

データ同化の 基礎と活用に向けて

京都大学 産官学連携本部 菊地亮太

データ同化の可能性

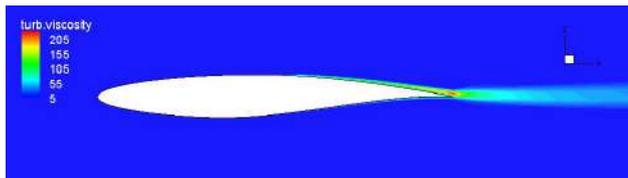
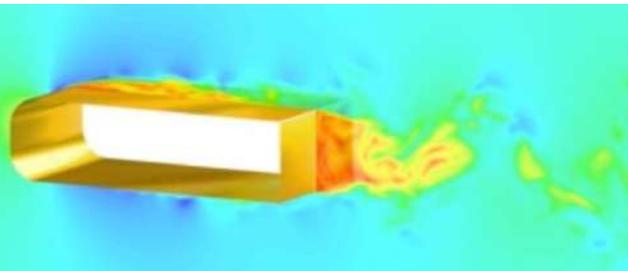
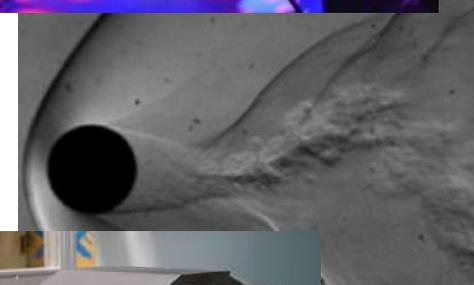
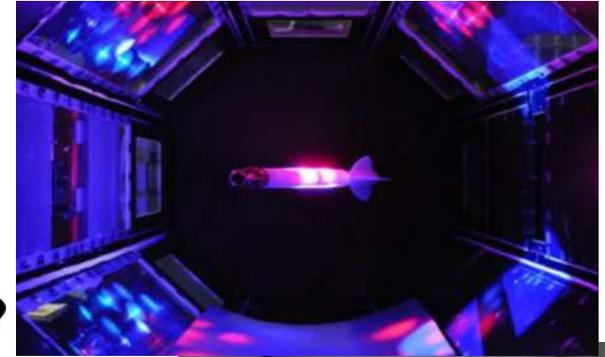
シミュレーション



データ同化



観測・計測値



出典: 東北大学 大林・焼野研究室

<http://www.ifs.tohoku.ac.jp/edge/>

データ同化はシミュレーションと観測・計測値を融合し、
両者の有効性を最大限に引き出す

本日の講義の目的と目標

物理学の研究の進め方:「実験・観測研究」と「数値解析・シミュレーション」は重要な柱



実験・観測



数値解析
シミュレーション

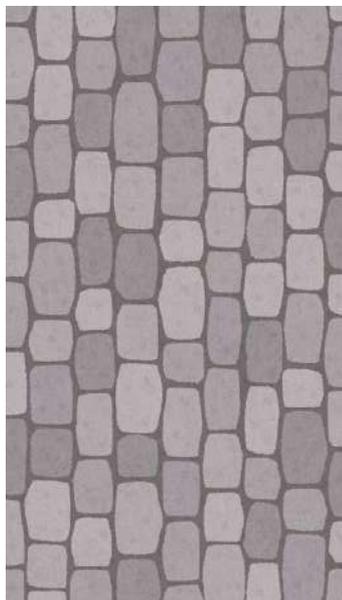
本日の講義の目的と目標

「私は実験家だから...」「私は数値解析家だから...」

⇒何となく壁があるように考えている人たちは少なくはない？



実験・観測



数値解析
シミュレーション

本日の講義の目的と目標

データ同化のコンセプト

⇒「実験とシミュレーションの融合」



実験・観測



数値解析
シミュレーション

本日の講義の目的と目標

今回の講義の目的

- データ同化の主な流れについて理解してもらう
 - 基礎的なカルマンフィルタによるデータ同化を実施する際の流れを図解して共有
- データ同化が解決しようとしている課題について理解する
 - どういう場面で使えそうかということについて理解を深めるために、データ同化を活用する主な目的について解説

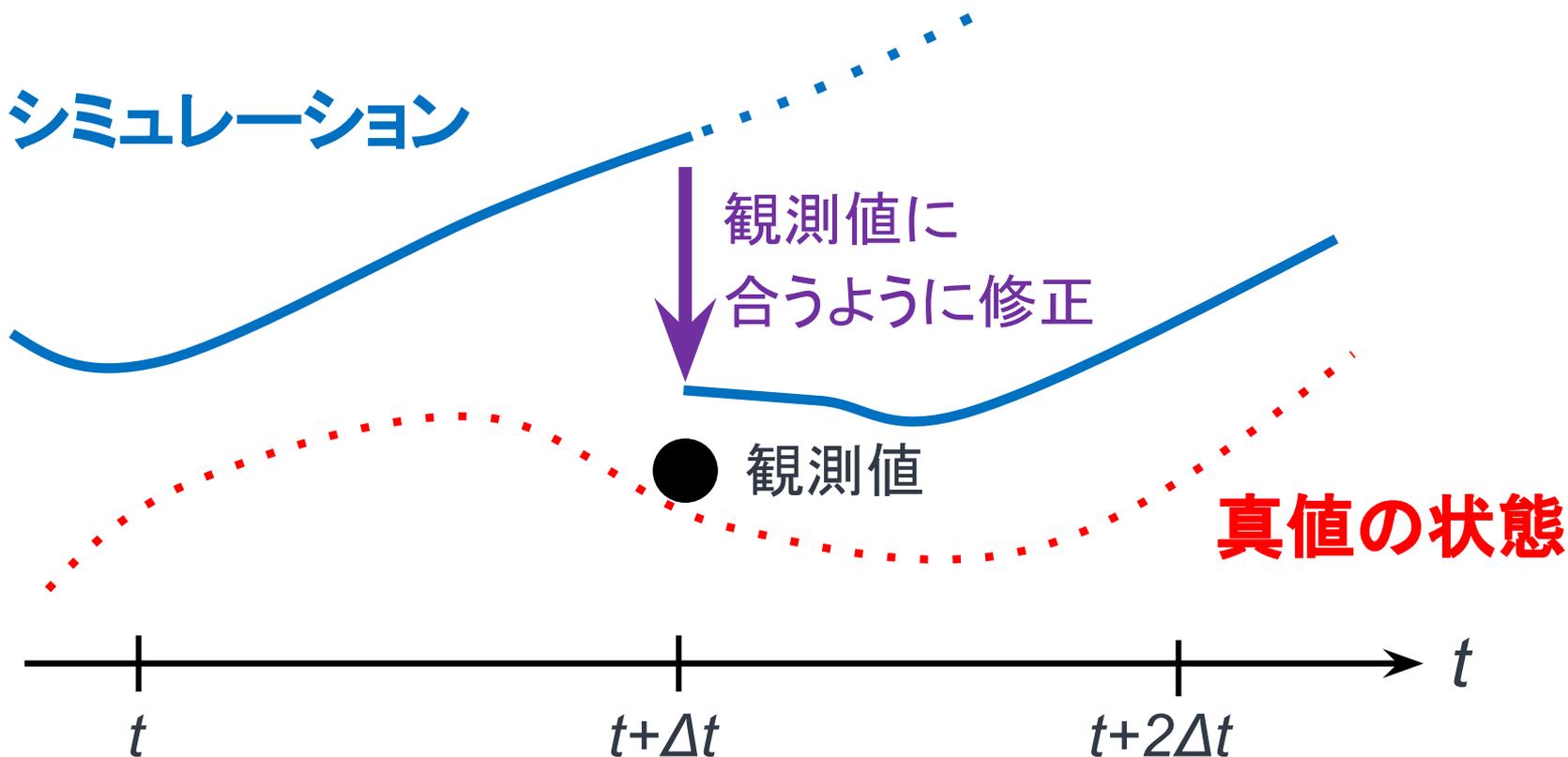
今回の講義の目標

- 自分の研究の中で使えるかも？と思ってもらう機会を提供
- 「実験家」「数値解析家」のような隔てる考え方以外もあると知ってもらう

データ同化手法の 基礎

データ同化とは？

シミュレーション



シミュレーションと観測はそれぞれ“不確実性”を持つ

⇒ 現実に近い情報を推定する

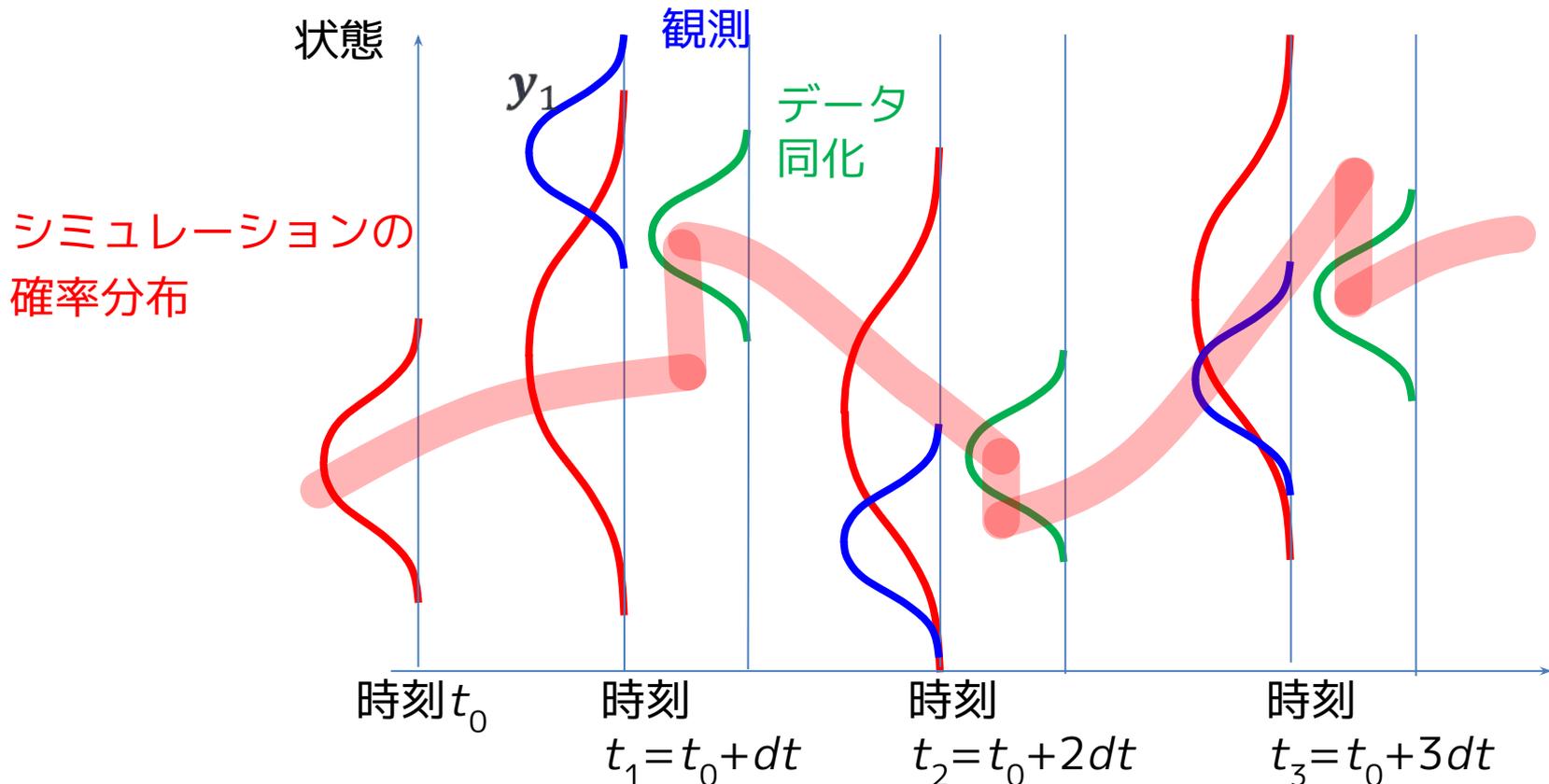
データ同化の種類

- 古典的な方法
 - 直接挿入：観測値を直接、シミュレーションの格子点に挿入
 - ナッジング：予測値と観測値の差に、係数をかけて強制項として追加
- 3次元データ同化
 - 最適内挿法：誤差標準偏差が最小になるように統計的に内挿
 - 3次元変分法：力学的なバランスを考慮しつつ、内挿
- 4次元データ同化
 - **カルマンフィルタ：正規分布・線形システム・線形観測を仮定**
 - アンサンブルカルマンフィルタ：非線形システム・非線形観測を受容
 - 粒子フィルタ：システム、観測、確率分布への仮説を必要としない
 - 4次元変分法：モデルの力学を考慮し、時間的・空間的に内挿

カルマンフィルタ

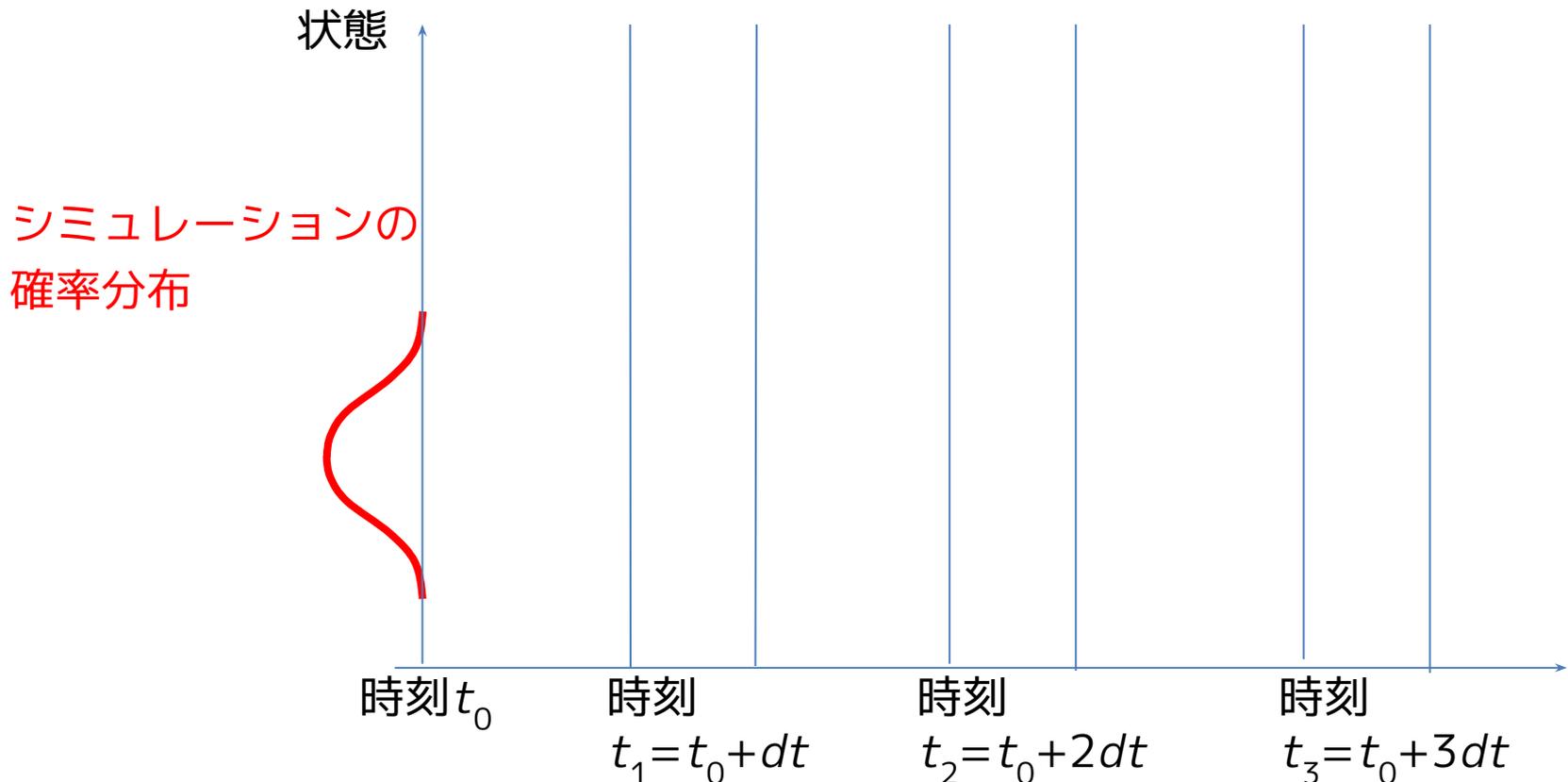
- シミュレーションの確率分布を、観測値に応じて更新することで推定値を得る

- シミュレーションの確率分布は正規分布だと仮定



カルマンフィルタ

- (1)初期条件を与える
 - シミュレーションの初期条件と誤差共分散行列の初期条件

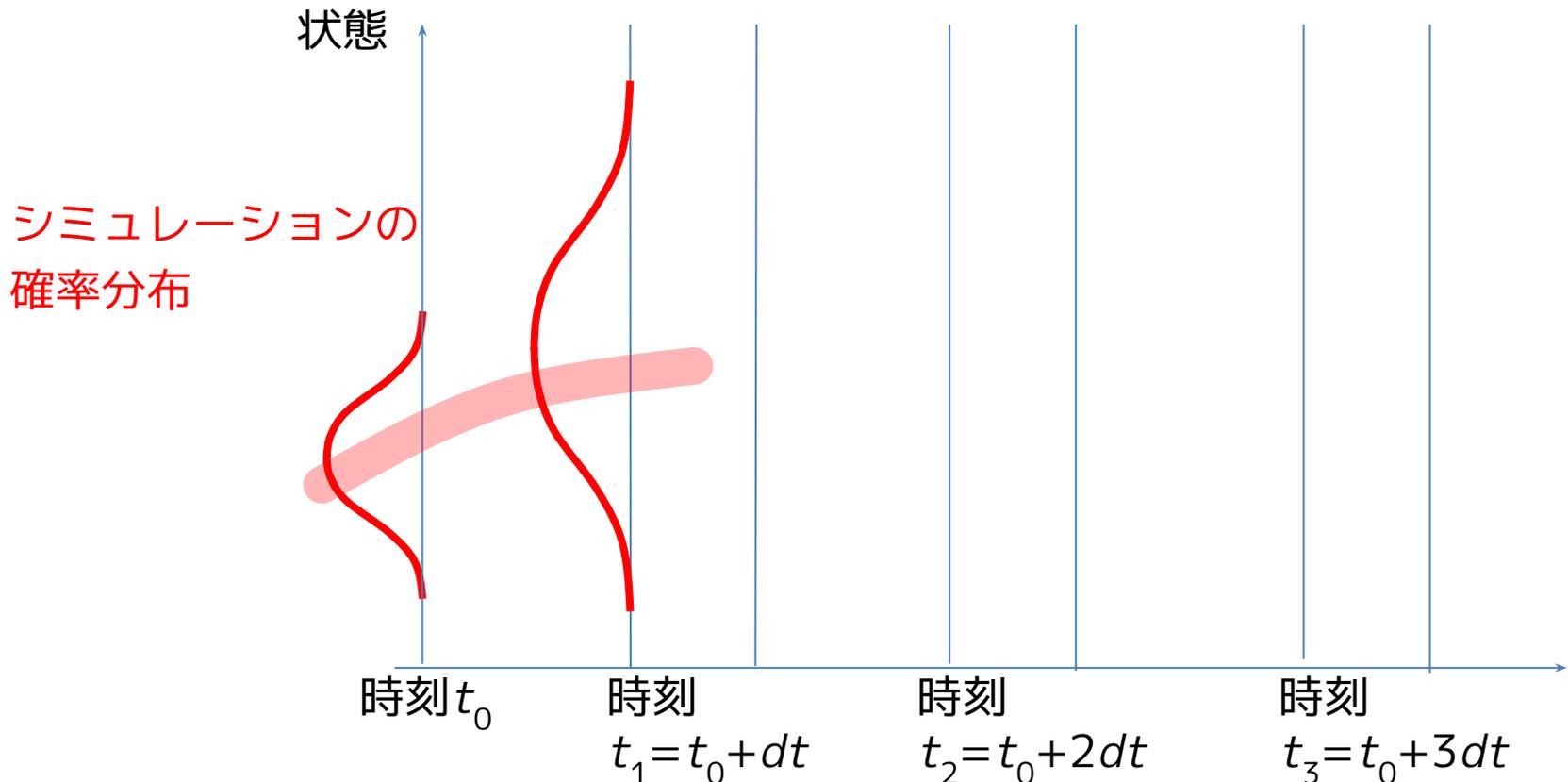


カルマンフィルタ

- (2) 予測計算を進める
 - 平均値と誤差共分散行列の予測計算を実施

$$\mathbf{x}_{t|t-1} = F_t \mathbf{x}_{t-1|t-1} \cdots (1)$$

$$V_{t|t-1} = F_t V_{t-1|t-1} F_t^T \cdots (2)$$

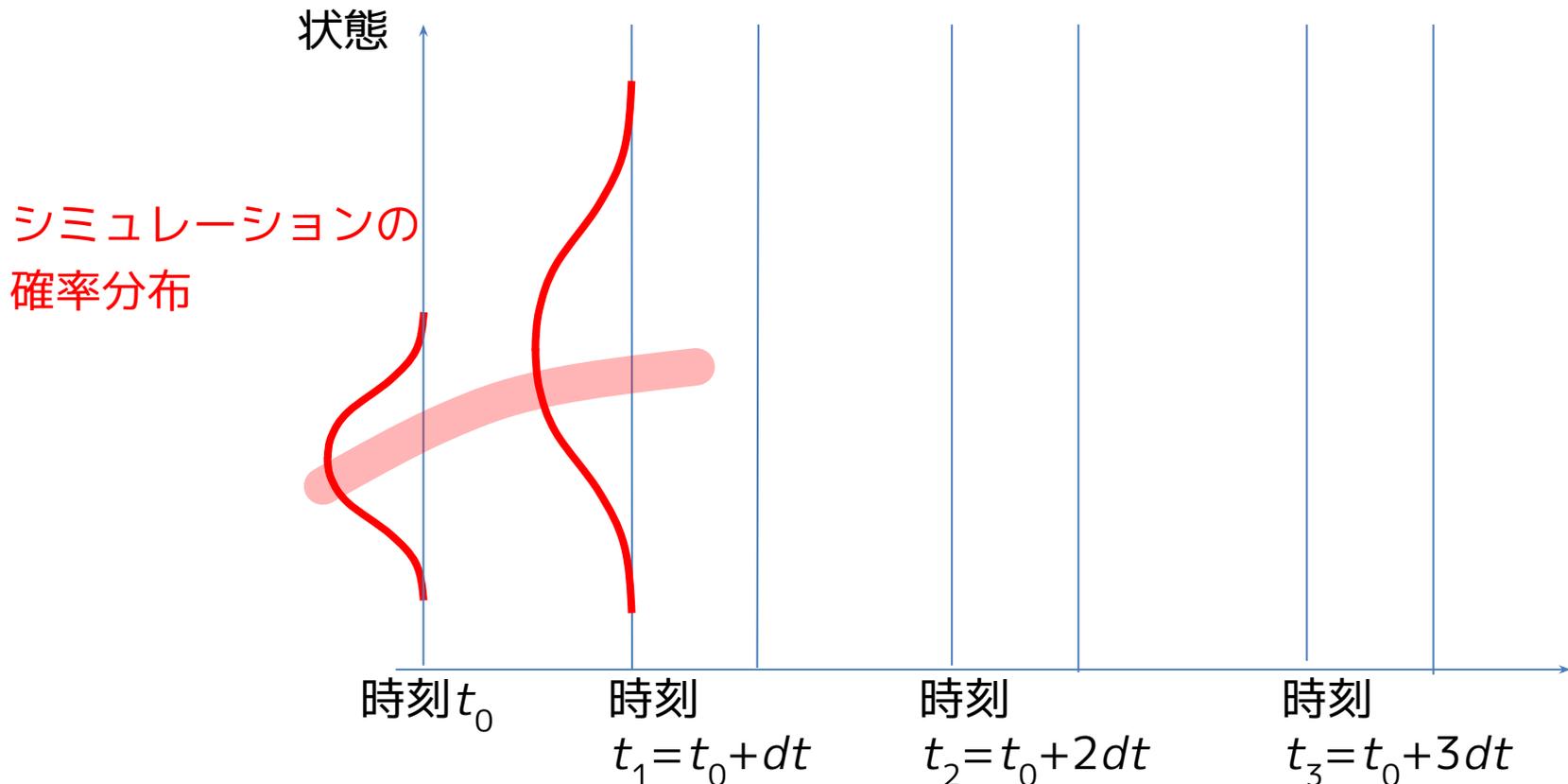


カルマンフィルタ

$$K_t = V_{t|t-1} H_t^T (H_t V_{t|t-1} H_t^T + R_t)^{-1} \dots (3)$$

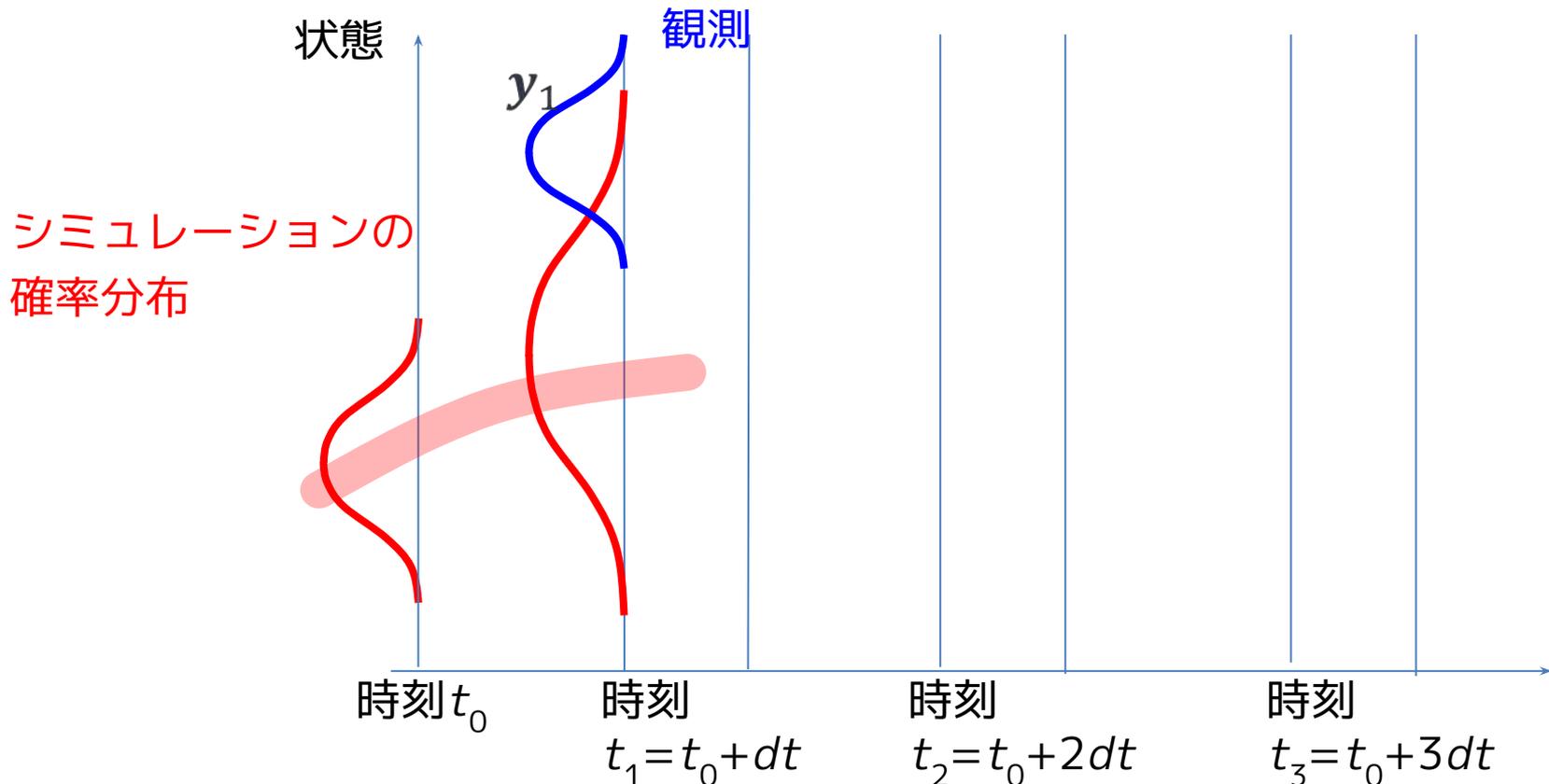
● (3)カルマンゲインを計算

- 不確実性が大きい部分は、観測値を良く使おう：カルマンゲイン大きめ
- 不確実性が小さい部分は、予測を信じよう：カルマンゲイン小さめ



カルマンフィルタ

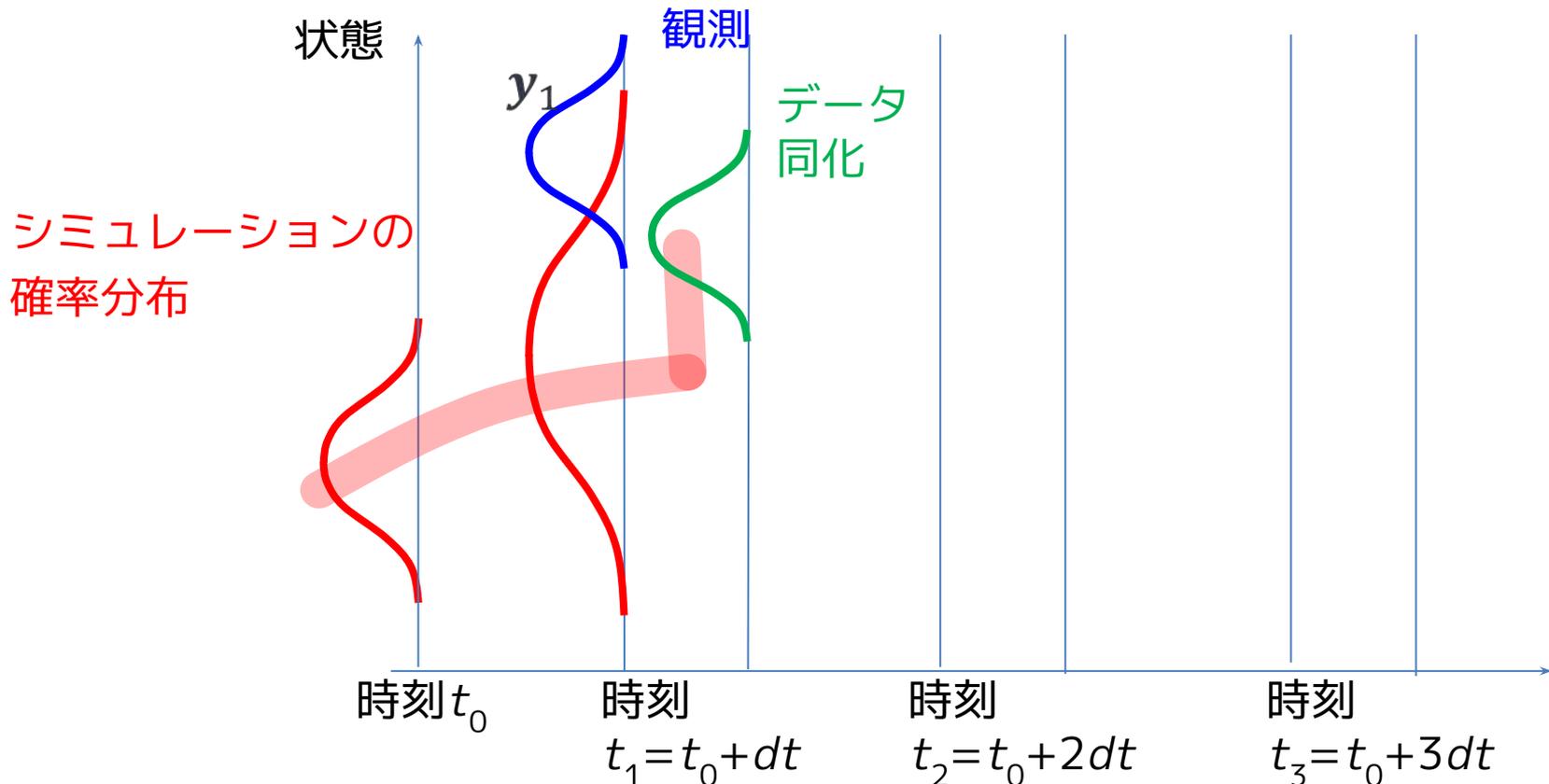
- (4)観測値を取得
 - 観測値にも、平均値と分散（不確実性）がある



カルマンフィルタ

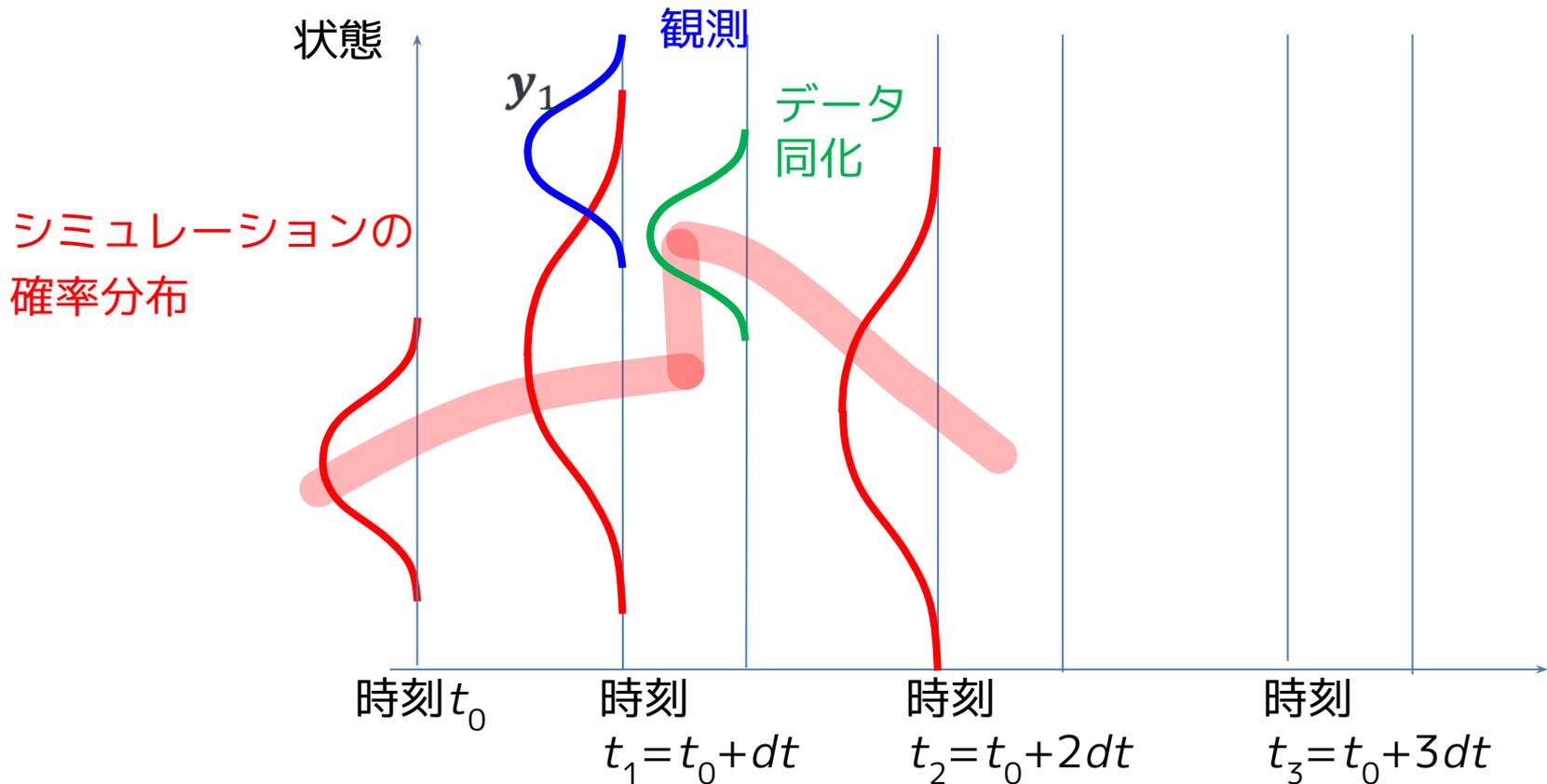
$$\mathbf{x}_{t|t} = \mathbf{x}_{t|t-1} + K_t(\mathbf{y}_t - H_t\mathbf{x}_{t|t-1}) \dots (4)$$

- (5)カルマンゲインを使って、シミュレーションを更新
 - 観測値とシミュレーションの差とカルマンゲインの行列積が更新量



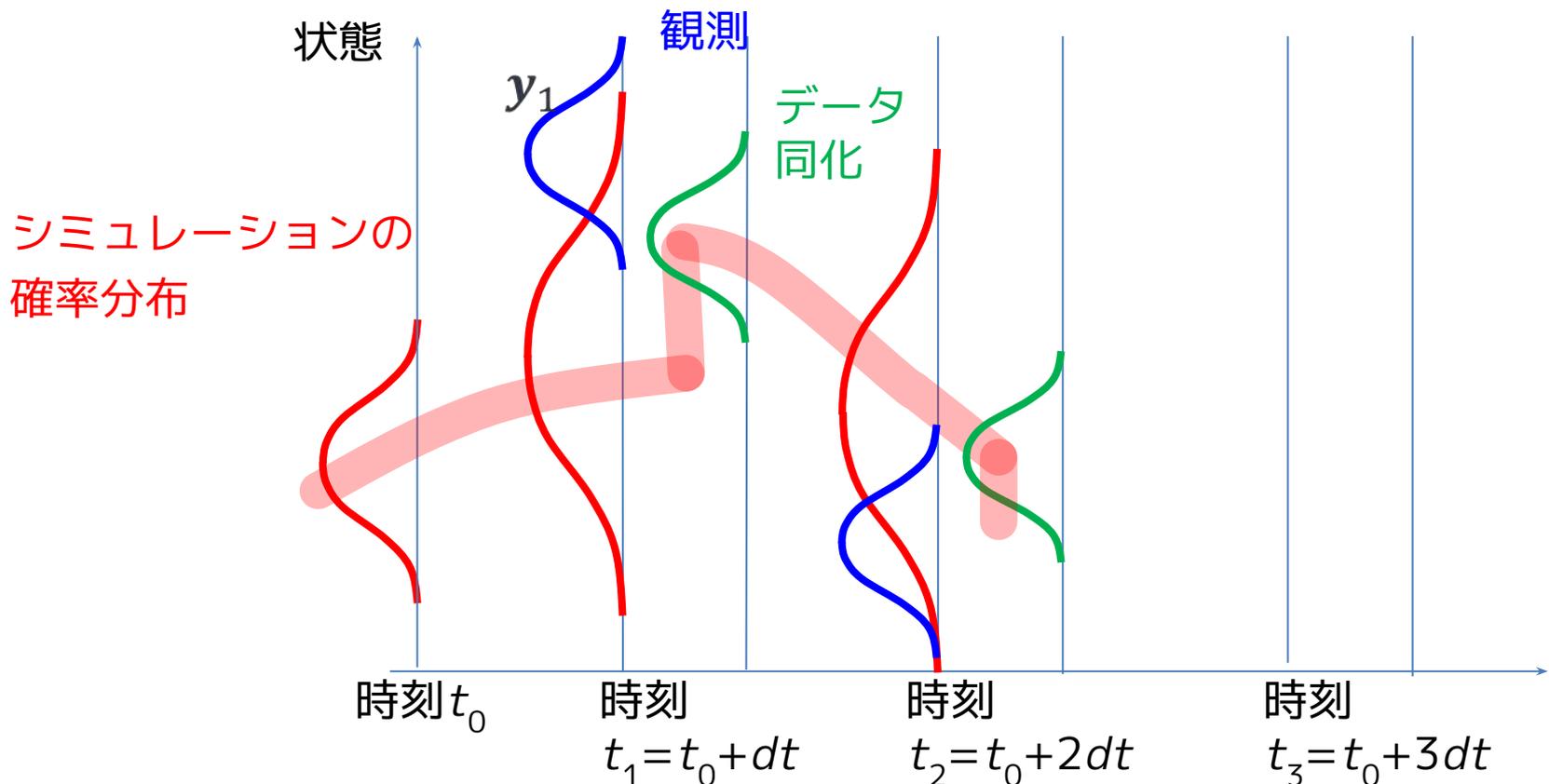
カルマンフィルタ

- (6)同様の動作を繰り返す中で、シミュレーションと観測を融合していく



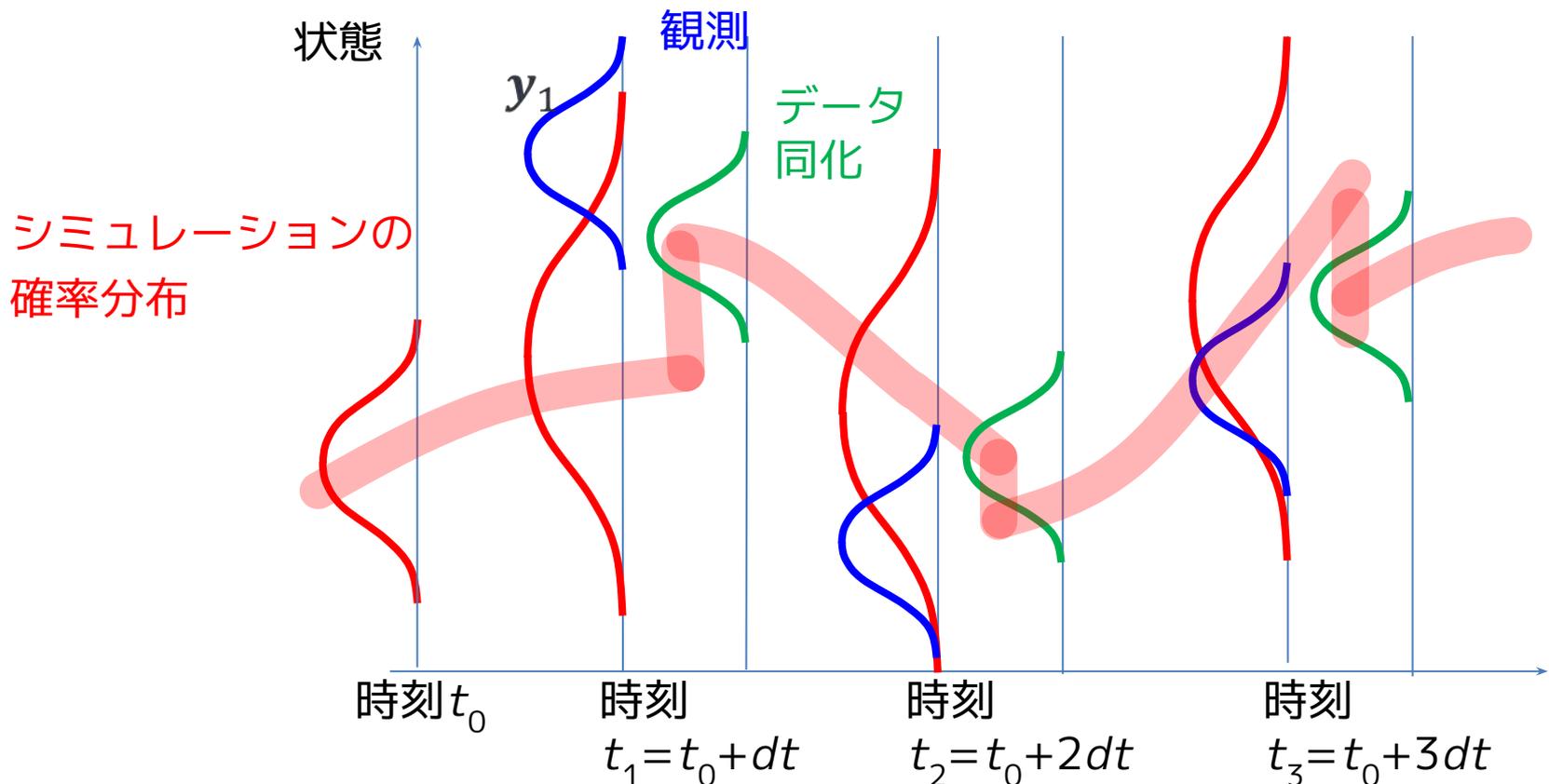
カルマンフィルタ

- (6)同様の動作を繰り返す中で、シミュレーションと観測を融合していく

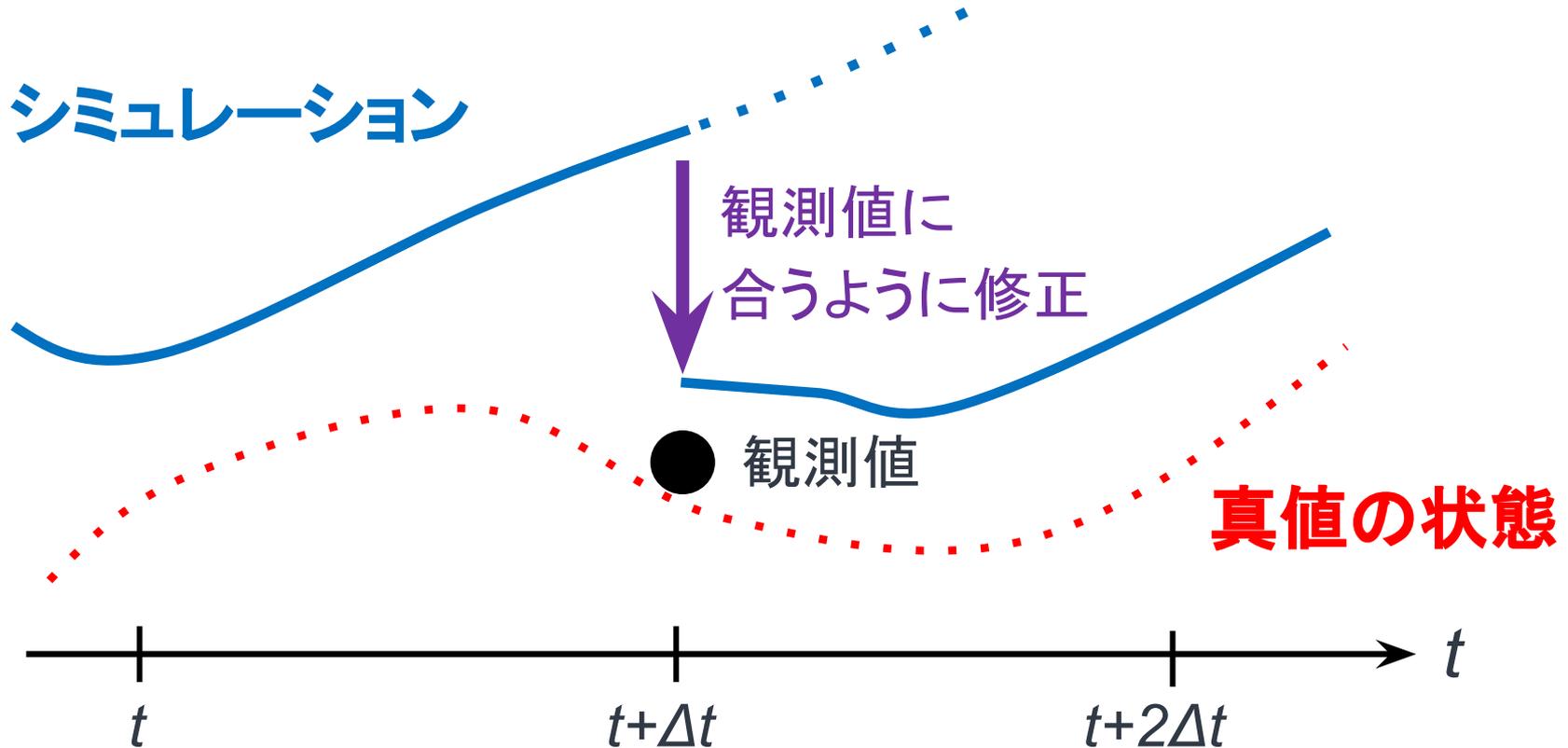


カルマンフィルタ

- (6)同様の動作を繰り返す中で、シミュレーションと観測を融合していく



データ同化の基礎まとめ



シミュレーションと観測はそれぞれ“不確実性”を持つ
⇒現実に近い情報を推定する

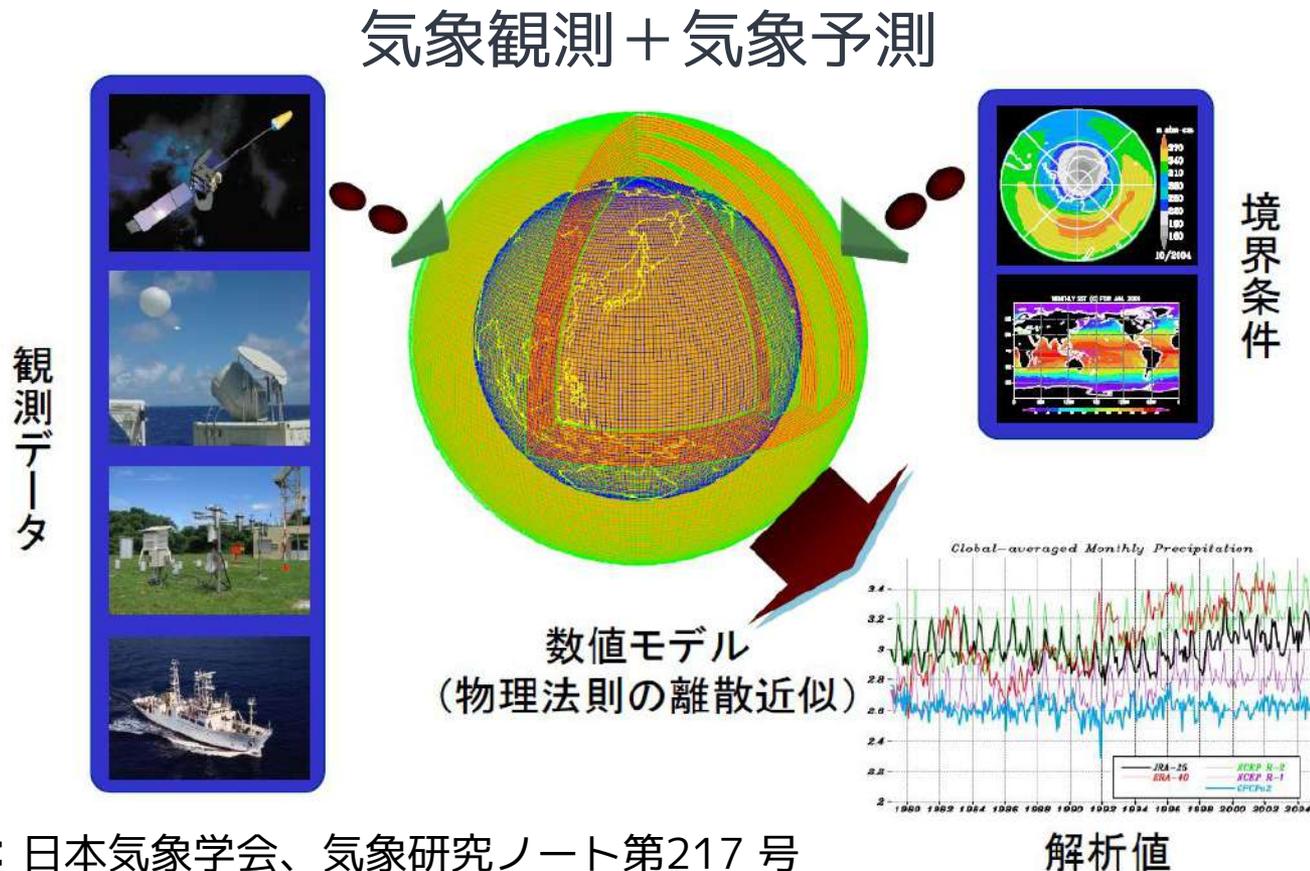
データ同化手法の 活用

データ同化の主な効果・目的

1. 予測や解析を行うため最適な初期・境界条件を求める
 - a. 流体計算、気象予測、構造計算etc
 - b. 現実に即した計算を実施するために
2. モデルに含まれる未知のパラメータを求める
 - a. 乱流モデル、反応モデル、物理過程etc
 - b. モデルの改善や改良
3. 観測・実験値の未知の誤差や観測過程を求めたい
 - a. 観測値の誤差構造の推定、誤差モデルの推定
 - b. 実験計画にフィードバック
4. 観測や実験データがない場所や時間の物理量を求める
 - a. データの内挿、データベース作成
 - b. 調査や研究の基盤的なデータベース（再解析）

データ同化の主な効果・目的

データ同化の適用例：初期・境界条件の推定



引用：日本気象学会、気象研究ノート第217号
「気象学におけるデータ同化」

データ同化の主な効果・目的

データ同化の適用例：パラメータ推定

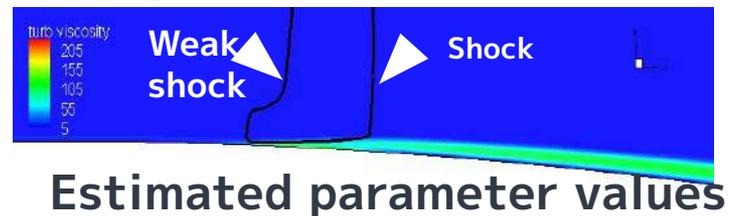
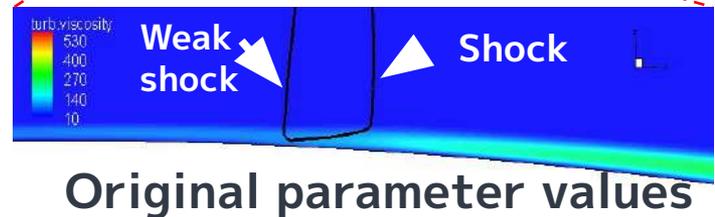
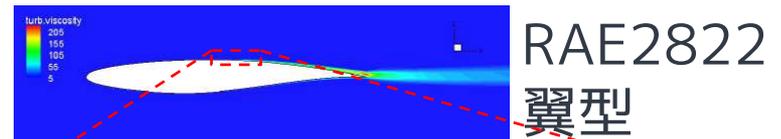
Spalart-Allmaras turbulence model (without f_{t2} term)

$$\frac{\partial \rho \hat{v}}{\partial t} + \frac{\partial \rho u_j \hat{v}}{\partial x_j} = \underbrace{\rho C_{b1} \hat{S} \hat{v}}_{\text{(Production)}} - \underbrace{\frac{M_\infty}{Re_\infty} C_{w1} f_w \left(\frac{\hat{v}}{d}\right)^2}_{\text{(Destruction)}} + \frac{M_\infty}{Re_\infty} \frac{\rho}{\sigma} \left[\frac{\partial}{\partial x_j} \left((v + \hat{v}) \frac{\partial \hat{v}}{\partial x_j} \right) + \underbrace{C_{b2}}_{\text{(Diffusion)}} \frac{\partial \hat{v}}{\partial x_i} \frac{\partial \hat{v}}{\partial x_i} \right]$$

風洞計測 + 流体計算

シミュレーションに含まれる
不確実性を減らすことで
高精度な数値計算を実現

Kato, Hiroshi, et al. "A data assimilation methodology for reconstructing turbulent flows around aircraft." *Journal of Computational Physics* 283 (2015): 559-581.

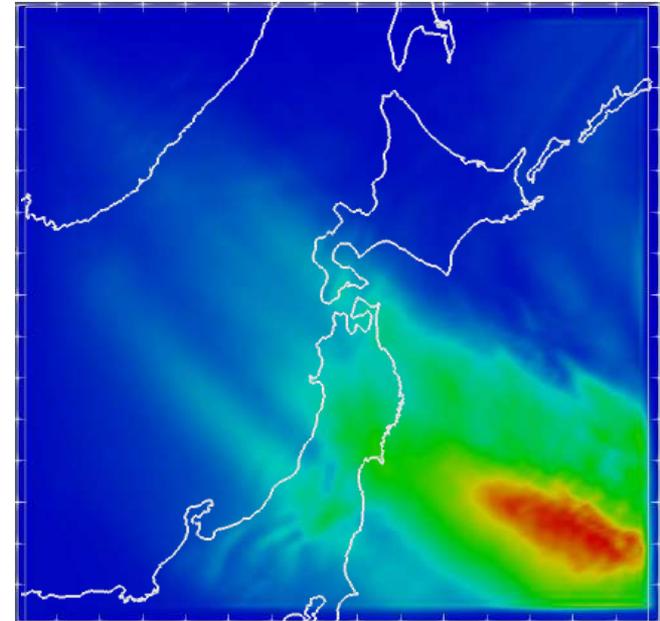


データ同化の主な効果・目的

データ同化の適用例：観測すべき点の探索

気象観測 + 気象予測

計算領域



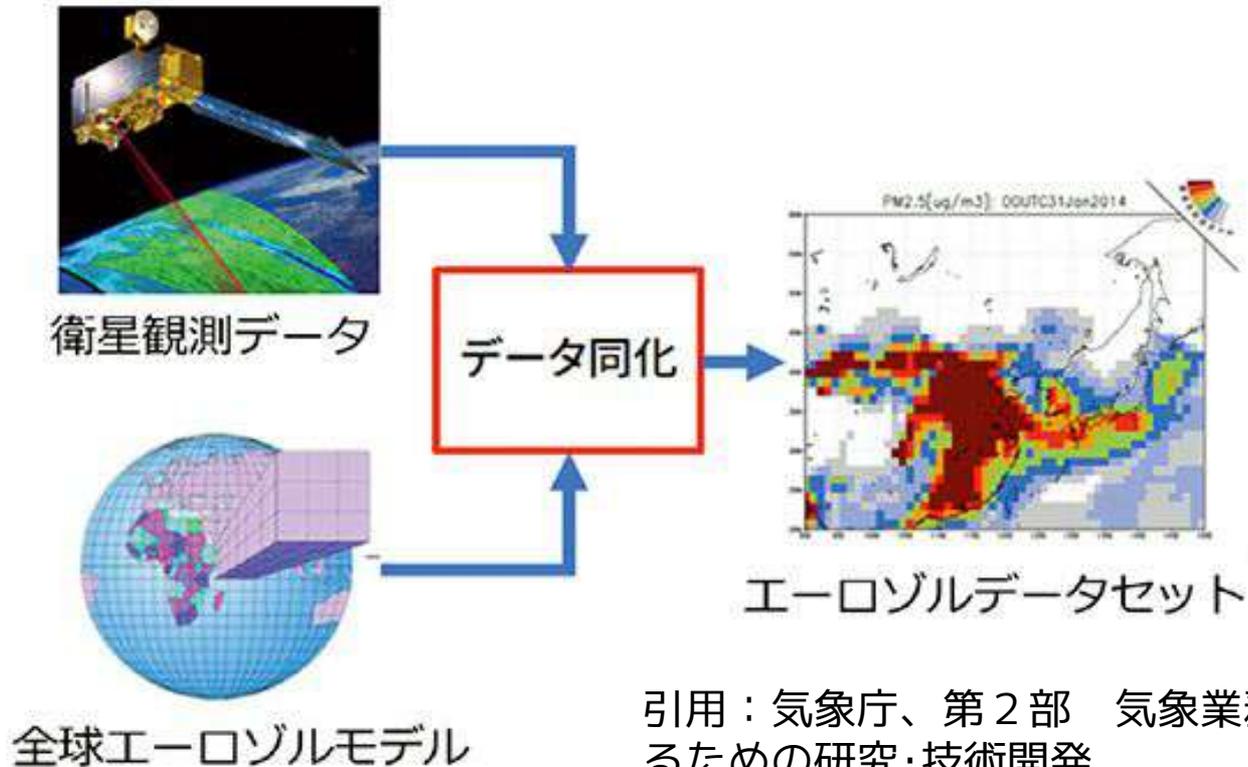
計算領域での
可観測性分布

Yoshimura, Ryoichi, et al. "Application of observability Gramian to targeted observation in WRF data assimilation." *Tellus A: Dynamic Meteorology and Oceanography* 72.1 (2020): 1-11.

データ同化の主な効果・目的

データ同化の適用例：均一なデータセット作成

エアロゾル再解析データセットの概要



引用：気象庁、第2部 気象業務を高度化するための研究・技術開発

<https://www.jma.go.jp/jma/kishou/books/hakusho/2018/index4.html>

本講義のまとめ

- 「実験・観測値」
 - メリット：現実の信頼できるデータ
 - デメリット：大量の計測を用意することや、時空間的に密な計測が出来ない場合がある
- 「数値解析・シミュレーション」
 - メリット：時空間的に密な情報や、精細な情報を取得可能
 - デメリット：初期境界条件や、パラメータに不確実性がある
- 「データ同化」：お互いを補い合い、より良い現象把握・予測が可能に
 - 初期境界条件の推定
 - 未知パラメータの推定
 - 観測・実験値の不確実性への貢献
 - 統計的なデータベースの作成