

## 2. データ同化法の変遷

# データ同化の歴史(1)

データ同化は数値天気予報(NWP)から始まった。

- 1922 Richarsonが手計算で数値天気予報を試みたが失敗(145hPa/6hの気圧変化を予測)
- 1950 Charneyらが電子計算機ENIACにより数値予報実験に成功(1日予報に1日要した)

初期値:主観(手書き)解析から客観(計算機)解析へ

- 1954 スウェーデン気象水理局が数値予報業務開始
- 1959 気象庁が数値予報業務開始(世界で3番目)

客観解析法:関数あてはめ法、逐次修正法、最適内挿法

(後二者では、モデル予測値を第一推定値に採用)

# データ同化の歴史(2)

## 非定時観測データへの対応

- 1969 Charneyらがデータ同化を提唱
- 1979 全球大気研究計画第1回全球実験(GARP/FGGE)  
⇒ 全球大気観測システムの確立

## データ同化の発展

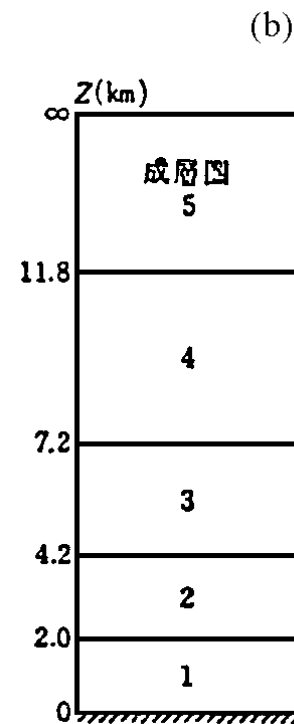
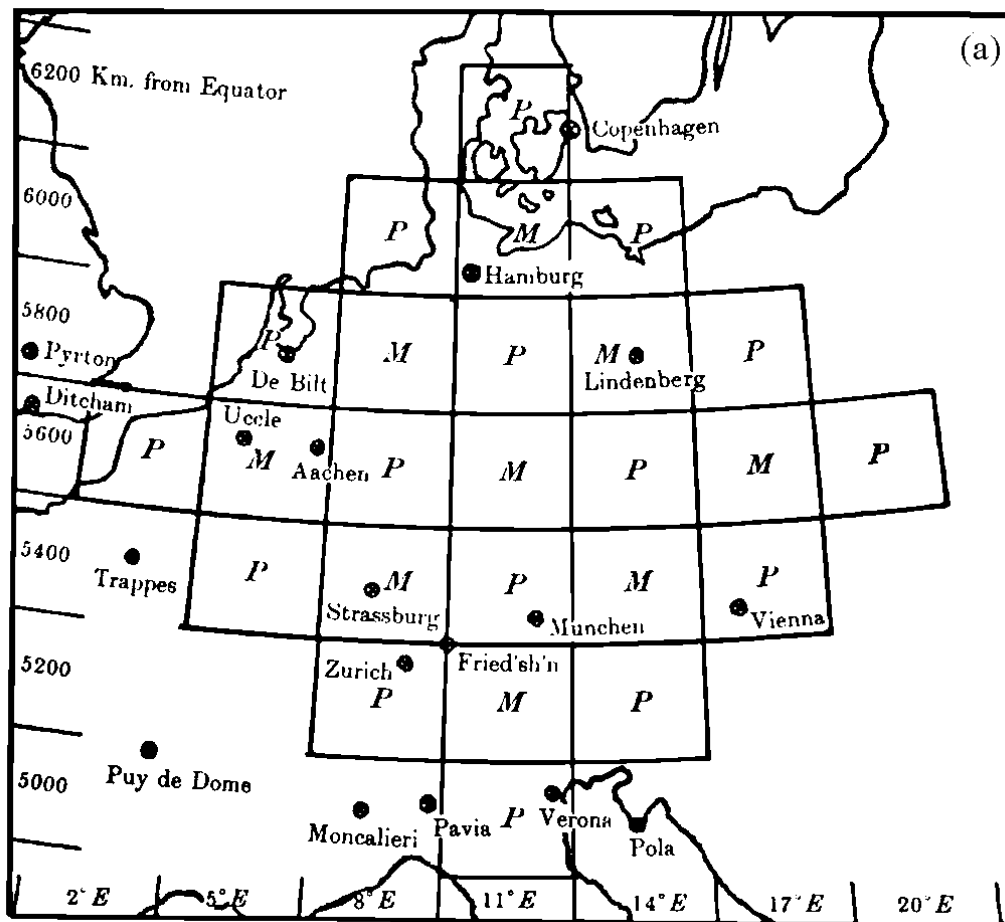
- 1980年代～ 変分法やカルマンフィルタの研究
- 1990年代～ 全球大気長期再解析の実施

## 他の分野では

- 1980年代～ 海洋データ同化の研究
- 1990年代後半～ 大気微量成分データ同化の研究

# Richardsonの数値予報実験

(1910年5月20日07UTC)



(Richardson, 1922)

# Charneyらによる最初の数値予報

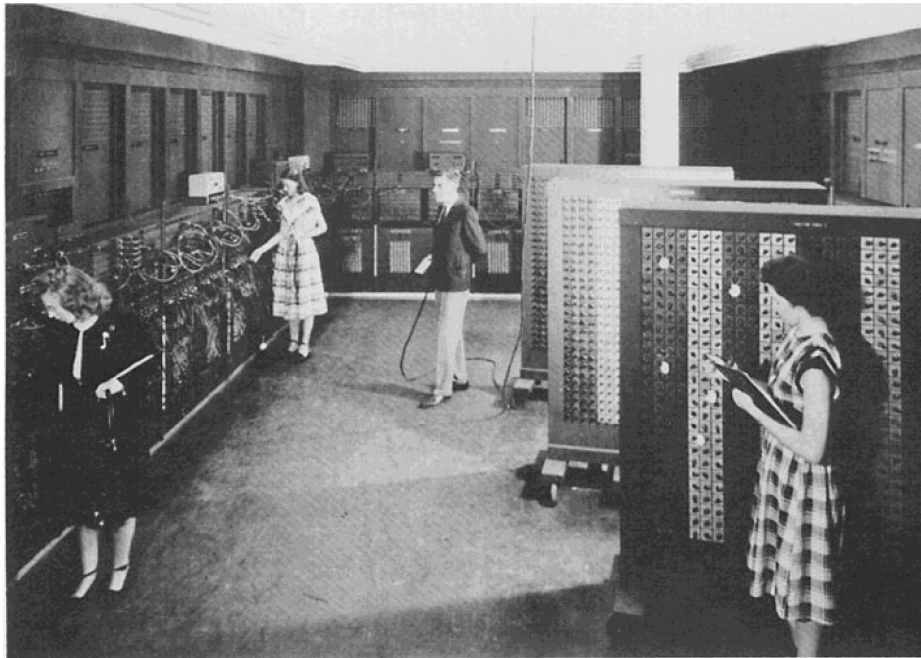
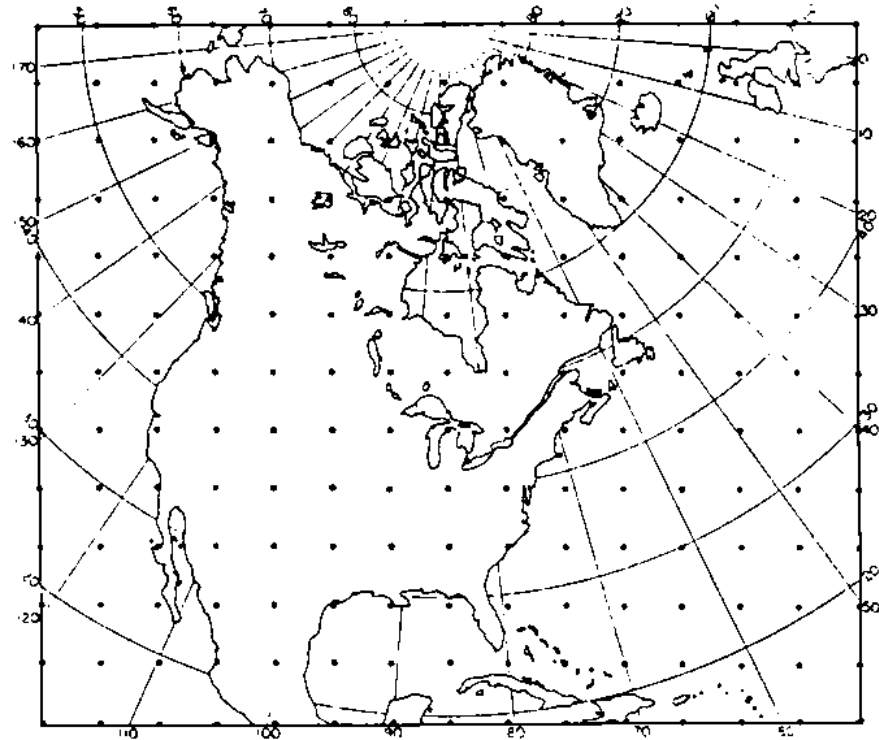


FIG. 4. Electronic Numerical Integrator and Computer (ENIAC), 7 July 1948.  
Ballistic Research Laboratory, Aberdeen Proving Ground, Aberdeen, Maryland.

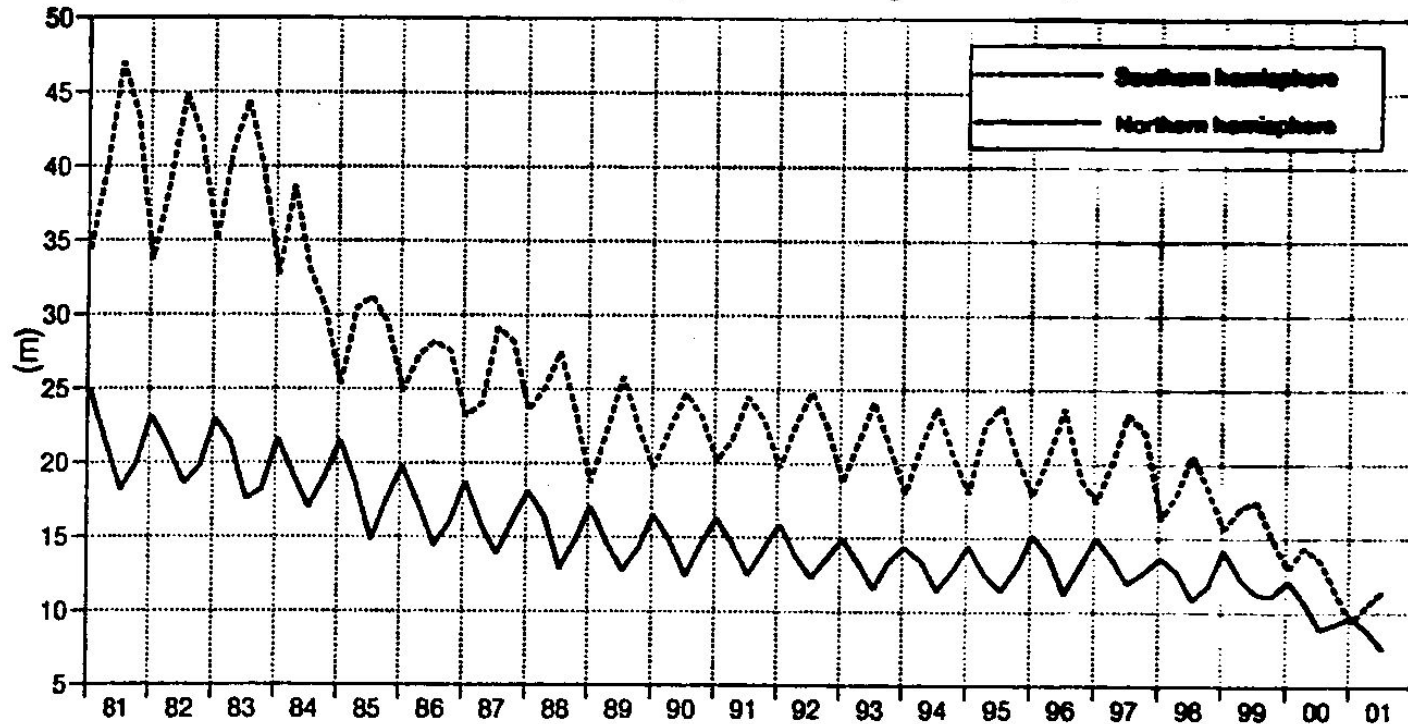
最初の数値予報に使われた計算機、ENIAC。  
Platzman et al. (1979)より



最初の数値予報に使われたモデルの格子配置。  
Charney et al. (1950)より

# ECMWFにおける 500hPa高度の予報精度の推移

R.m.s. error of day-1 500hPa height forecasts

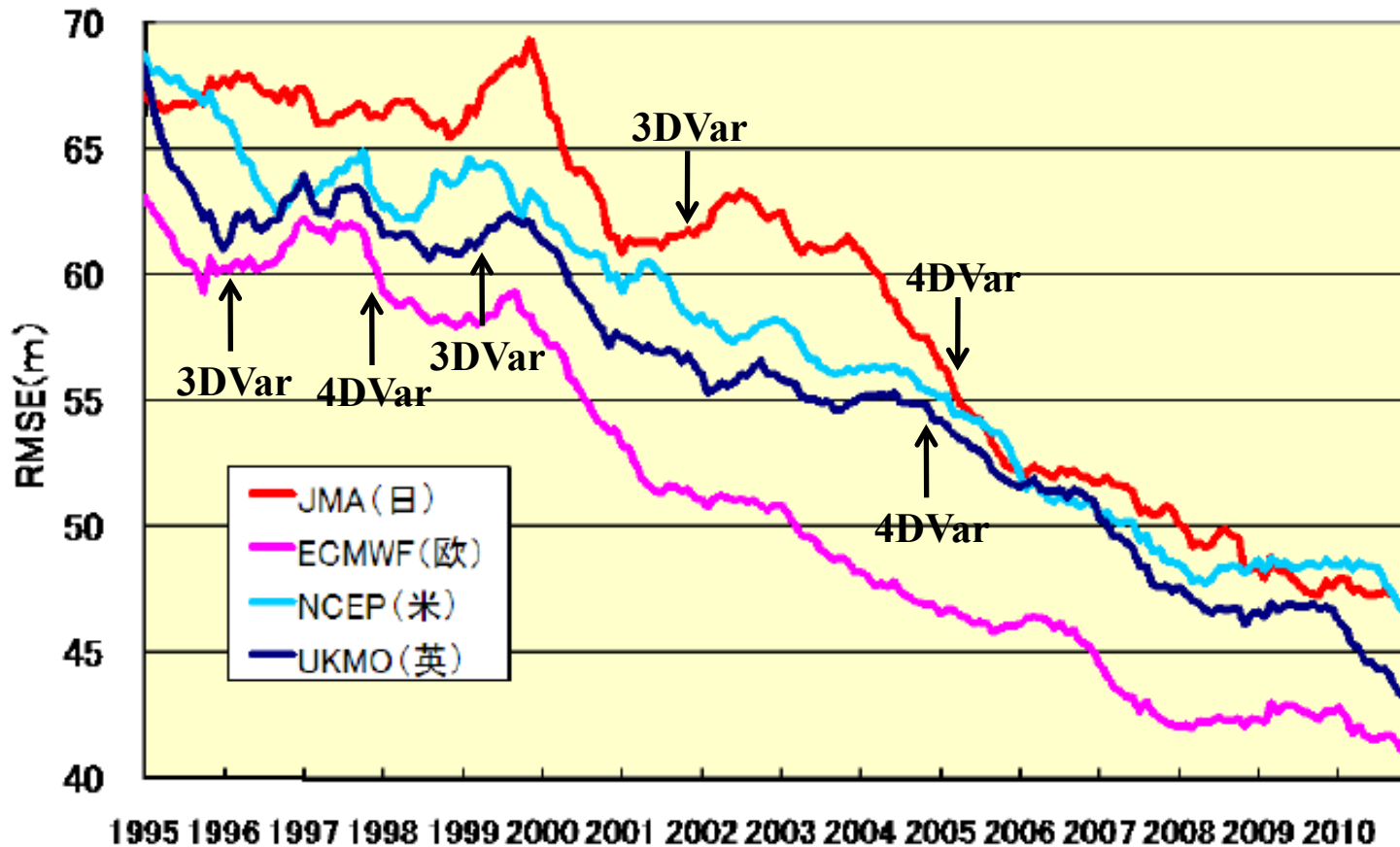


- ・衛星輝度温度データ同化
- ・3次元変分法導入
- ・4次元変分法導入
- ・モデル物理過程改良など
- ・同化システムの誤差統計量改訂など

(Simmons and Hollingsworth, 2002)

# 全球数値予報の精度の推移

図6-3 2010年11月までの北半球500hPa高度予報誤差(120時間予報)



# 関数あてはめ法

## Weierstrassの近似定理

閉区間の連続関数は、多項式によって任意の精度で近似できる。

Panofsky (1949), Gilchrist and Cressman (1954)

気象要素の空間分布を有限個の基底関数の線形結合で表し、その係数を観測データから最小二乗法で求める。解析領域を小領域に分割し、基底関数として多項式を用いるなどの方法がある。地衡風バランスや傾度風バランスを考慮することもできる。

ただし、予測値を用いないのでデータ同化とはいえない。

## 問題点

- 観測データの分布に大きな地域的偏りがあると、基底関数の個数を増やして観測データに近づけるほど、データ空白域の解析場が不自然になる。
- 基底関数の個数があまりに少ないと、観測データから得られる小さなスケールの現象が表現されない。
- 基底関数の選択と観測データの分布によっては、係数を求めるための逆行列計算に計算不安定が生じる。



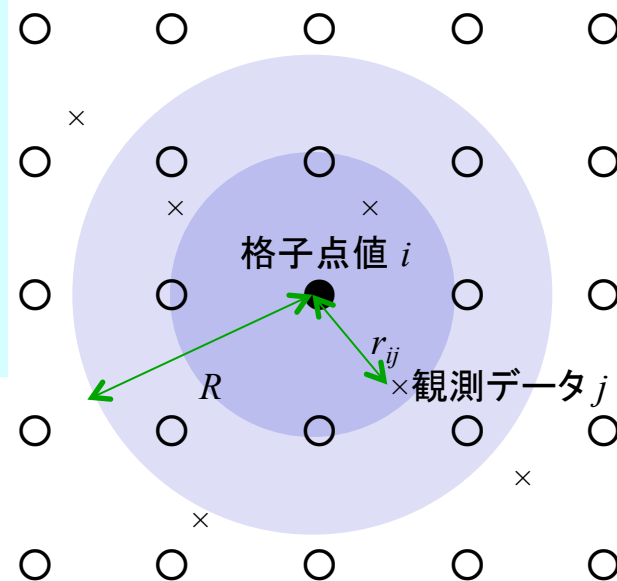
# 逐次修正法

Bergthorsson and Doos (1955)

ある範囲内の観測データに経験的な重みを与えて、各格子点の予測値を修正する。範囲を縮小させながらこの修正を反復し、解析値とする。

$$\begin{aligned}x_i^{a(0)} &= x_i^f \\x_i^{a(k)} &= x_i^{a(k-1)} + \frac{\sum_{j=1}^J w^{(k-1)}(r_{ij})(x_j^o - H_j(\mathbf{x}^{a(k-1)}))}{\sum_{j=1}^J w^{(k-1)}(r_{ij}) + \lambda^2} \\(k &= 1, 2, 3, \dots)\end{aligned}$$

$w(r)$  : 重み ( $r$  : 格子点と観測点の距離)  
 $H$  : 格子点値を観測点に内挿する演算子  
 $\lambda$  : 観測値と予測値の誤差標準偏差の比



Cressman (1959)

$$w(r) = \begin{cases} \frac{R^2 - r^2}{R^2 + r^2} & (r \leq R) \\ 0 & (r > R) \end{cases}$$

Barnes (1964)

$$w(r) = \exp\left(-\frac{r^2}{2R^2}\right)$$

# 最適内挿法(1)

Eliassen (1954), Gandin (1963)

観測データの重みを経験的に与えるのではなく、解析誤差標準偏差が最小になるように統計的に計算する。

$$x_i^a = x_i^f + \sum_{j=1}^J w_{ij} (x_j^o - H_j(\mathbf{x}^f))$$

解析誤差

$$\Delta x_i^a = \Delta x_i^f + \sum_{j=1}^J w_{ij} (\Delta x_j^o - \Delta H_j^f)$$

$$\Delta x_i^a := x_i^a - x_i$$

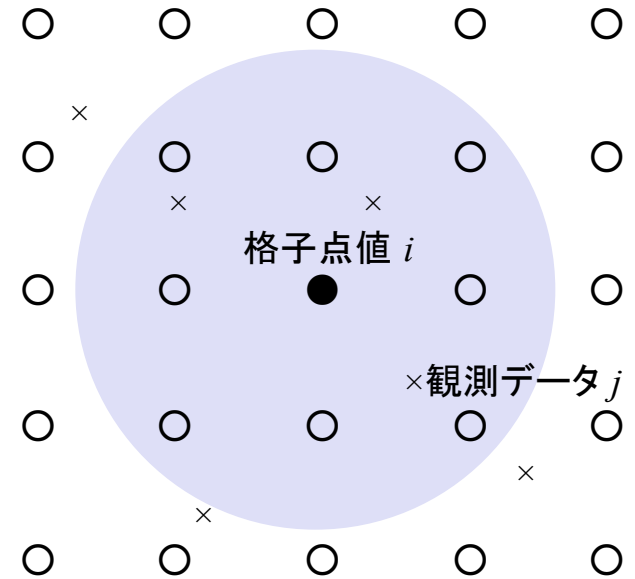
$$\Delta x_i^f := x_i^f - x_i$$

$$\Delta x_j^o := x_j^o - H_j(\mathbf{x})$$

$$\Delta H_j^f := H_j(\mathbf{x}^f) - H_j(\mathbf{x})$$

解析誤差分散

$$\left(\sigma_i^a\right)^2 = \left(\sigma_i^f\right)^2 - 2 \sum_{j=1}^J w_{ij} \left\langle \Delta x_i^f \Delta H_j^f \right\rangle + \sum_{j=1}^J \sum_{k=1}^J w_{ij} w_{ik} \left( \left\langle \Delta x_j^o \Delta x_k^o \right\rangle + \left\langle \Delta H_j^f \Delta H_k^f \right\rangle \right)$$



# 最適内挿法(2)

解析誤差分散の期待値が最小になるように、重み  $w_{ij}$  を決める。

$$\sum_{k=1}^J \left( \langle \Delta x_j^o \Delta x_k^o \rangle + \langle \Delta H_j^f \Delta H_k^f \rangle \right) w_{ik} = \langle \Delta x_i^f \Delta H_j^f \rangle \quad (j=1,2,\dots,J)$$

観測誤差の間には相関がないものとし、それらの標準偏差を  $\sigma_j^o$  とする。予測値の誤差標準偏差は水平面内で局所的に一様で  $\sigma^f$  とし、2点間の予測誤差の間の相関係数を  $\rho_{ij}$  とする。

$$\sum_{k=1}^J (\lambda_j \lambda_k \delta_{jk} + \rho_{jk}) w_{ik} = \rho_{ij} \quad (j=1,2,\dots,J)$$

$$\lambda_j := \frac{\sigma_j^o}{\sigma^f}$$

最適内挿法は当初は水平2次元内挿であったが、その後、空間3次元内挿に拡張された。また、地衡風平衡に基づいて、風とジオポテンシャル高度を同時に解析する**多変量解析**が導入され、1970年代にはいと次第に主流の客観解析法になった。Gustafsson(1981)の第5.1表によれば、当時の10の現業数値予報センターのうち、7つのセンターで最適内挿法が採用されていた。

# 多変量解析(1)

地衡風平衡などを利用して、風の観測データからジオポテンシャル高度の推定、あるいはその逆の推定などを行う。例えば、格子点*j*における東西風の観測データから、格子点*i*の高度を推定すると

$$z_i^a = z_i^f + \frac{\sigma^{(z)f} \sigma^{(u)f}}{(\sigma^{(u)o})^2 + (\sigma^{(u)f})^2} \rho_{ij}^{(zu)} (u_j^o - u_j^f)$$

必要な誤差統計量は、地衡風平衡が成り立つとして高度の誤差統計量から計算する。

$$\begin{aligned} \sigma^{(z)f} \sigma^{(u)f} \rho_{12}^{(zu)} &= \left\langle \Delta z^f(x_1, y_1) \Delta u^f(x_2, y_2) \right\rangle \\ &= -\frac{g}{f_0} \left\langle \Delta z^f(x_1, y_1) \frac{\partial \Delta z^f}{\partial y_2}(x_2, y_2) \right\rangle \\ &= -\frac{g}{f_0} \frac{\partial}{\partial y_2} \left\langle \Delta z^f(x_1, y_1) \Delta z^f(x_2, y_2) \right\rangle \\ &= -\frac{g}{f_0} (\sigma^{(z)f})^2 \frac{\partial}{\partial y_2} \rho_{12}^{(zz)} \end{aligned}$$

# 多変量解析(2)

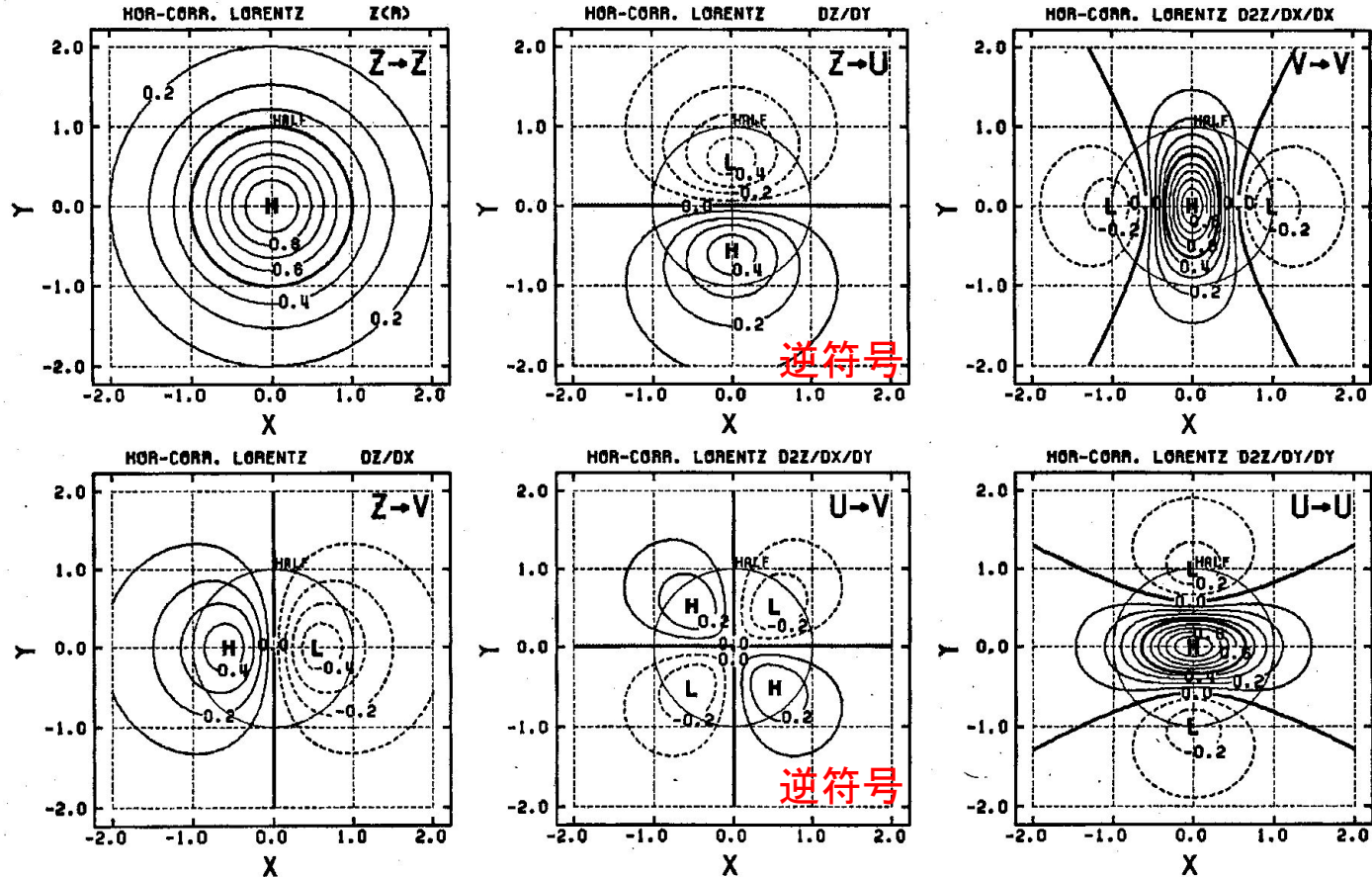


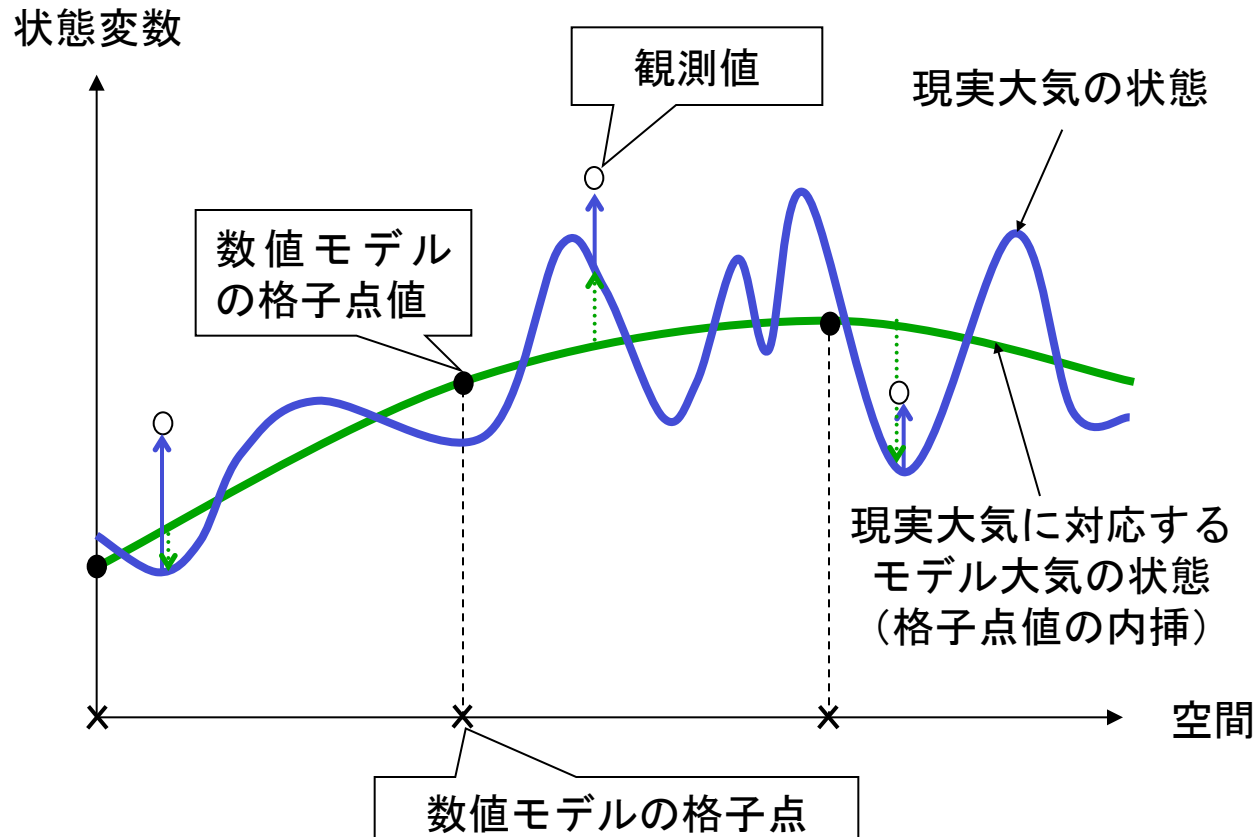
図3.3.3 ローレンツ型重相関係数の形状。高度と風の変量解析の際にはこのような重相関係数が必要となる。解析要素と観測データ要素との組み合わせに応じて、この図に示す解析インクリメントの反応が生じる。図のX, Y座標はいずれも、高度の誤差相関が0.5に落ちる距離を1（細い円で示す）として規格化されている。

# 観測データの代表性誤差

観測誤差の分解

$$\Delta x_j^o = (x_j^o - x_j^t) + (x_j^t - H_j(\mathbf{x})) \quad (j : \text{観測点の位置})$$

測定誤差 ↑ 代表性誤差 ↑



# 誤差統計量の推定(1)

## D値(Departure)

$$\mathbf{d} := \mathbf{x}^o - \mathbf{H}(\mathbf{x}^f) = \Delta \mathbf{x}^o - \Delta \mathbf{H}^f$$

$$\langle \mathbf{d} \mathbf{d}^T \rangle = \langle \Delta \mathbf{x}^o (\Delta \mathbf{x}^o)^T \rangle + \langle \Delta \mathbf{H}^f (\Delta \mathbf{H}^f)^T \rangle$$

### 仮定

- 予測誤差は水平面内で局所的に一様である。
- 予測誤差の相関係数が2点間の距離のみの関数である。
- 観測誤差が互いに無相関である。

このとき、必要な誤差統計量はD値統計から次のように算出できる。

$$(\sigma_j^o)^2 + (\sigma^f)^2 = \langle d_j^2 \rangle$$

$$(\sigma^f)^2 \rho(r_{jk}) = \langle d_j d_k \rangle \quad (j \neq k)$$

$$(\sigma^f)^2 = \lim_{r_{jk} \rightarrow 0} \langle d_j d_k \rangle$$

# 誤差統計量の推定(2)

同じ種類の観測データだけでD値統計を取れば、観測誤差標準偏差は一定とみなせるので、次の式が成り立つ。

$$\frac{\langle d_j d_k \rangle}{\langle d_j^2 \rangle} = \frac{(\sigma^f)^2}{(\sigma^o)^2 + (\sigma^f)^2} \rho(r_{jk}) \quad (j \neq k)$$

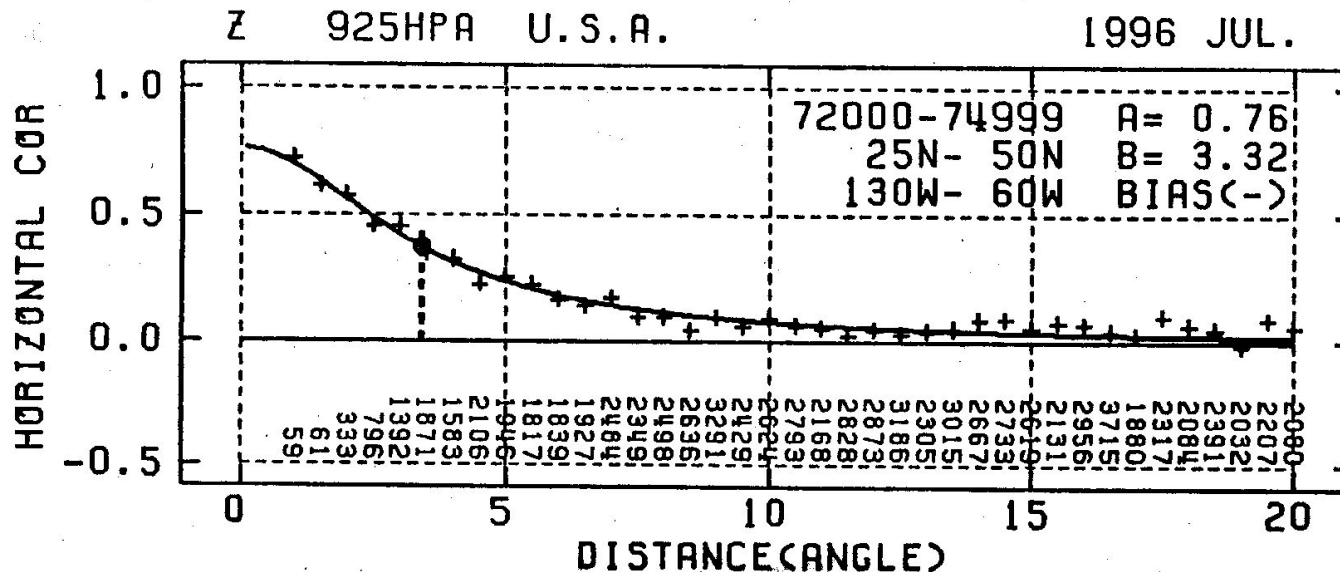


図3.3.2 高度の実際の誤差相関例。1996年7月におけるアメリカ合衆国上空の数十地点のラジオゾンデ観測から求めた925hPa高度の相関(+で表示)。横軸は距離(角度)で、縦軸が相関の値を示す。グラフの下部にプロットされている数字は、相関計算に用いた観測対の総計である。また、実線はフィッティングさせたモデル関数である。

(多田, 1997)



# 最適内挿法の問題点

- 空間内挿に基づいた方法であるため、数値モデルの状態変数と同じ物理量の観測データしか同化できない。したがって、衛星による輝度温度データなどは、逆解析によって気温や湿度にあらかじめ変換する必要があるが、それによってさらなる仮定が導入される。
- 予測誤差の統計量には長期間の平均的な値が使われるため、それが循環場の状態や観測データの分布によって変動することが考慮されない。
- 観測データの分布が解析時刻ごとに大きく変動しないことを仮定しているため、衛星観測などの非定時の観測データを適切に扱えない。
- 計算上の問題から、各格子点ごとに同化する観測データの範囲が狭く、そのために解析値が力学的にバランスすることが保証されない。

これらの問題点を解消するものとして、4次元変分法やアンサンブルカルマンフィルタなどの高度なデータ同化法が導入された。

# 変分法の最適内挿法に対する利点

## 3次元変分法

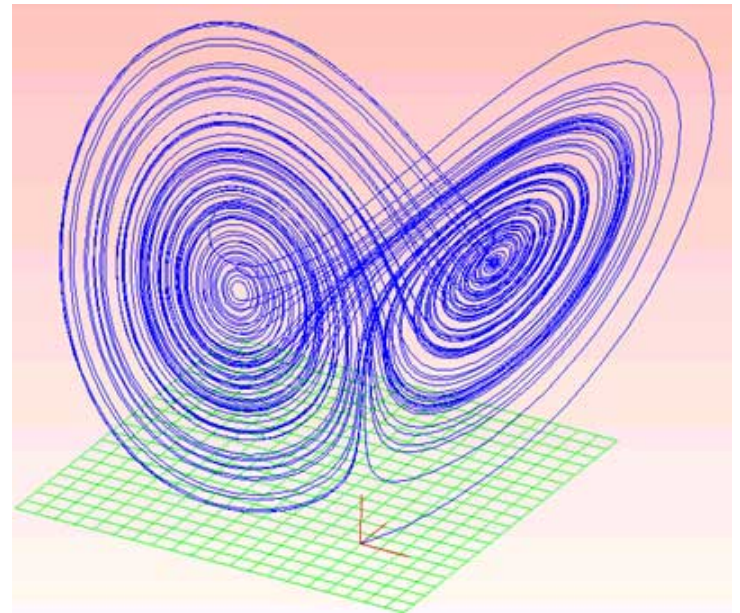
- ・数値モデルの状態変数から算出可能な任意の物理量の観測データを同化できる
- ・すべての解析値を一括して算出するため、力学的なバランスに優れた解析値が得られる

## 4次元変分法(上記に加えて)

- ・数値モデルによって表現される物理法則をデータ同化に直接利用できる
- ・非定時観測データをその時刻のデータとして同化できる

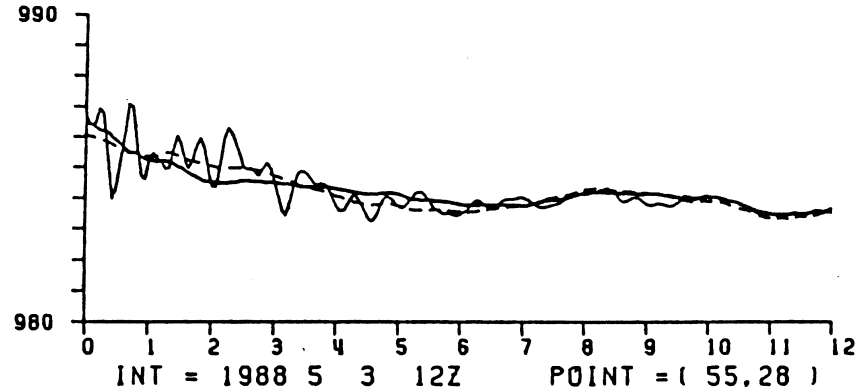
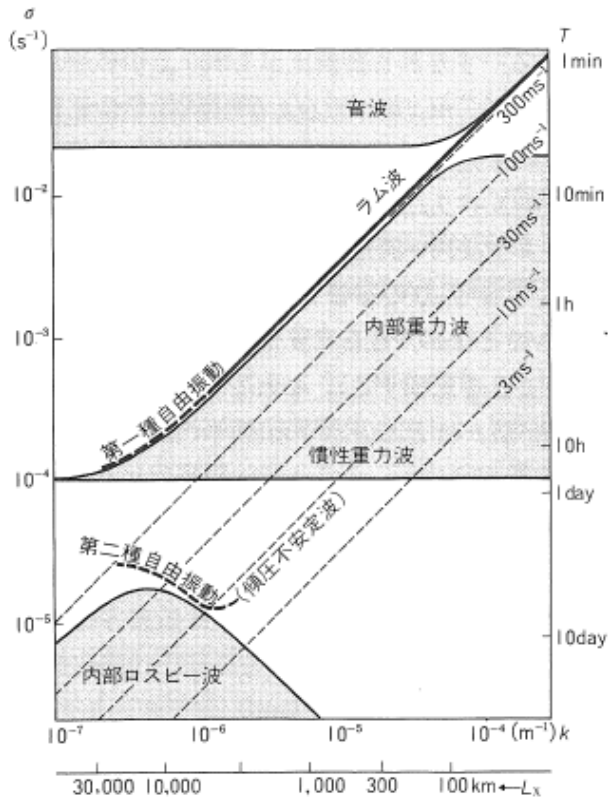
# 解析値の力学的バランス

- 現実の大気や海洋の運動を散逸カオス系と見なせるとすると、外力の急激な変動がない限り、状態変数で張られる相空間において、それらの状態はアトラクターのごく近傍を運動していると考えられる。これを力学的にバランスした状態という。
- それらの推定値である解析値が力学的にバランスしていないと、解析値を初期値にして数値モデルを時間積分したときに、現実には見られない過渡的な運動が発生する。
- 数値モデルが大きなバイアスを持っているためにそのアトラクターが現実のアトラクターからずれていると、解析値の力学的バランスを保証するうえで困難が生じる。



# 初期値化 (Initialization)

数値天気予報は、当初は順圧モデルや準地衡風モデルなど、重力波を含まない数値モデルを用いていたが、1960年代からより精度の高いプリミティブ方程式モデルを採用し始めた。このため、解析値に含まれる重力波ノイズを効果的に取り除く方法が必要とされた。これを初期値化とよぶ。



気象庁領域モデルのある格子点における地上気圧の初期値からの時間変化。非線形ノーマルモード・イニシャリゼーションを行わなかった場合(細実線)と行った場合(太実線:カットオフ周期12時間, 破線:6時間)。横軸と縦軸の単位は、それぞれhourとhPa。(Takano et al., 1990)

# 非線形 Normal Mode Initialization

Machenhauer (1977)

数値モデルの方程式を高周波重力波成分に射影する。

$$\frac{d\mathbf{x}_G}{dt} = \mathbf{L}\mathbf{x}_G(t) + \mathbf{N}(\mathbf{x}_G(t), \mathbf{x}_{\bar{G}}(t), t)$$

ここで、 $\mathbf{x}_G$  と  $\mathbf{x}_{\bar{G}}$  はそれぞれ高周波重力波成分とそれ以外の成分、 $\mathbf{L}$  は正則な定数行列、右辺第2項は非線形項など残りの成分を表す。

高周波重力波の発生を抑えるために、解析値に含まれる  $\mathbf{x}_G(0)$  を次式を満たす  $\mathbf{x}_G^*(0)$  で置き換える。

$$\mathbf{L}\mathbf{x}_G^*(0) + \mathbf{N}(\mathbf{x}_G^*(0), \mathbf{x}_{\bar{G}}(0), 0) = 0$$

この非線形方程式を次のように反復法で解く。

$$\mathbf{x}_G^{*(0)}(0) = \mathbf{0}$$

$$\mathbf{x}_G^{*(k)}(0) = -\mathbf{L}^{-1}\mathbf{N}(\mathbf{x}_G^{*(k-1)}(0), \mathbf{x}_{\bar{G}}(0)) \quad (k = 1, 2, \dots)$$

# デジタルフィルタと IAU

## デジタルフィルタ

Lynch and Huang (1992), Lynch (1997)

デジタルフィルタを利用すると、高周波成分をより効果的に除去できる。

$$\mathbf{x}^*(0) = \sum_{n=-N}^N h_n \mathbf{x}(n\Delta t)$$

$\Delta t$  : 数値モデルのタイムステップ

$h_n$  : 適当なローパスフィルタの重み係数

データ同化では、Lanczos filter やChebyshev filterがよく利用される。

## IAU (Intermittent Analysis Update)

Bloom et al (1996)

観測データによる予測値の修正を、解析時刻の予測値にいきなり加えるのではなく、時間積分しながら少しずつ加えることによって力学的なバランスを保証する。

# ナッジング (Nudging) (1)

Hoke and Anthes (1976)

衛星やレーダーなどによる時間的に連続な観測データを同化する経験的な方法。観測データのある格子点で、観測データに近づけるための強制項を数値モデルに追加して時間積分する。

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(\mathbf{x}, t) \Rightarrow \frac{d\hat{\mathbf{x}}}{dt} = \mathbf{F}(\hat{\mathbf{x}}, t) + \gamma(\mathbf{x}^o - \hat{\mathbf{x}}) \quad (\gamma > 0)$$

数値モデルが線形で1変数の場合に、最適な $\gamma$ の値を求めてみる。

数値モデル:  $\frac{dx}{dt} = \alpha x \quad (\alpha < 0)$

現実システム:  $\frac{dx}{dt} = \alpha x + \varepsilon_t, \quad \langle \varepsilon_t \rangle = 0, \quad \langle \varepsilon_t \varepsilon_{t'} \rangle = Q\delta(t - t')$

観測データ:  $x^o = x + \delta_t, \quad \langle \delta_t \rangle = 0, \quad \langle \delta_t \delta_{t'} \rangle = R\delta(t - t')$

ナッジング:  $\frac{d\hat{x}}{dt} = \alpha \hat{x} + \gamma(x^o - \hat{x}), \quad \hat{x}(0) = x(0) + \eta$

# ナッジング(2)

真値

$$x(t) = x(0)e^{\alpha t} + \int_0^t \varepsilon_\tau e^{\alpha(t-\tau)} d\tau$$

解析値

$$\begin{aligned} \hat{x}(t) &= \hat{x}(0)e^{(\alpha-\gamma)t} + \gamma \int_0^t x^o(\tau) e^{(\alpha-\gamma)(t-\tau)} d\tau \\ &= x(t) + \eta e^{(\alpha-\gamma)t} + \int_0^t (\gamma \delta_\tau - \varepsilon_\tau) e^{(\alpha-\gamma)(t-\tau)} d\tau \end{aligned}$$

解析誤差分散

$$\begin{aligned} \sigma^2(t) &:= \langle (\hat{x}(t) - x(t))^2 \rangle = \frac{\gamma^2 R + Q}{2(\gamma - \alpha)} + \left[ \langle \eta^2 \rangle - \frac{\gamma^2 R + Q}{2(\gamma - \alpha)} \right] e^{-2(\gamma - \alpha)t} \\ &\rightarrow \frac{\gamma^2 R + Q}{2(\gamma - \alpha)} \quad ((\gamma - \alpha)t \rightarrow \infty) \end{aligned}$$

これを最小にする  $\gamma$  の値と、そのときの解析誤差分散

$$\gamma_m = \sqrt{\alpha^2 + \frac{Q}{R}} + \alpha$$

$$\sigma_m^2 = \gamma_m R$$



# ナッジング(3)

連続時間カルマンフィルタ: 線形システムに対する最適推定を与える。

$$\frac{d\mathbf{x}}{dt} = \mathbf{F}(t)\mathbf{x} + \boldsymbol{\varepsilon}_t, \quad \langle \boldsymbol{\varepsilon}_t \rangle = 0, \quad \langle \boldsymbol{\varepsilon}_t (\boldsymbol{\varepsilon}_{t'})^T \rangle = \mathbf{Q}(t)\delta(t-t')$$

$$\mathbf{x}^o = \mathbf{x} + \boldsymbol{\delta}_t, \quad \langle \boldsymbol{\delta}_t \rangle = 0, \quad \langle \boldsymbol{\delta}_t (\boldsymbol{\delta}_{t'})^T \rangle = \mathbf{R}(t)\delta(t-t')$$

解析値と解析誤差共分散行列

$$\frac{d\hat{\mathbf{x}}}{dt} = \mathbf{F}(t)\hat{\mathbf{x}} + \mathbf{P}\mathbf{R}(t)^{-1}(\mathbf{x}^o - \hat{\mathbf{x}})$$

$$\frac{d\mathbf{P}}{dt} = \mathbf{F}(t)\mathbf{P} + \mathbf{P}\mathbf{F}(t)^T + \mathbf{Q}(t) - \mathbf{P}\mathbf{R}(t)^{-1}\mathbf{P}$$

例題の1変数の場合は $\mathbf{F}, \mathbf{Q}, \mathbf{R}$  が定数であることから、定常解では

$$0 = 2\alpha\sigma^2 + Q - \frac{\sigma^4}{R} \quad \therefore \sigma^2 = \left( \sqrt{\alpha^2 + \frac{Q}{R}} + \alpha \right) R$$

最適な  $\gamma$  を用いたナッジングによる解析値と一致する。

多変数の場合には、ナッジングを連続時間カルマンフィルタのように最適化することは一般に困難である。