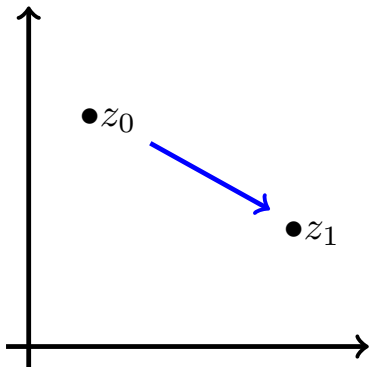
## 非逐次

- 全ての観測  $x_0, x_1, \dots$  から状態  $z_0, z_1, \dots$  を一括で推定
- 平滑化・最適化

# 確率の扱い

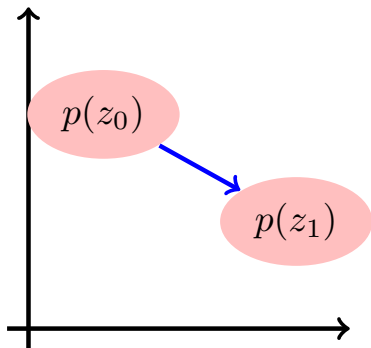
Simulation model

$$z_{t+1} = f(z_t)$$



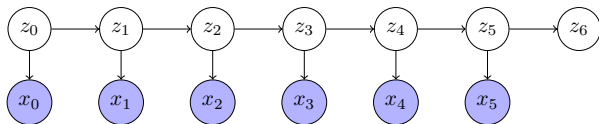
System model

$$p(z_{t+1}) = \int dz_t p(z_{t+1}|z_t)p(z_t)$$



# 逐次データ同化

Simulation



Observation

$$\begin{aligned} p(\mathbf{z}_{t+1} | \mathbf{x}_t, \mathbf{x}_{t-1}, \dots) &= \int d\mathbf{z}_t p(\mathbf{z}_{t+1}, \mathbf{z}_t | \mathbf{x}_t, \dots) \\ &= \int d\mathbf{z}_t p(\mathbf{z}_{t+1} | \mathbf{z}_t, \mathbf{x}_t, \dots) p(\mathbf{z}_t | \mathbf{x}_t, \dots) \\ &= \int d\mathbf{z}_t p(\mathbf{z}_{t+1} | \mathbf{z}_t) p(\mathbf{z}_t | \mathbf{x}_t, \dots) \\ &\propto \int d\mathbf{z}_t p(\mathbf{z}_{t+1} | \mathbf{z}_t) p(\mathbf{x}_t | \mathbf{z}_t) p(\mathbf{z}_t | \mathbf{x}_{t-1}, \dots) \end{aligned}$$