

データ同化 - データを活かしたシミュレーションと予測の方法

明治大学 総合数理学部現象数理学科/大学院先端数理科学研究科
中村和幸 (knaka@isc.meiji.ac.jp)

データ同化とは…

- ▶ 気象学・海洋学の分野で発達
- ▶ 物理数値シミュレーションモデルと観測データを統合する手法
 - ▶ シミュレーションのみでは適切に物理現象を再現できない
 - ▶ (例えば) 気象予測には適切な初期値の構成が必要
 - ▶ そもそもシミュレーションモデルには、モデル化されない力学、未知境界条件・パラメータといった不確かさが存在
 - ▶ 観測データは物理的・社会的制約により得られる情報に限界がある

▶ 観測データを使い数値シミュレーション内の変数を修正

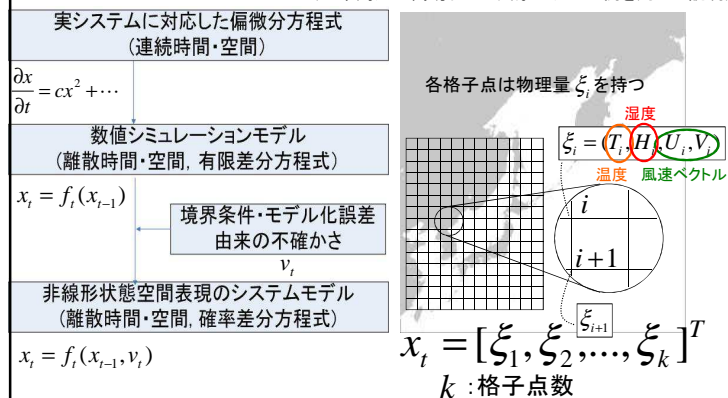


データ同化の狙い・可能とするところ

- ▶ 観測できない物理変数の推定
 - ▶ 予測のための初期条件の構成
 - ▶ 予報精度の向上を目指す
 - ▶ 現業の天気予報ですでに行われている
 - ▶ 感度解析
 - ▶ 効率のよい計測点・データの設計
 - ▶ シミュレーションモデルの改善
-
- ▶ 経験的パラメータの推定
 - ▶ 境界条件の推定
 - ▶ (シミュレーションモデルの選択)

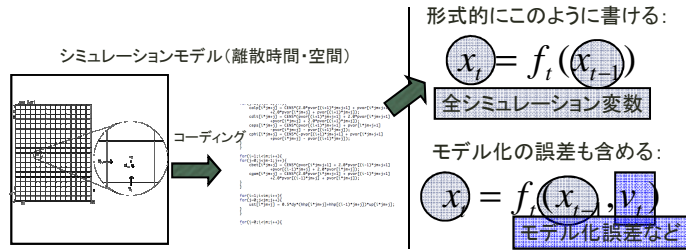
方程式からシステムモデルへ

(日本周辺の簡易化した気象モデルの例を用いて説明)



シミュレーションモデルとシステムモデル

- ▶ シミュレーションモデルの「誤差」、初期・境界条件などによる状態の誤差が反映されていない
- ▶ このような誤差まで含めたモデルを考え、システムモデルを定式化



観測情報と観測モデル

- ▶ ほとんどの場合、観測情報はシミュレーションの情報に比べて圧倒的に不足。ダイナミクスを伴う逆問題。
- ▶ さらに、時点間で独立な「観測ノイズ」もある
- ▶ 観測情報は、「その時点の全物理変数(=全シミュレーション変数)、および「観測ノイズ」が与えられれば、説明できる」という定式化

全観測変数

観測ノイズ

$$y_t = h_t(x_t, w_t)$$

全シミュレーション変数

$\dim(x_t) \gg \dim(y_t)$

$10^4 \sim 10^9$ $10^1 \sim 10^6$

状態空間モデルと推定アルゴリズム

▶ (非線形)状態空間モデル

全シミュレーション変数

モデル化誤差など

$$x_t = f_t(x_{t-1}, v_t)$$

全観測変数

観測ノイズ

$$y_t = h_t(x_t, w_t)$$

$\dim(x_t) \gg \dim(y_t)$

- ▶ この形で表現できると、以下の二つが可能となる
 - ▶ 観測 y_t を得る毎の x_t の推定(オンライン推定)
 - ▶ 一定時間幅の全ての y_t を用いて、その区間の x_t を推定(バッチ)
 - ▶ どちらも、ベイズアンモデリングの観点で捉えられる

データ同化アルゴリズム一覧

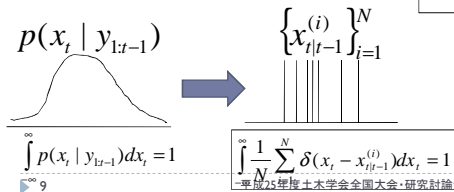
- ▶ Kalman filter
 - ▶ Extended Kalman filter
 - ▶ Ensemble Kalman filter (EnKF)
 - ▶ 気象・海洋分野で主流
 - ▶ さまざまな変種あり(EAKF,ETKF,...)
 - ▶ Particle filter
 - ▶ SIR filter の他、SIS filter もある
 - ▶ 理論的にいい面がある
- 逐次型
- ▶ 4DVAR
 - ▶ 気象・海洋分野におけるもう一つの主流(現業で使われている)
- 変分(非逐次)型
- ▶ 3DVAR 1時点の補間と隠れ変数の推定のみ
 - ▶ Nudging, OI, ... 原始的

粒子近似: 分布を実現値集合で近似する

$$p(x_t | y_{1:t-1}) \cong \frac{1}{N} \sum_{i=1}^N \delta(x_t - x_{t|t-1}^{(i)})$$

$$p(x_t | y_{1:t}) \cong \frac{1}{N} \sum_{i=1}^N \delta(x_t - x_{t|t}^{(i)})$$

δ 関数を使った近似になっている



i 番目の実現値

$x_{t|t}^{(i)}$

時刻

使用した観測の最後の時刻

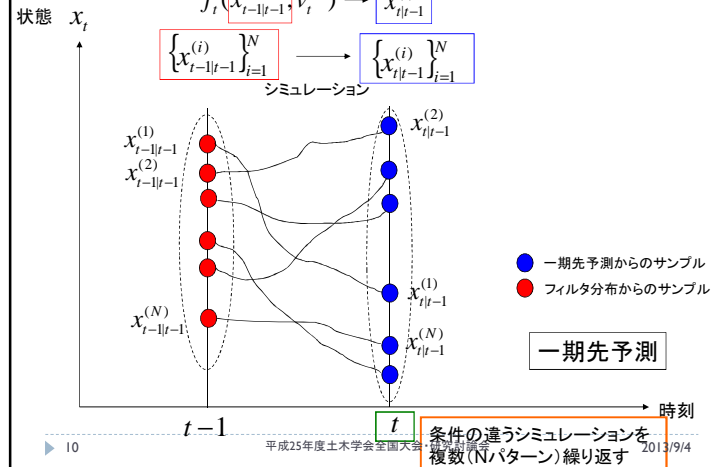
N : 実現値の個数

一期先予測 (EnKF, PF(SIR, SIS) 共通)

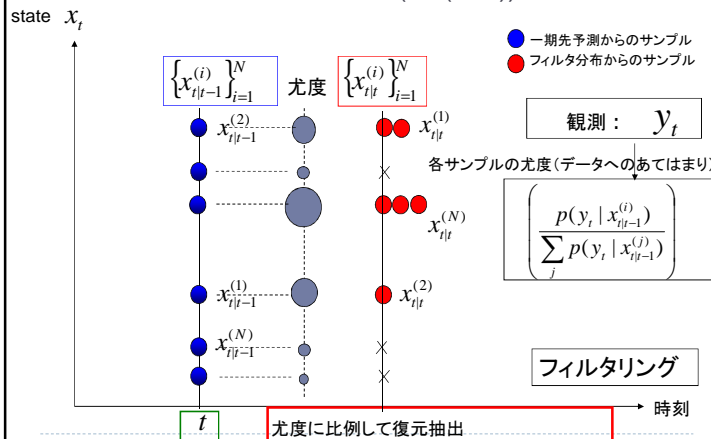
$$f_t(x_{t-1|t-1}^{(i)}, v_t^{(i)}) \rightarrow x_{t|t-1}^{(i)}$$

$$\{x_{t-1|t-1}^{(i)}\}_{i=1}^N \rightarrow \{x_{t|t-1}^{(i)}\}_{i=1}^N$$

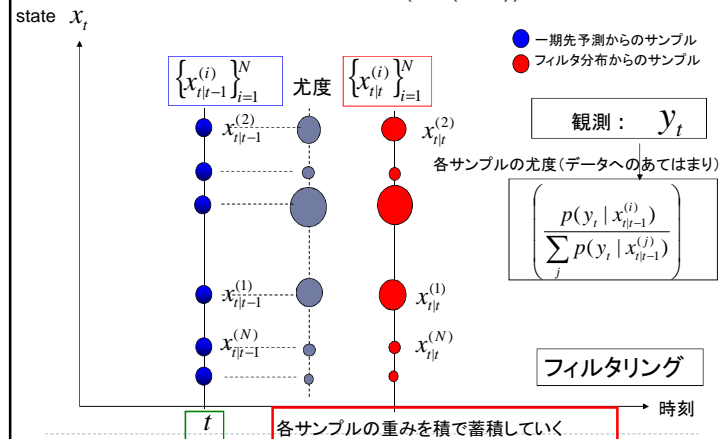
シミュレーション



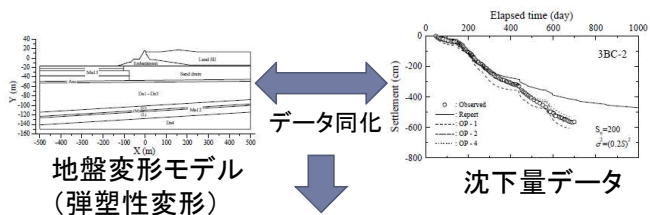
フィルタリング(PF(SIR))



フィルタリング(PF(SIS))



研究紹介1 神戸空港・地盤沈下の予測



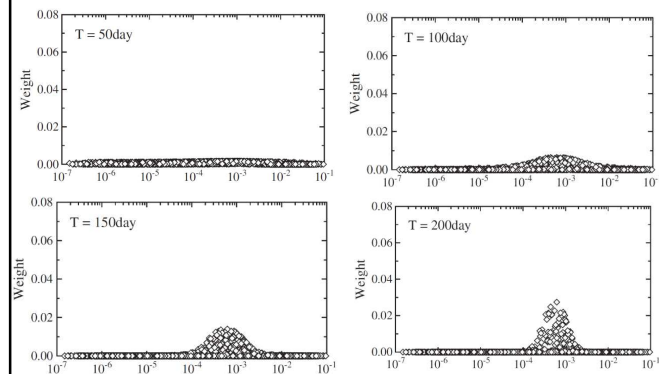
地盤変形モデル
(弾塑性変形)

沈下量データ

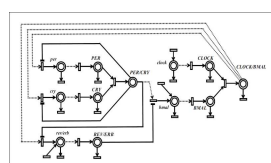
- ・直接見ることができない地中の土の状態がわかる
- ・予測精度の向上で、中途での意思決定が可能に

(珠玖隆行(岡大), 村上章, 藤澤和謙(京大), 西村伸一(岡大)各氏との共同研究)

研究紹介1(続):透水パラメータの逐次推定

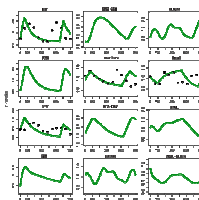


研究紹介2 マウスの体内時計モデルのデータ解析



現実の系を表すには不完全
未知パラメータ

データ同化



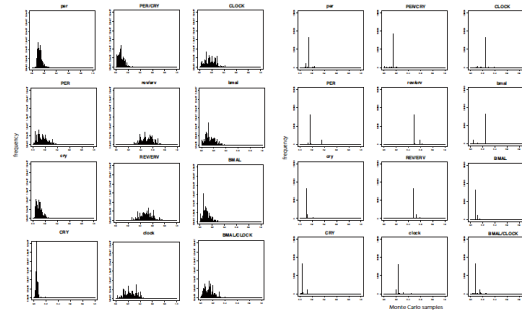
ノイズ, 欠測など

マウスの体内時計(概日周期)のメカニズム
の解明と予測

(長崎正朗(東北大), 宮野悟(東大), 吉田亮, 樋口知之(統数研)各氏との共同研究)

研究紹介2(続):パラメータの統計的推定と並列・ハイパフォーマンスコンピューティングの活用

1億サンプルによる近似 10万サンプルによる近似

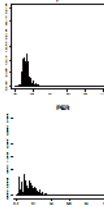


並列計算を活用することで、
分布をより正確に近似・推定ができる

予測精度が上がるだけでなく、
興味ある現象が起こる確率を
正確に出すことができる

なぜ分布なのか？

- ▶ 確率分布で捉えることで...
 - ▶ どの範囲にどの確率で入るか考えられるので、リスク評価に用いることができる
 - ▶ さらに、「入ってほしくない」範囲の各階級に損失を定義することにより、トータルコストの計算が可能となる
- ▶ そもそも、大規模空間では、不確実性は必ず現れる
 - ▶ 各パラメータが10通りの値を取るとして、10個あったら 10^{10} (=100億)通り
 - ▶ 全ての場合をデータから集めるのは不可能



金融工学における商品価値評価・意思決定との関連

- ▶ 金融市場は不確実性を持つため、確率的な評価がなされてきた
 - ▶ 各種確率解析による金融商品の将来価格の評価
 - ▶ たとえば、ブラックショールズモデルでは「株価は幾何ブラウン運動に従う」と仮定
 - ▶ 将来の利益または損失の期待値で現在価値を評価
 - すなわち、(将来の利益または損失) × (その確率) の和
 - ▶ データ同化により将来状態の「確率」を予測することで、期待価値によって「現在の価値」を評価可能
- ➡ 不確実性のある現象を定量的に捉え、意思決定するための方法

ベイズ推定と確率的モデリング(1)

現象Xが発生した条件下で
データYが得られる確率

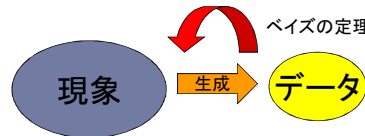
現象Xが発生する
「もともとの」確率

$$p(X | Y) = \frac{p(Y | X)p(X)}{p(Y)}$$

データYが得られた時に
現象がXである確率

$$p(Y)$$

データYの生成確率 (必要なのは $p(Y|X)$ と $p(X)$)

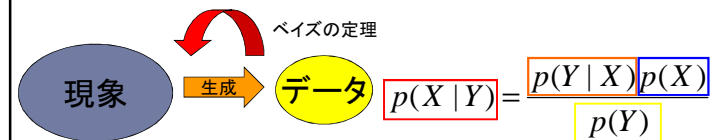


データ生成モデルと現象の発生確率を与えれば、データから現象の説明ができるようになる(データ同化も同じ仕組み)

□ : 事前知識やシミュレーション
□ : 観測を表す式

ベイズ推定

ベイズ推定と確率的モデリング(2)



生成モデル $p(Y | X)$ と事前分布 $p(X)$ の適切な構成 (=モデリング) を通じて、有用な事後分布 $p(X | Y)$ による情報抽出をめざす。

大量データ・大規模計算の時代の解析人材育成

- ▶ 数理学・計算機科学を知識と研究の基盤とし、応用分野に挑戦していく人材はまだ少ない
 - ▶ 数理学・情報科学側の問題
 - ▶ 応用研究の位置づけ
 - ▶ 応用分野の広さによる教育の難しさ
 - ▶ 研究機会創出の困難性
 - ▶ 問題を「はっきりさせる」までに時間がかかる
 - 地盤変形の共同研究には、方向性をはっきりさせるのに半年かかっている
 - ▶ 集中的にコミットできる環境が必要だが、...
 - ▶ プロジェクトベースのポストク雇用とキャリアパスの関係

明治大学における取り組み

- ▶ 「社会に貢献する数理学」の観点から、次の三本柱に関する研究・教育を進めている
 - ▶ 現象から数式を作る「モデリング」
 - ▶ 数式を計算機上で解いて再現する「シミュレーション」
 - ▶ 上記結果や数式の解析から現象を理解する「数理解析」
- ▶ 研究：先端数理学インスティテュート(研究機関)・GCOE「現象数理学の形成と発展」
 - ▶ 各分野研究者との交流
 - ▶ テーマ別の研究集会・大学院集中講義
- ▶ 教育：大学院先端数理学研究科(2011年度開設)・総合数理学部現象数理学科(2013年度開設)
 - ▶ 数理と情報を軸に上記を学ぶカリキュラム
 - ▶ 各分野と数理学分野の相互交流可能な人材を育成

まとめ

- ▶ データ同化は大規模シミュレーションとデータを統合
 - ▶ 基礎方程式がしっかりしている場合に良く使われる
 - ▶ 非線形性により、難しさが変わってくる
- ▶ 予測のための大規模シミュレーション・大規模データ解析には、新しい道具が必要
 - ▶ 確率的モデリングが必須
 - ▶ ベイズ推論により、不確実なものを不確実なまま捉えてデータと融合できる
 - ▶ 期待値を計算することで、意思決定まで考慮にいれた議論が可能となる