データ同化の基礎について

同志社大学理工学部
山根　省三

はじめに

データ同化　・・・　観測と数値モデルを組み合わせて、
実際の状態を推定する方法。
数学的には統計的推定論の応用と
見なせる（淡路 ほか、2009）。

気象の分野では、

- 数値天気予報の初期値作成
- 再解析データの作成

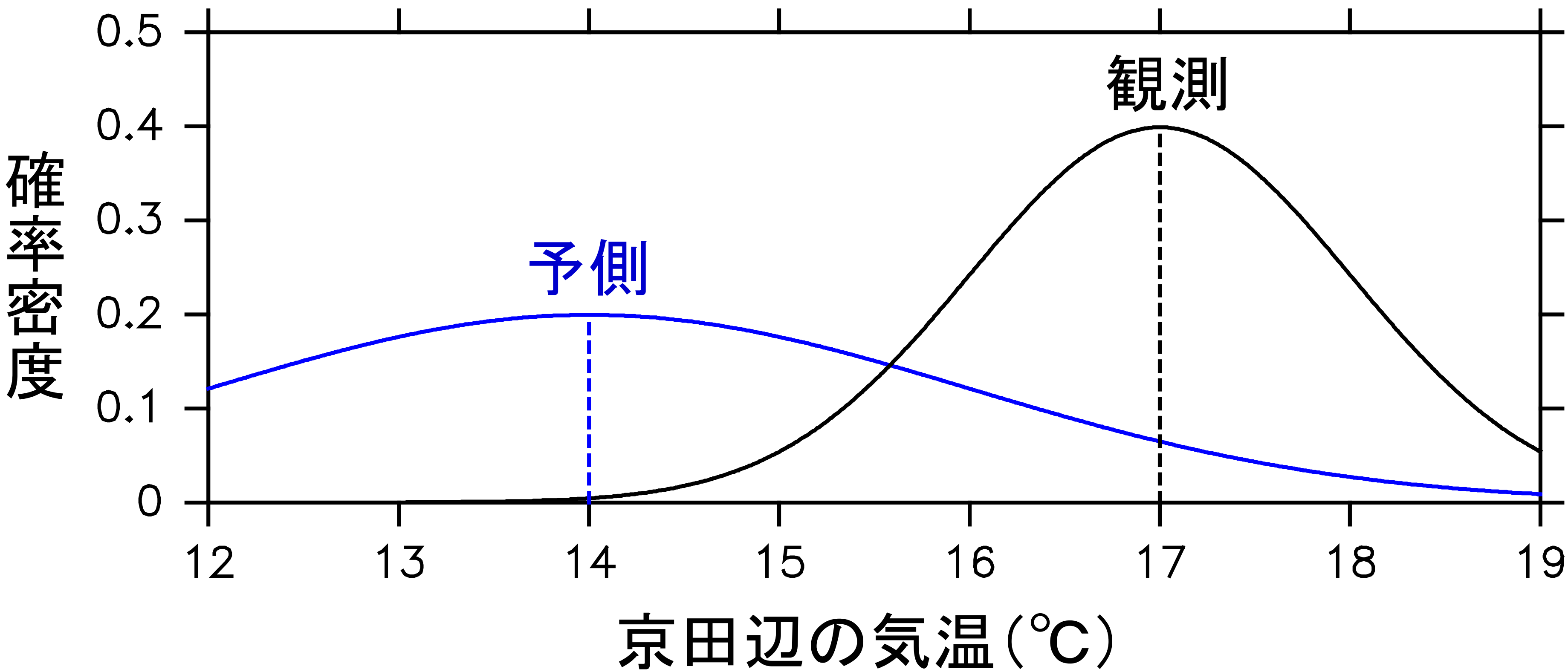
などに用いられる。

データ同化の手法：

最適内挿法、変分法（3次元、4次元）、カルマン・フィルタ

予側値 14℃　予測誤差の標準偏差 2℃
観測値 17℃　観測誤差の標準偏差 1℃

誤差は正規分布に従うものとする。

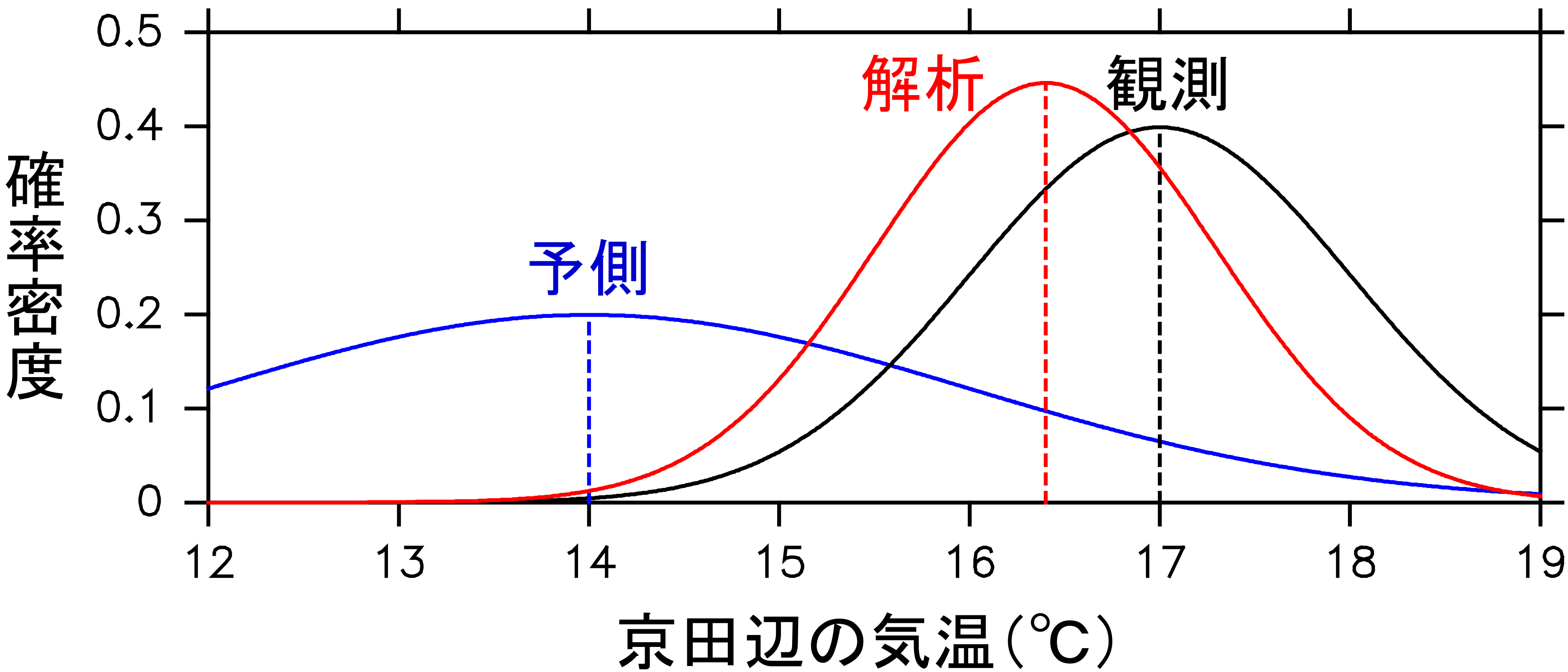


予側値 14℃　予測誤差の標準偏差 2℃

観測値 17℃　観測誤差の標準偏差 1℃

⇩

解析値 16.4℃　解析誤差の標準偏差 $\sqrt{0.8}$ ℃



平均 μ 、標準偏差 σ の正規分布:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

予側値 x_f 　予測誤差の標準偏差 σ_f
観測値 x_o 　観測誤差の標準偏差 σ_o 　とする。

予測と観測の確率密度分布の積が最大となる x が最適な推定値(解析値)となる。

$$\Rightarrow \quad \frac{(x-x_f)^2}{2\sigma_f^{\ 2}} + \frac{(x-x_o)^2}{2\sigma_o^{\ 2}}$$ (指数部分の逆符号のもの)

が最小となる x が最適な推定値(解析値)

変形すると、

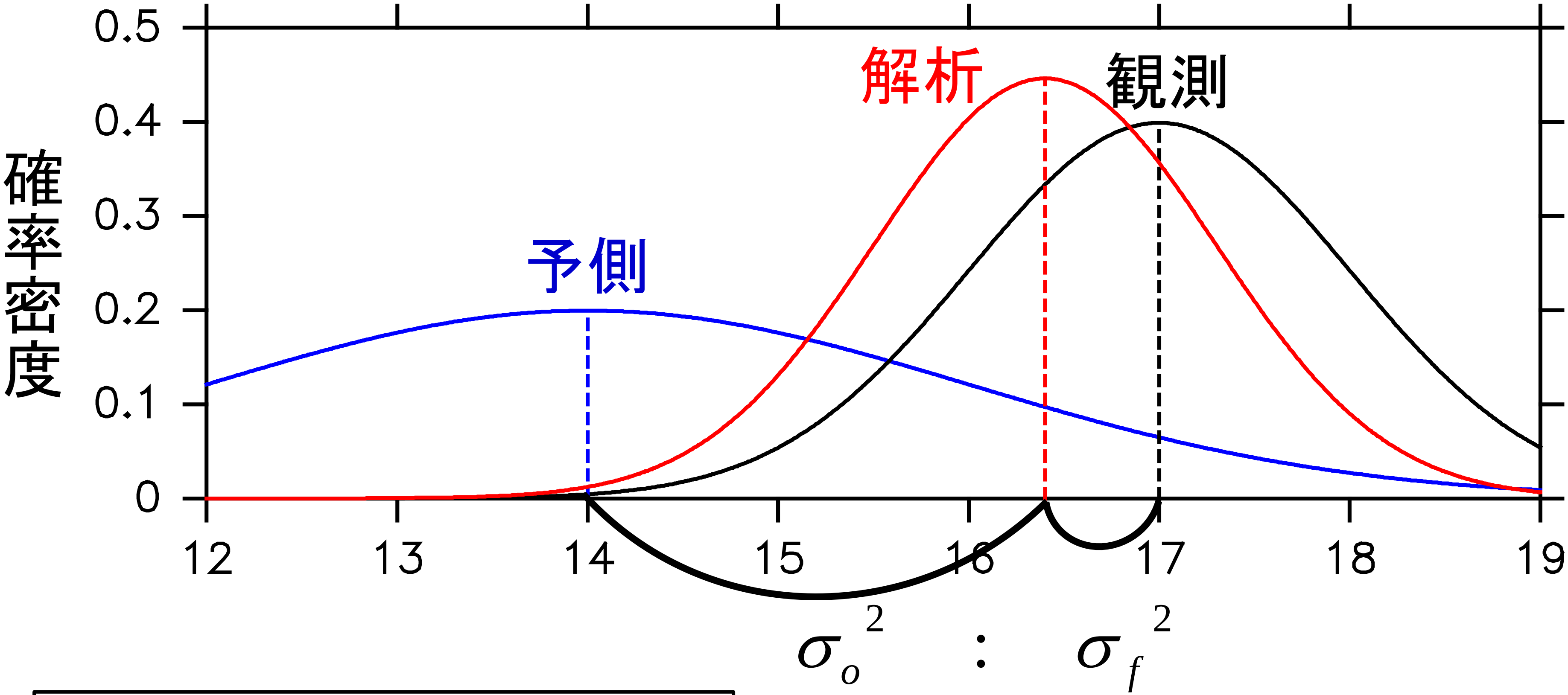
$$\frac{(x-x_f)^2}{2\sigma_f{}^2}+\frac{(x-x_o)^2}{2\sigma_o{}^2}=\frac{\left(x-\dfrac{\sigma_o{}^2x_f+\sigma_f{}^2x_o}{\sigma_f{}^2+\sigma_o{}^2}\right)^2}{2\left(\dfrac{\sigma_f{}^2\sigma_o{}^2}{\sigma_f{}^2+\sigma_o{}^2}\right)}+\frac{\left(x_f-x_o\right)^2}{\sigma_f{}^2+\sigma_o{}^2}$$

なので、最適な推定値（解析値） x_a は、

$$x_a=\frac{\sigma_o{}^2x_f+\sigma_f{}^2x_o}{\sigma_f{}^2+\sigma_o{}^2}=x_f+\frac{\sigma_f{}^2}{\sigma_f{}^2+\sigma_o{}^2}\left(x_o-x_f\right)$$

また、その誤差の標準偏差 σ_aは、

$$\sigma_a{}^2=\frac{\sigma_f{}^2\sigma_o{}^2}{\sigma_f{}^2+\sigma_o{}^2}\quad\Leftrightarrow\quad\sigma_a{}^{-2}=\sigma_f{}^{-2}+\sigma_o{}^{-2}$$



$$\boxed{x_a = x_f + \frac{\sigma_f{}^2}{\sigma_f{}^2 + \sigma_o{}^2}\left(x_o - x_f\right)}$$

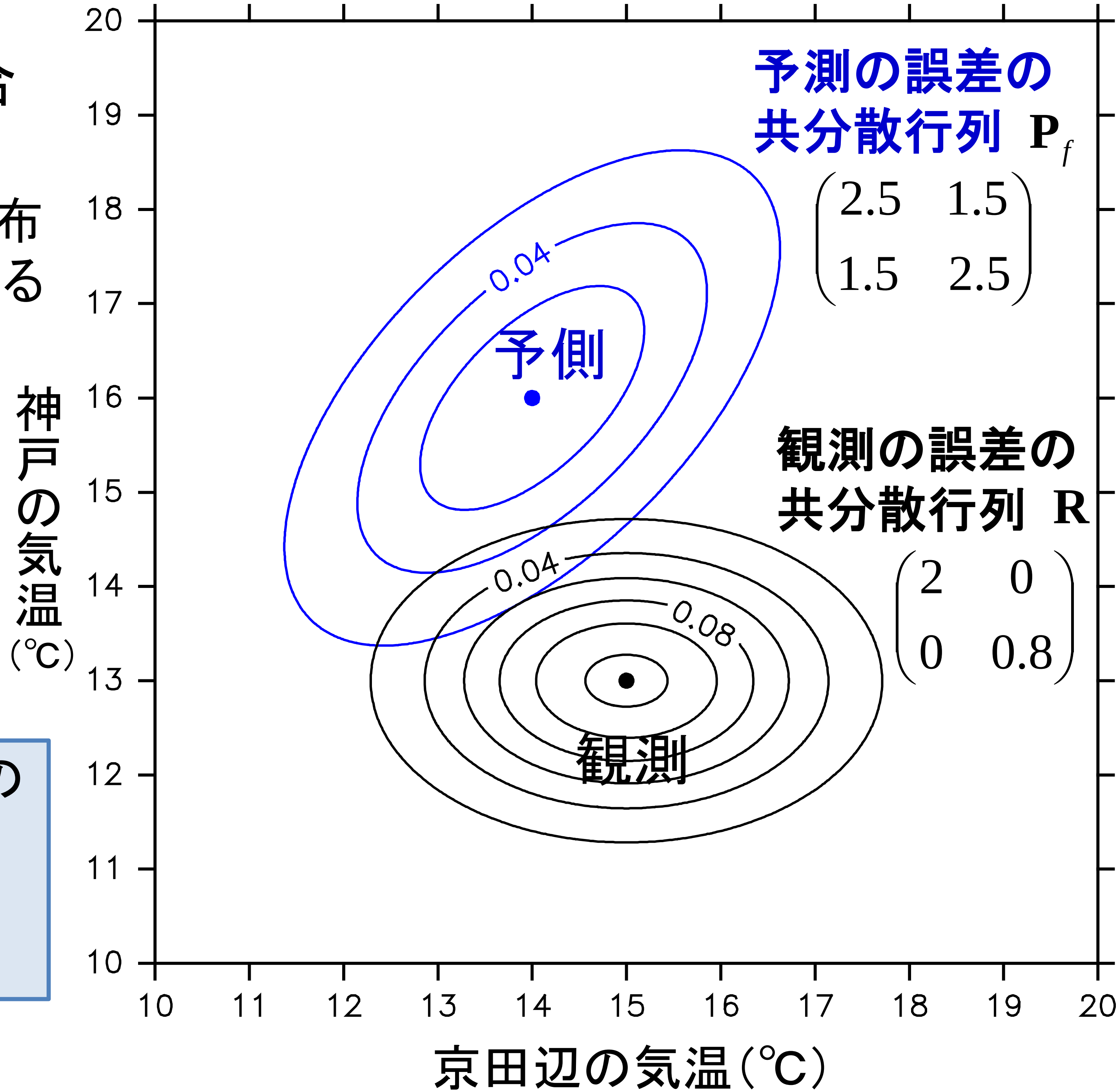
解析値は分散の比で決まる。

$$\boxed{\sigma_a{}^{-2} = \sigma_f{}^{-2} + \sigma_o{}^{-2}}$$

解析値の誤差の分散の逆数は、予測と観測の誤差の分散の逆数の和で与えられる。

2次元の場合

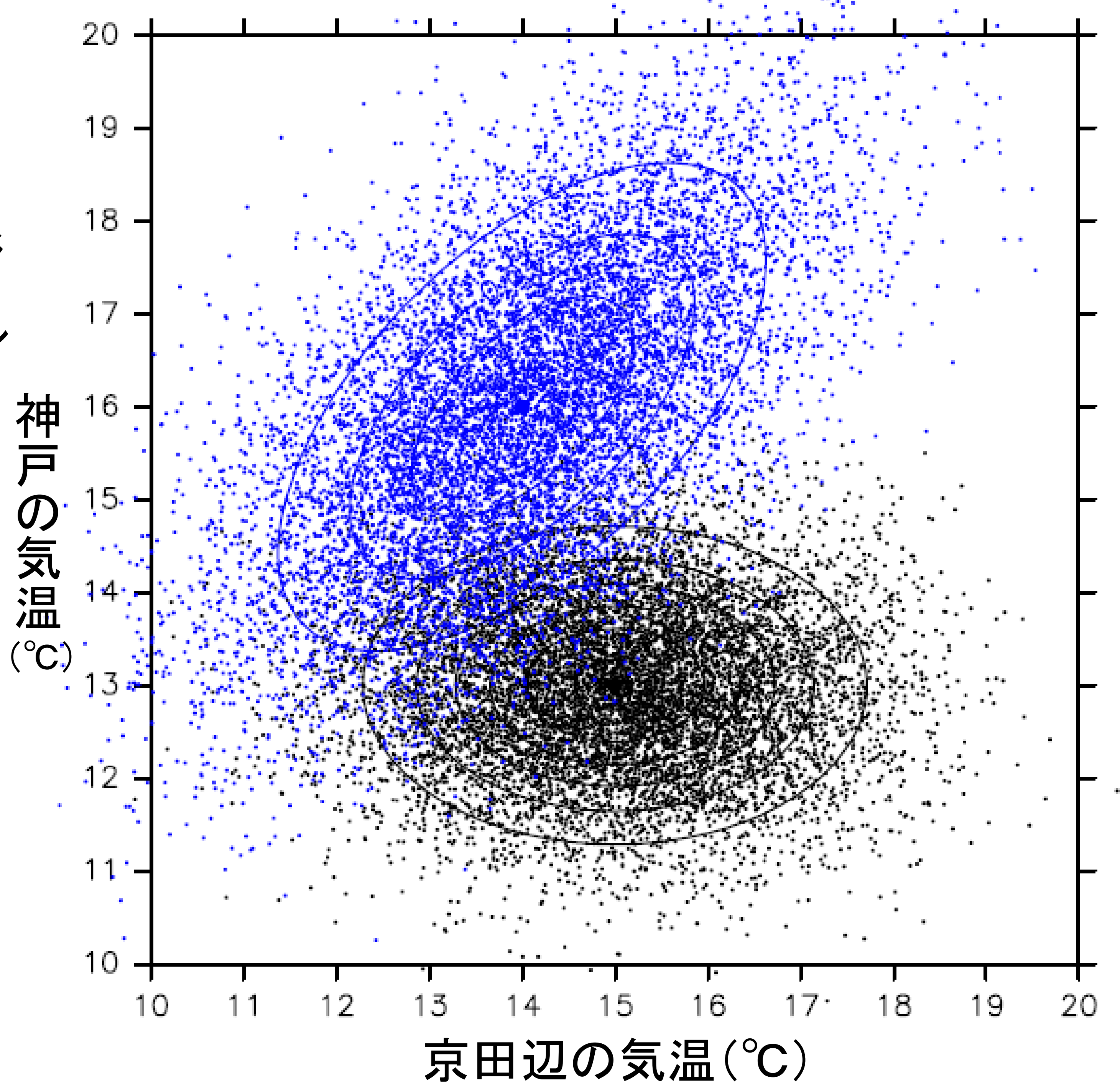
誤差は正規分布に従うものとする

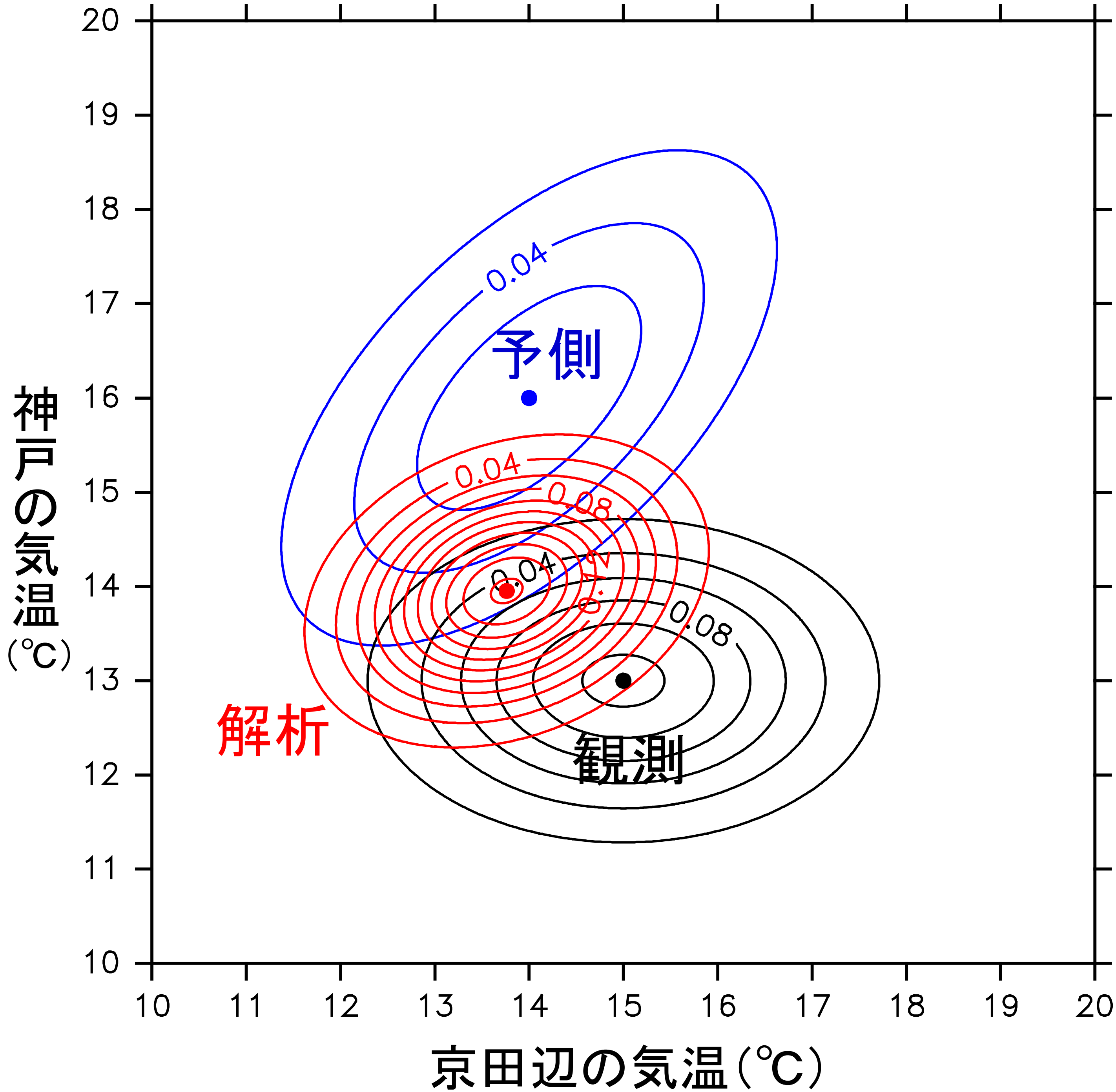


2地点の気温の予測の誤差の間に正の相関がある。

分布のイメージ

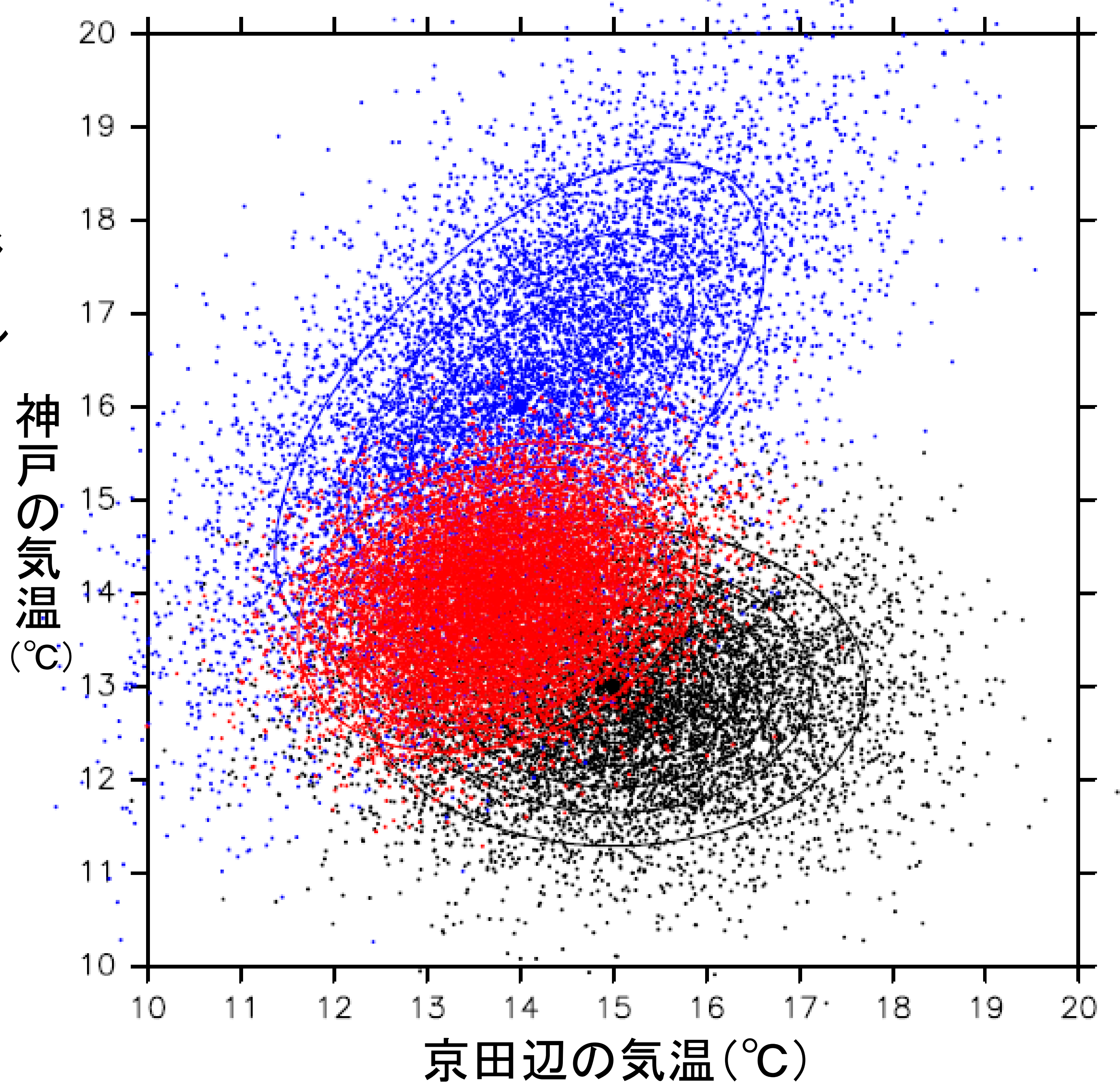
10000 サンプル




予側
解析
観測
0.04
0.04
0.08
0.12
0.04
0.08
神戸の気温(℃)
京田辺の気温(℃)
10
11
12
13
14
15
16
17
18
19
20

分布のイメージ

10000 サンプル



平均 $\boldsymbol{\mu}$ 共分散行列 $\mathbf{S}$ の n 次元正規分布:

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^n \sqrt{|\mathbf{S}|}} \mathrm{e}^{\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{S}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)}$$

予測値 $\mathbf{x}_f$ 予測値の誤差共分散行列 $\mathbf{P}_f$

観測値 $\mathbf{y}$ 観測値の誤差共分散行列 $\mathbf{R}$

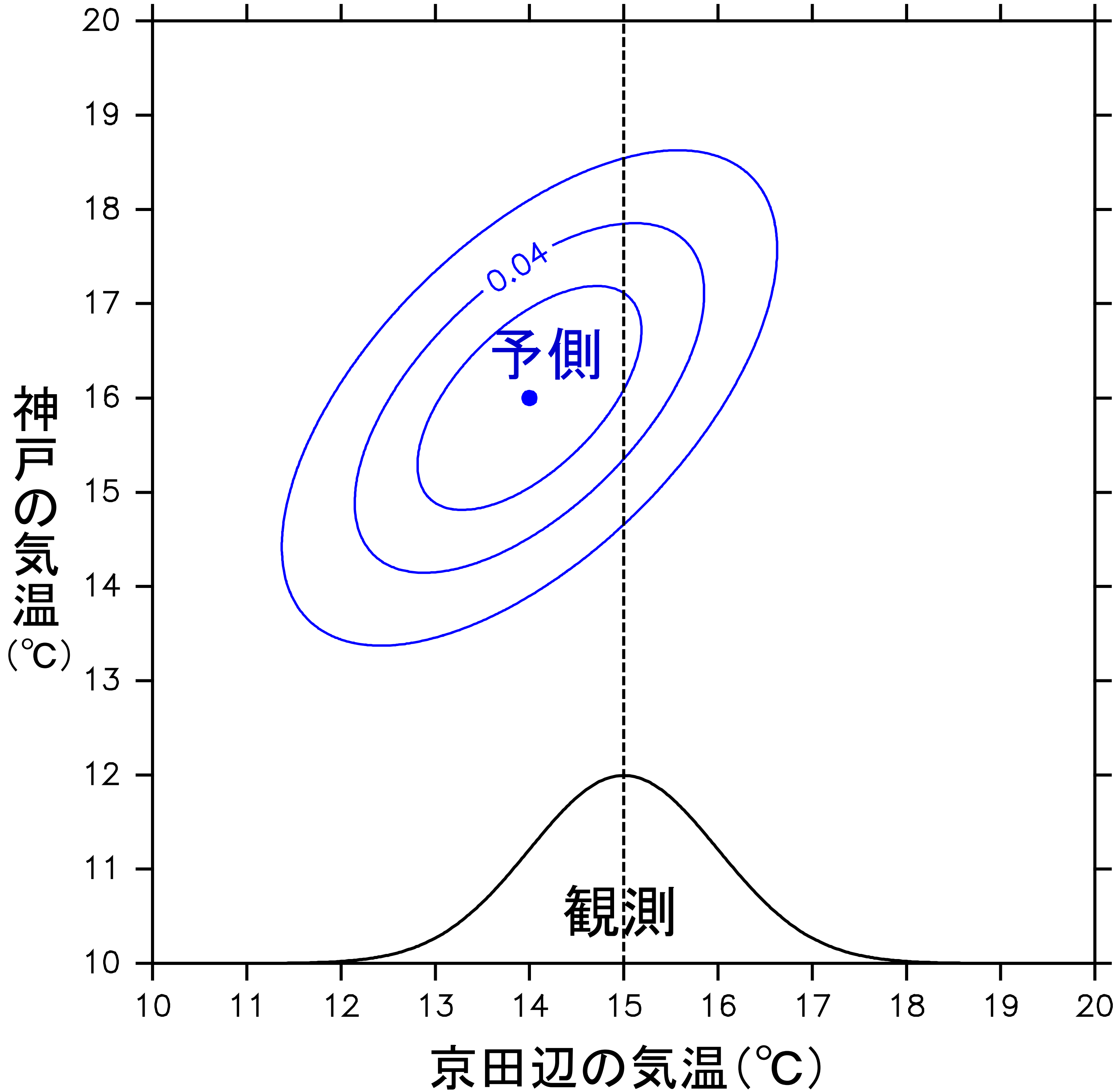
$\Downarrow$

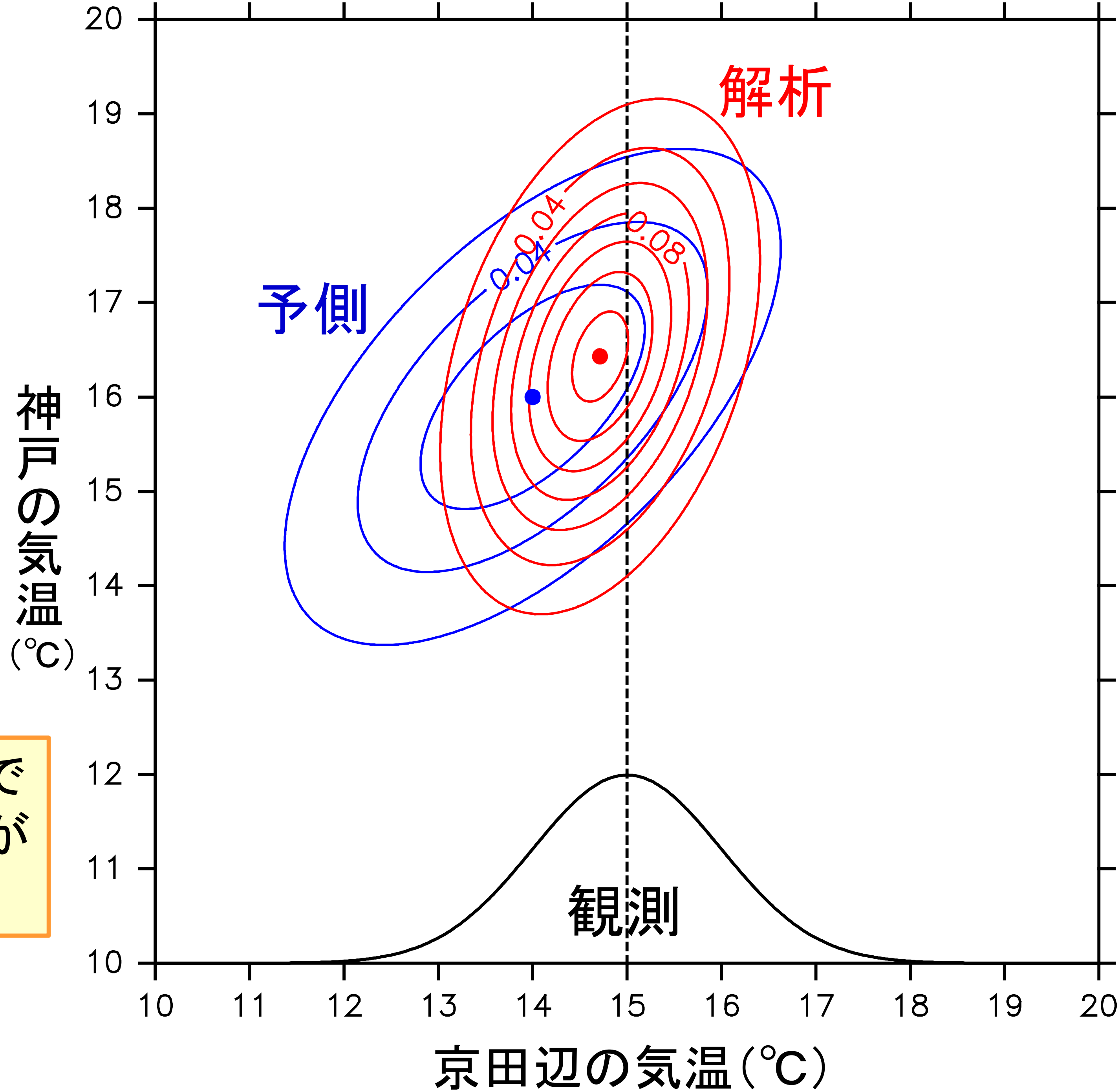
解析値 $\mathbf{x}_a$ 解析値の誤差共分散行列 $\mathbf{P}_a$

$$\mathbf{x}_a = \mathbf{x}_f + \mathbf{P}_f\left(\mathbf{P}_f + \mathbf{R}\right)^{-1}\left(\mathbf{y} - \mathbf{x}_f\right)$$

$$\mathbf{P}_a^{-1} = \mathbf{P}_f^{-1} + \mathbf{R}^{-1}$$

1地点だけで
気温を観測





1地点の観測で
2地点の気温が
修正される。

予測値や解析値を考えている世界と
観測値を考えている世界とが異なる場合は、
観測演算子 $\mathbf{H}$ でこの二つの世界をつなぐ。

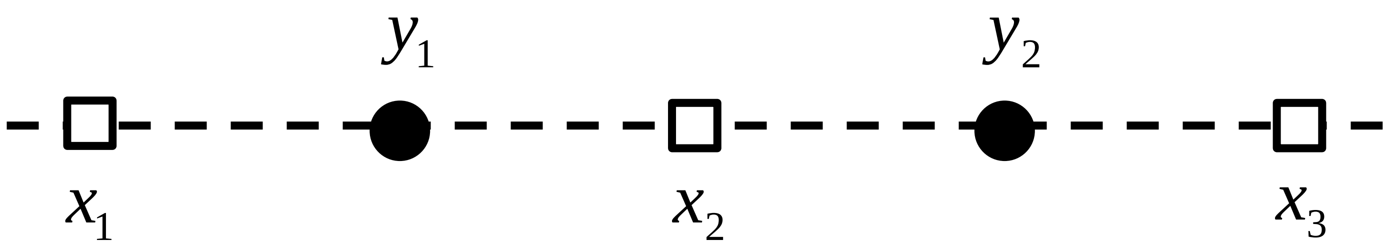
$$\mathbf{H} \quad \mathbf{x} \to \mathbf{y}$$

解析値の次元が n で、観測の数が m 個の場合、
$\mathbf{H}$ は $m \times n$ の行列。
（$\mathbf{H}$ は n 次元から m 次元への写像を表す。）

2地点のうち1地点で観測している場合の $\mathbf{H}$ は、

$$\mathbf{H} = \begin{pmatrix} 1 & 0 \end{pmatrix}$$

である。



（淡路ほか 2009 より）

3地点のうちの中間の2地点で観測した場合の $\mathbf{H}$ は、

$$\mathbf{H} = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \end{pmatrix}$$

となる。

$$\mathbf{H}\ \mathbf{x} = \begin{pmatrix} 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}$$

m 個の観測値の確率分布は、

$$\frac{1}{\left(\sqrt{2\pi}\right)^{m}\sqrt{\,|\mathbf{R}|}}\ \exp\left(-\frac{1}{2}(\mathbf{H}\mathbf{x}-\mathbf{y})^{T}\mathbf{R}^{-1}(\mathbf{H}\mathbf{x}-\mathbf{y})\right)$$

と表せる。よって最小値問題は、

$$J(\mathbf{x})=\frac{1}{2}\left(\mathbf{x}-\mathbf{x}_{f}\right)^{T}\mathbf{P}_{f}^{-1}\left(\mathbf{x}-\mathbf{x}_{f}\right)+\frac{1}{2}\left(\mathbf{H}\quad-\mathbf{y}\right)^{T}\mathbf{R}^{-1}\left(\mathbf{x}\mathbf{H}\quad-\mathbf{y}\right)$$

を最小とする $\mathbf{x}$ について考えることになる。

予測値 $\mathbf{x}_f$ 予測値の誤差共分散行列 $\mathbf{P}_f$

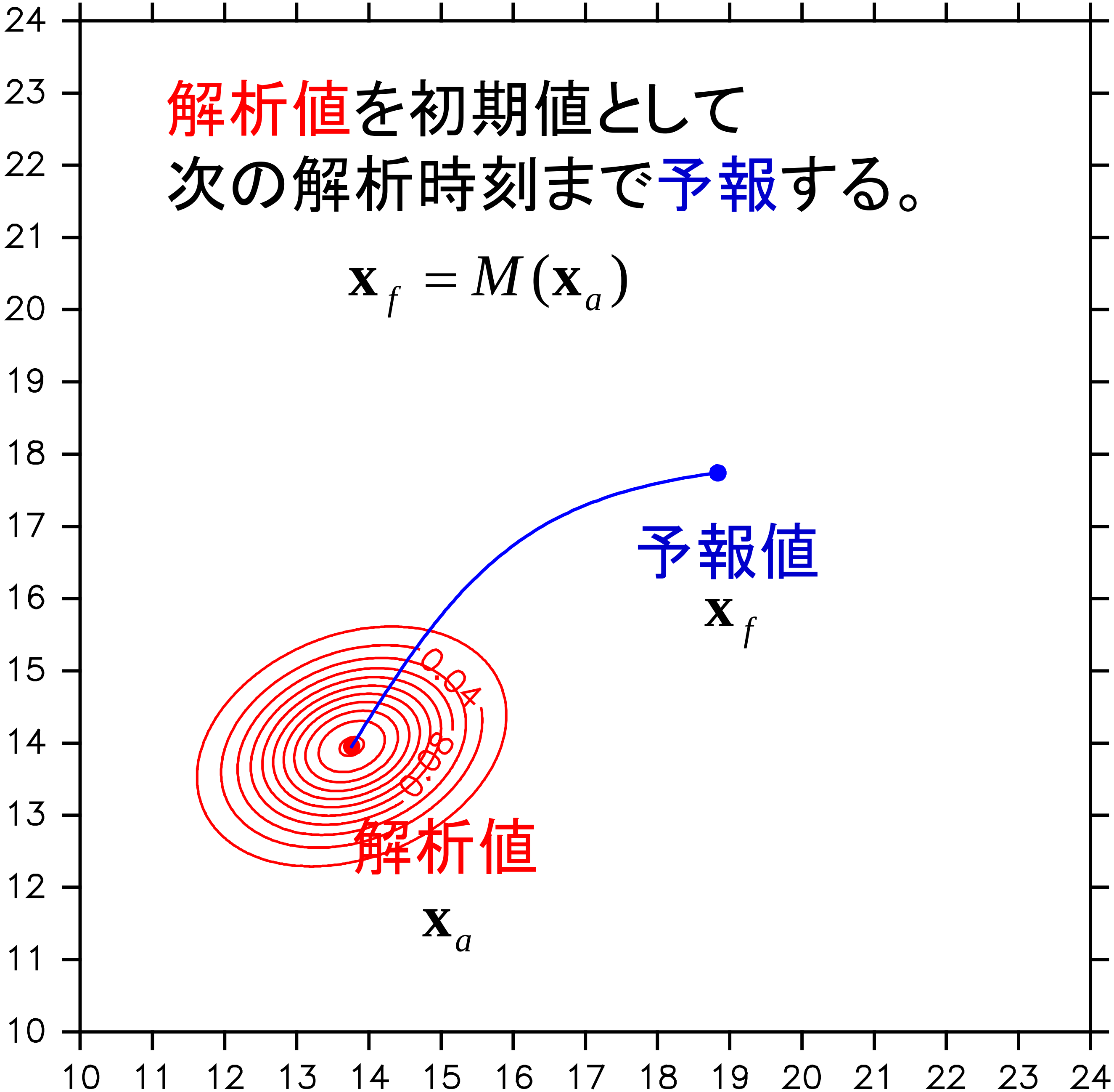
観測値 $\mathbf{y}$ 観測値の誤差共分散行列 $\mathbf{R}$

解析値 $\mathbf{x}_a$ 解析値の誤差共分散行列 $\mathbf{P}_a$

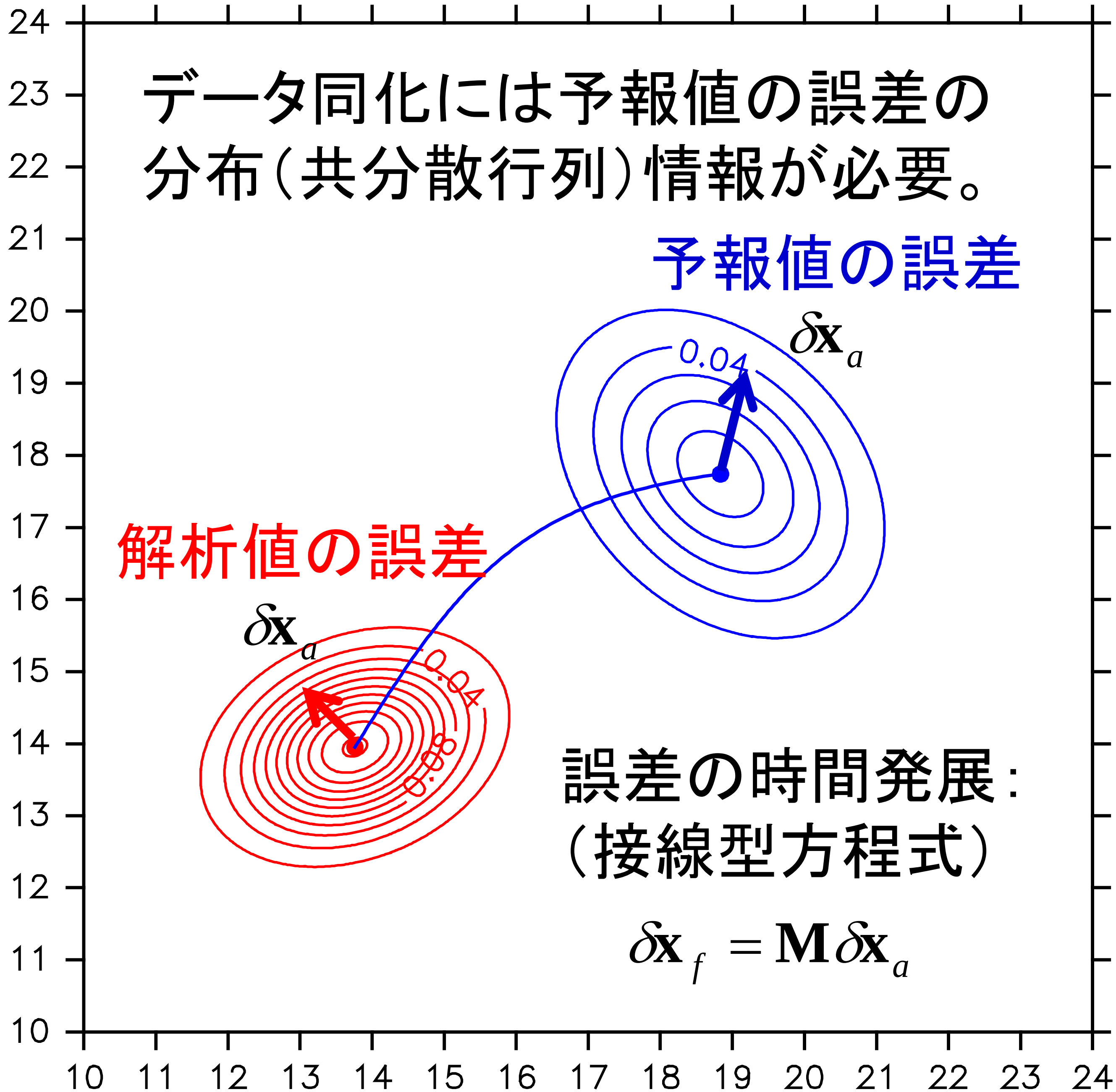
$$\mathbf{x}_a = \mathbf{x}_f + \mathbf{P}_f \mathbf{H}^T \left(\mathbf{H} \quad {}_f \mathbf{H}^T + \mathbf{R} \right)^{-1} \mathbf{P} \left(\mathbf{y} - \mathbf{H} \quad {}_f \right)$$

$$\mathbf{P}_a{}^{-1} = \mathbf{P}_f{}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}$$

数値予報の初期値作成の場合のデータ同化のサイクル



解析値を初期値として
次の解析時刻まで予報する。
$\mathbf{x}_f = M(\mathbf{x}_a)$
予報値
$\mathbf{x}_f$
解析値
$\mathbf{x}_a$
0.04
0.08
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24


データ同化には予報値の誤差の
分布（共分散行列）情報が必要。
予報値の誤差
$\delta \mathbf{x}_a$
0.04
解析値の誤差
$\delta \mathbf{x}_a$
0.04
0.08
誤差の時間発展:
（接線型方程式）
$\delta \mathbf{x}_f = \mathbf{M} \delta \mathbf{x}_a$
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

誤差の時間発展を線形と仮定すれば、

$$\delta\mathbf{x}_f = \mathbf{M}\delta\mathbf{x}_a$$ （接線形方程式）

解析誤差共分散行列は、

$$\mathbf{P}_a = \left\langle \delta\mathbf{x}_a \ \delta\mathbf{x}_a{}^T \right\rangle$$ （$\langle \quad \rangle$ は期待値を表す）

なので、次の時刻の予報誤差共分散行列は、

$$\mathbf{P}_f = \left\langle \delta\mathbf{x}_f \ \delta\mathbf{x}_f{}^T \right\rangle = \mathbf{M}\left\langle \delta\mathbf{x}_a \ \delta\mathbf{x}_a{}^T \right\rangle \mathbf{M}^T = \mathbf{M} \quad {}_a\mathbf{M}^T$$

から求めることができる。

実際には誤差の発展は線形ではなく、
また、予報モデルは完全ではないので、
予報誤差共分散行列 $\mathbf{P}_f$ は、

$$\mathbf{P}_f = \mathbf{M}\mathbf{P}_a\mathbf{M}^T + \mathbf{Q}$$

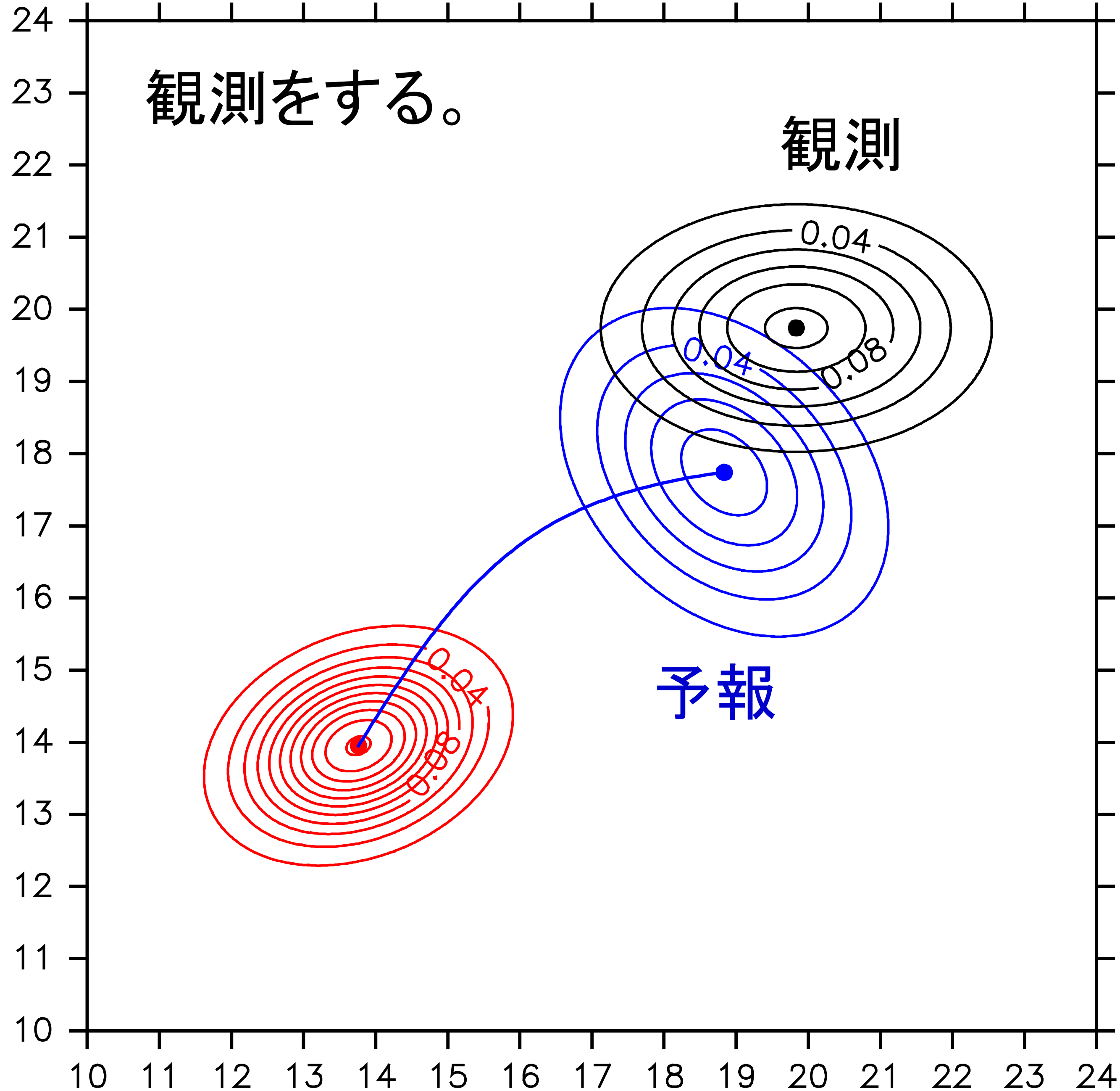
と記述される。 $\mathbf{Q}$ はよく分からない量である。

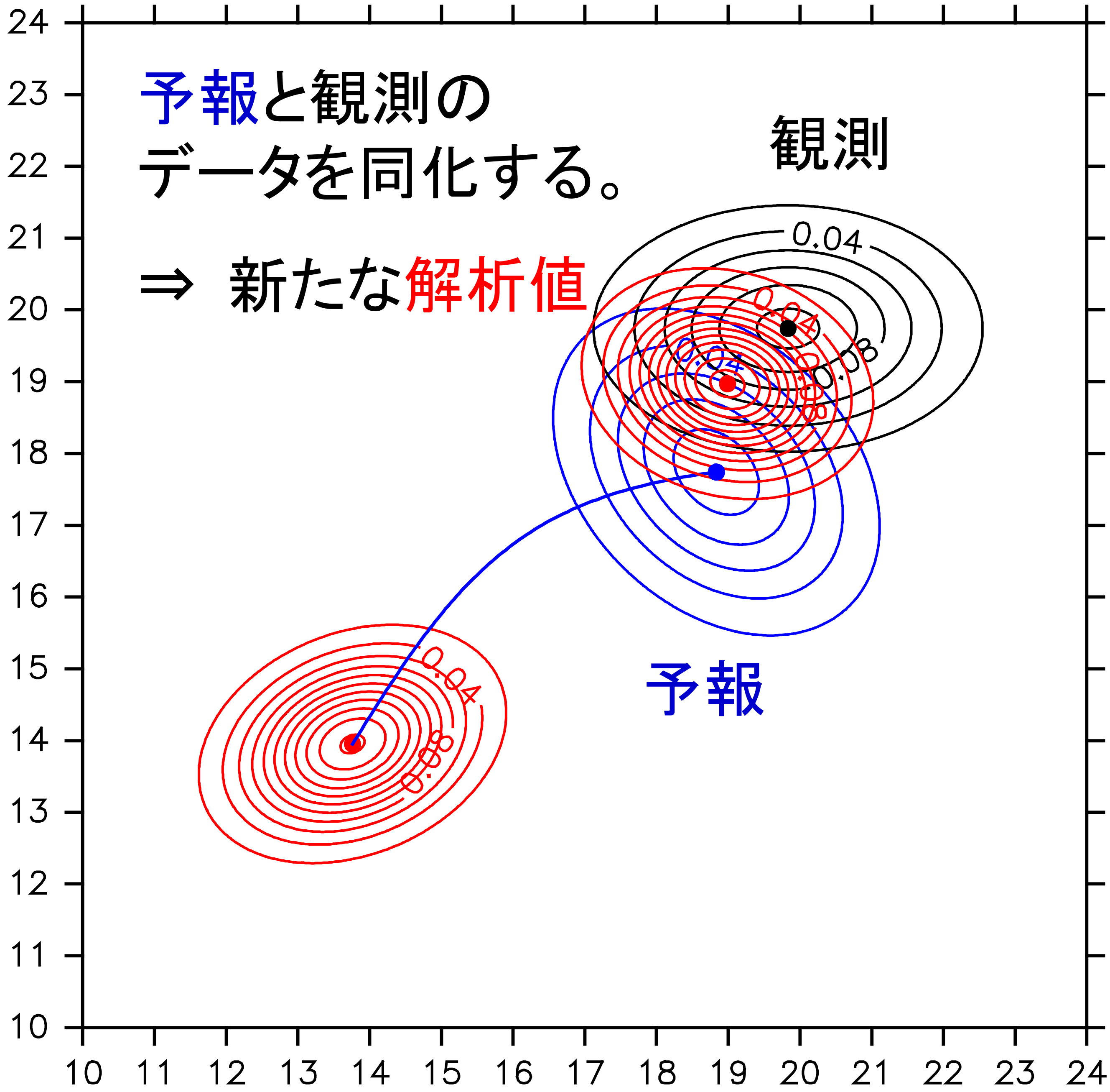
上式の代わりに、

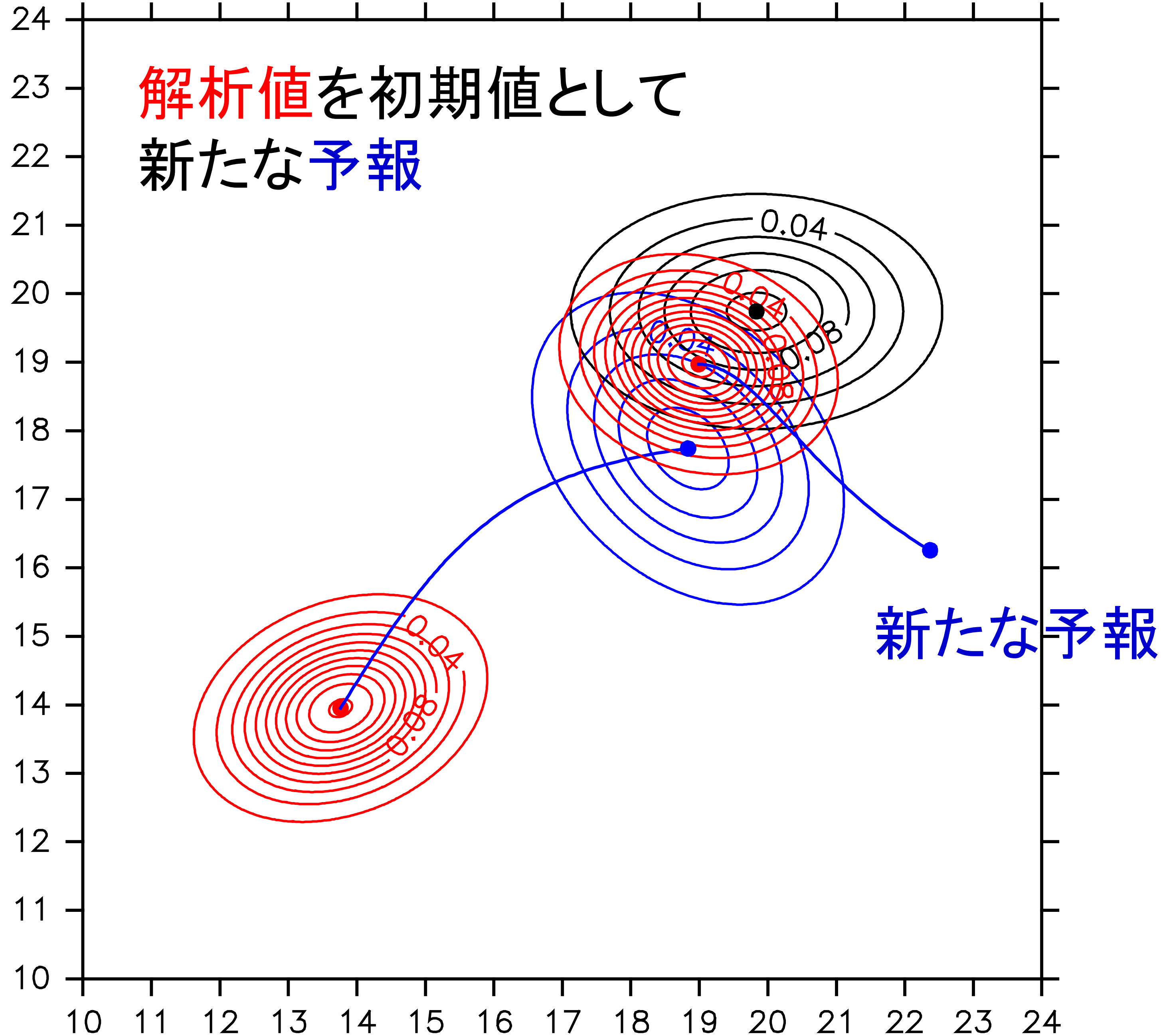
$$\mathbf{P}_f = \mathbf{M}\mathbf{P}_a\mathbf{M}^T(1+\alpha)$$

（$\alpha > 0$: イ　　　　　　　　ン

が用いられたりする。


観測をする。
観測
0.04
0.08
0.04
予報
0.04
0.08
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24


予報と観測の
データを同化する。
⇒ 新たな解析値
観測
予報
0.04
0.08


解析値を初期値として
新たな予報
新たな予報
0.04
0.08
0.04
0.08
0.04
0.04
0.08
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

データ同化には、共分散行列の情報が必要であるが、数値予報のような大きな次元では取り扱いが難しい。

最適内挿法、3次元変分法

・・・　予報誤差共分散行列 $\mathbf{P}_f$ は時間変化しないとする。

カルマン・フィルタ、4次元変分法

・・・　$\mathbf{P}_f$ の時間変化を取り扱う。

データ同化実験の例

Lorenz(1996)モデルを用いて

Lorenz(1996)モデル
（Lorenz and Emanuel　1998）

40変数のモデル

$$\frac{dX_j}{dt} = \left(X_{j+1} - X_{j-2}\right) X_{j-1} - X_j + F \qquad j = 1, 2, \ldots 4.$$

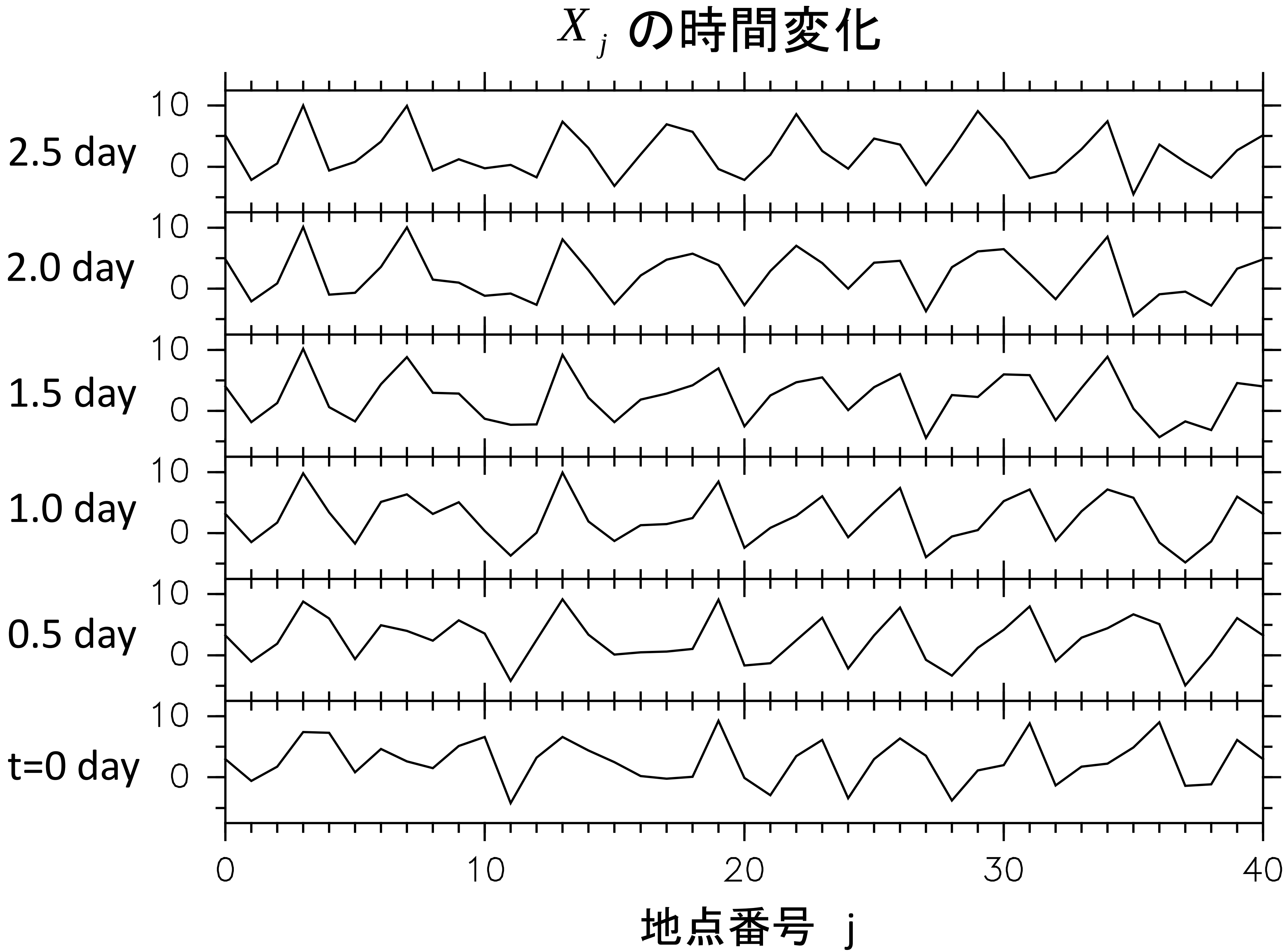
（サイクリック）

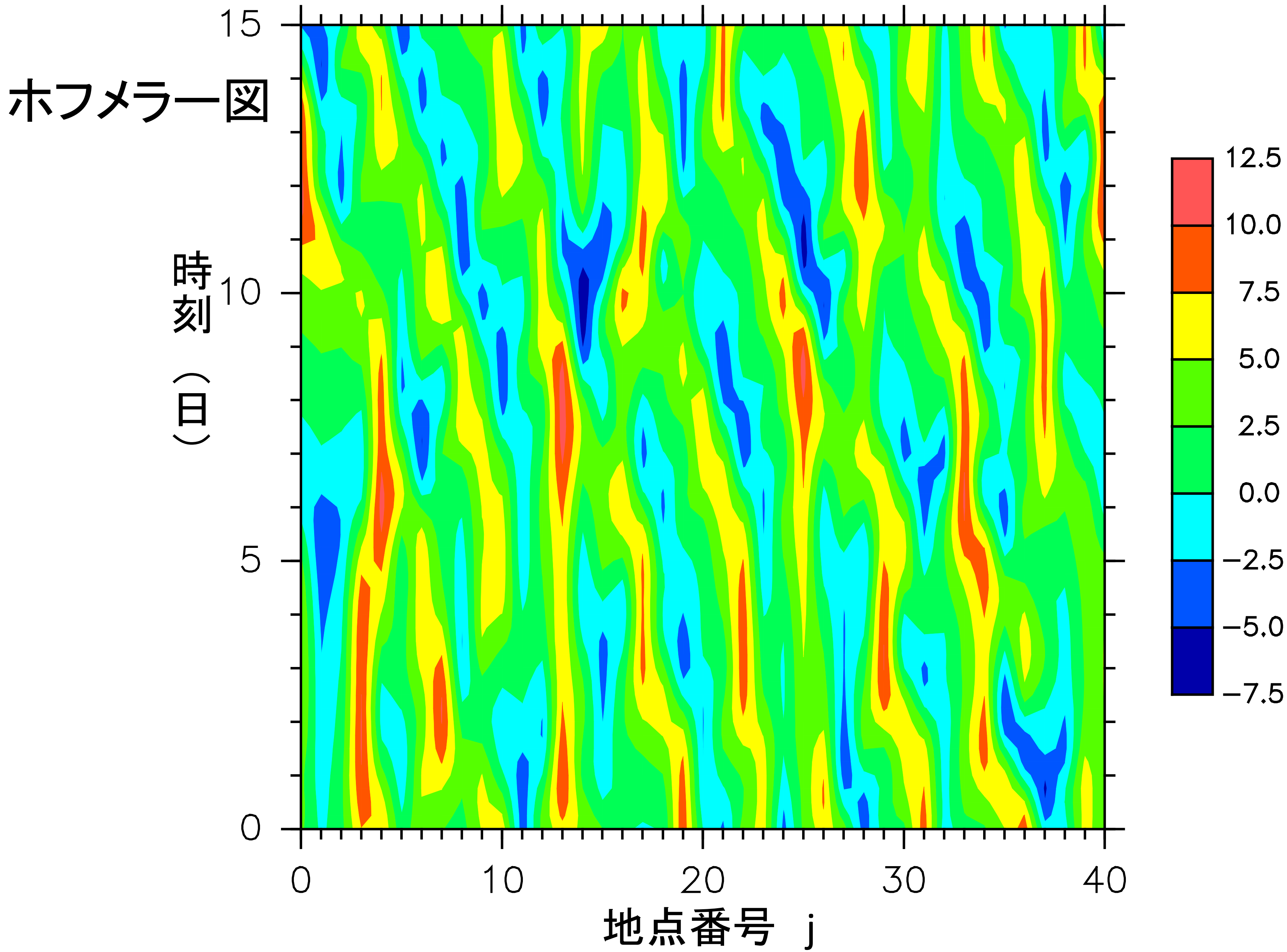
移流　　　散逸　強制

$t = 1$　を5日と考える。時間刻み幅 6時間。

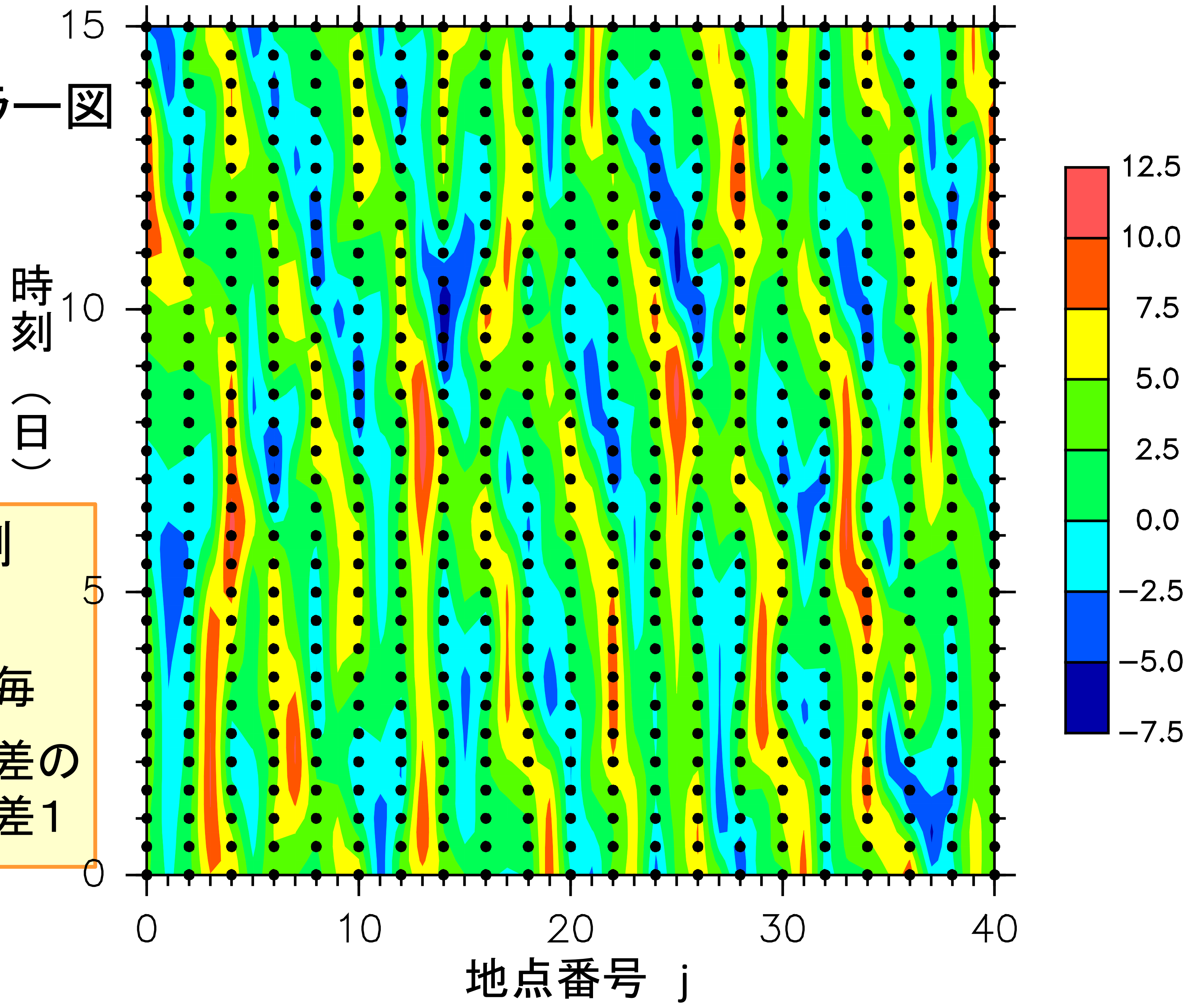
$F = 8$　のとき

- カオス解（約2.1日で摂動の大きさ2倍になる）
- 長時間積分の平均 2.3、標準偏差 3.6


X_j の時間変化
2.5 day
2.0 day
1.5 day
1.0 day
0.5 day
t=0 day
10
0
0
10
20
30
40
地点番号 j


ホフメラー図
時刻（日）
地点番号 j
0
5
10
15
0
10
20
30
40
12.5
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5

ホフメラー図



● 観測

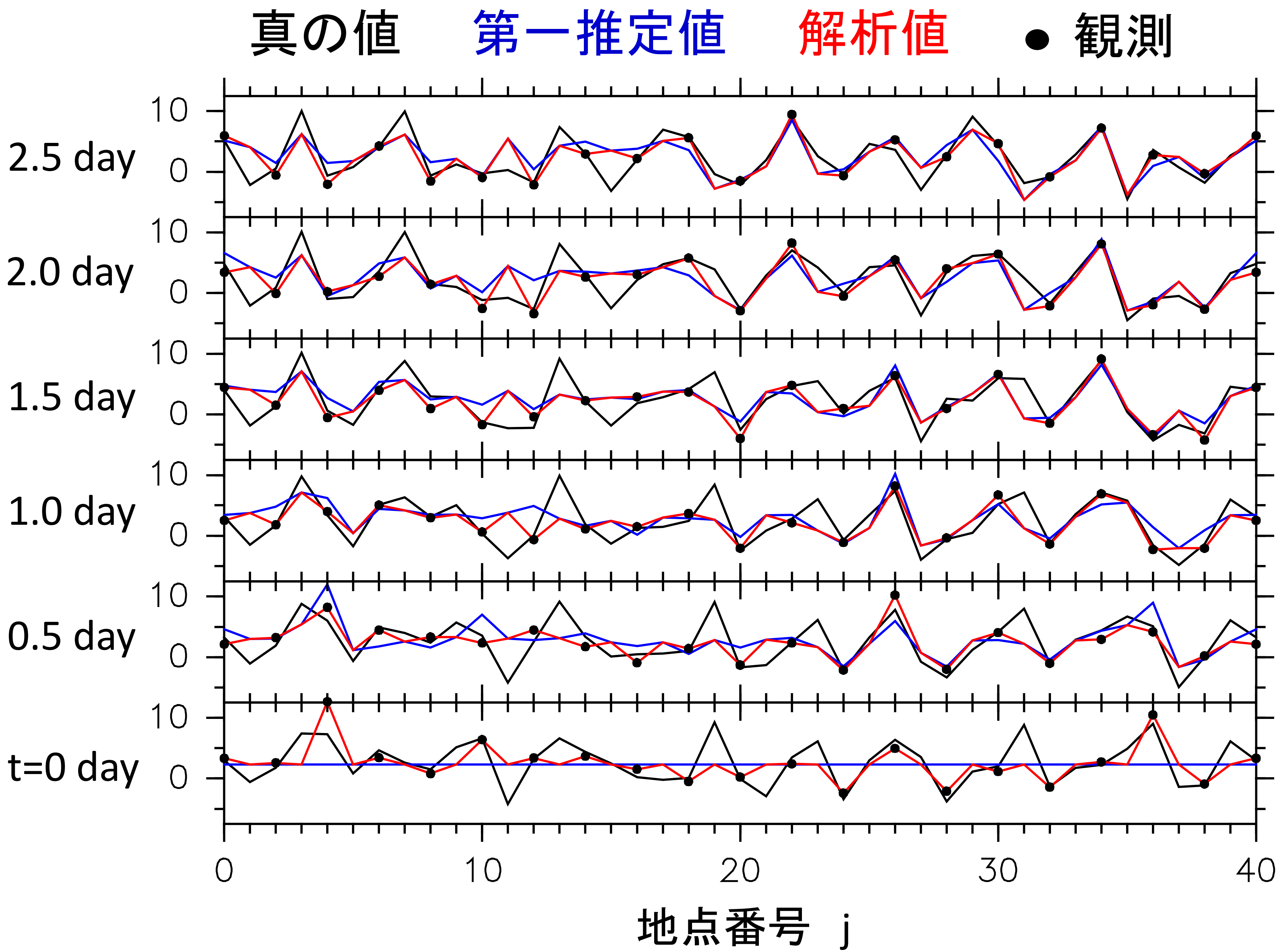
20地点

12時間毎

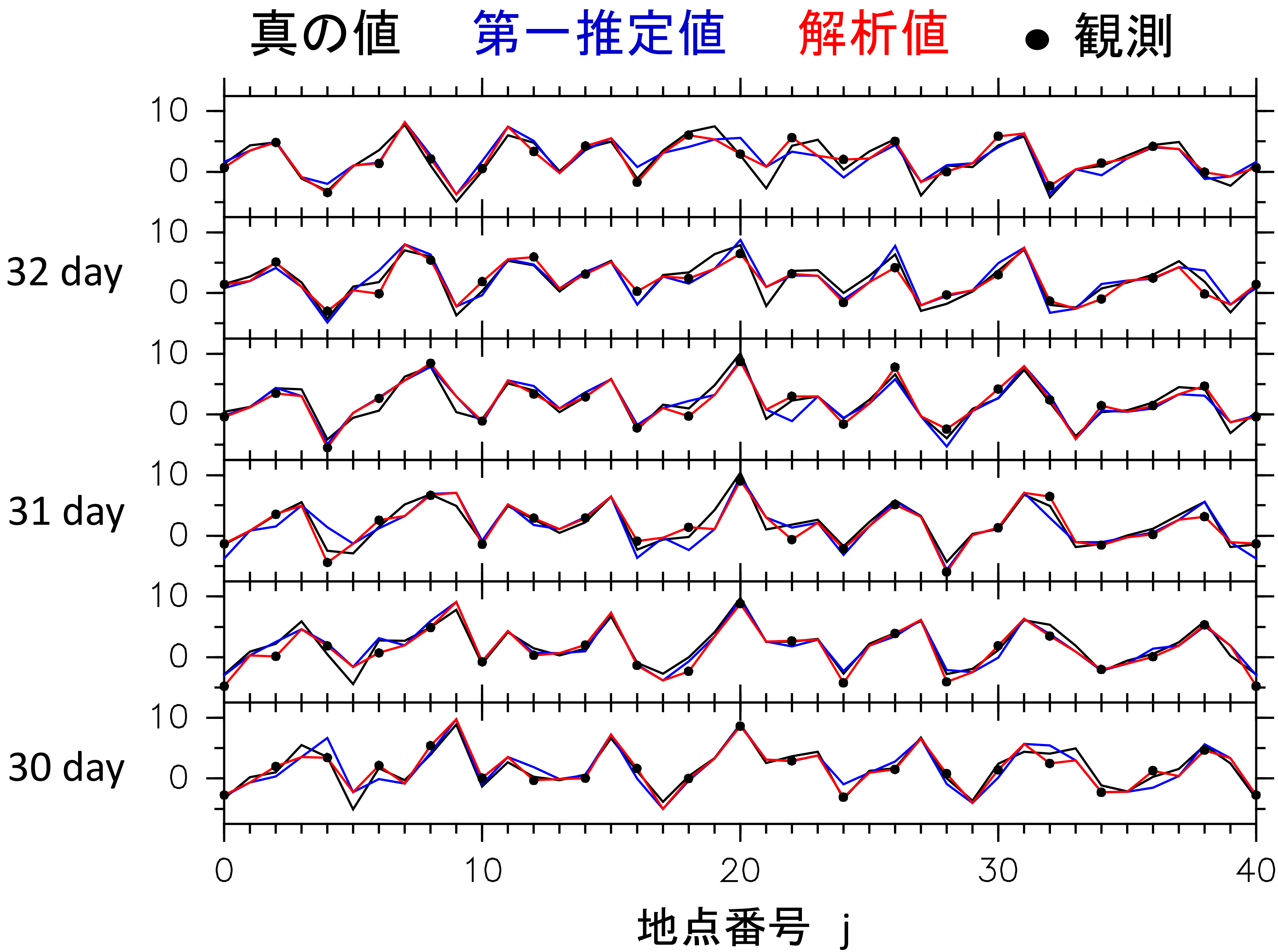
観測誤差の標準偏差1

実験1：　データ同化なし

観測のあった地点の予測値を観測値に置き換えて、
それを解析値とする。

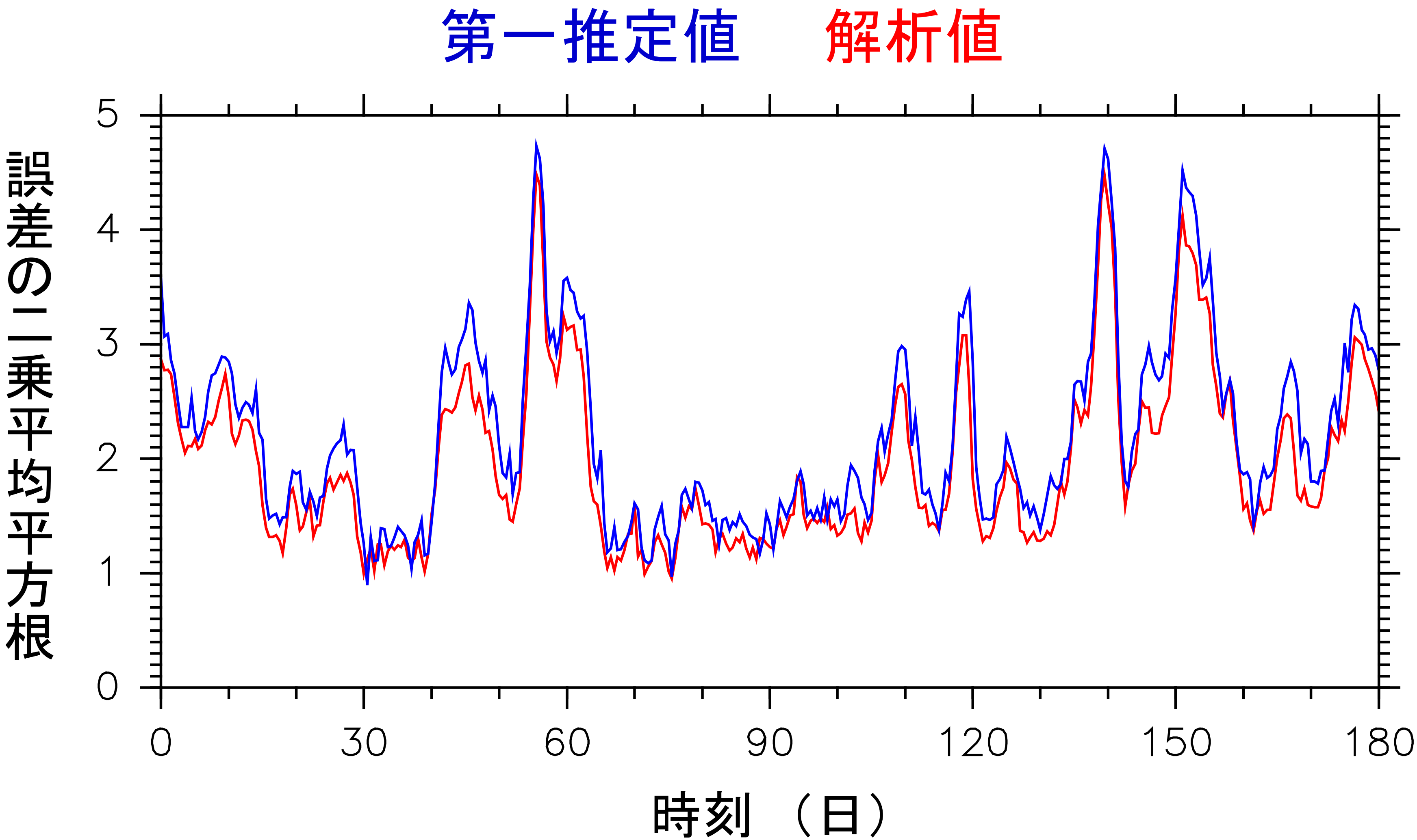


真の値
第一推定値
解析値
● 観測
2.5 day
2.0 day
1.5 day
1.0 day
0.5 day
t=0 day
10
0
0
10
20
30
40
地点番号 j



真の値
第一推定値
解析値
観測
32 day
31 day
30 day
10
0
0
10
20
30
40
地点番号 j

誤差(二乗平均平方根 RMSE)の時間変化

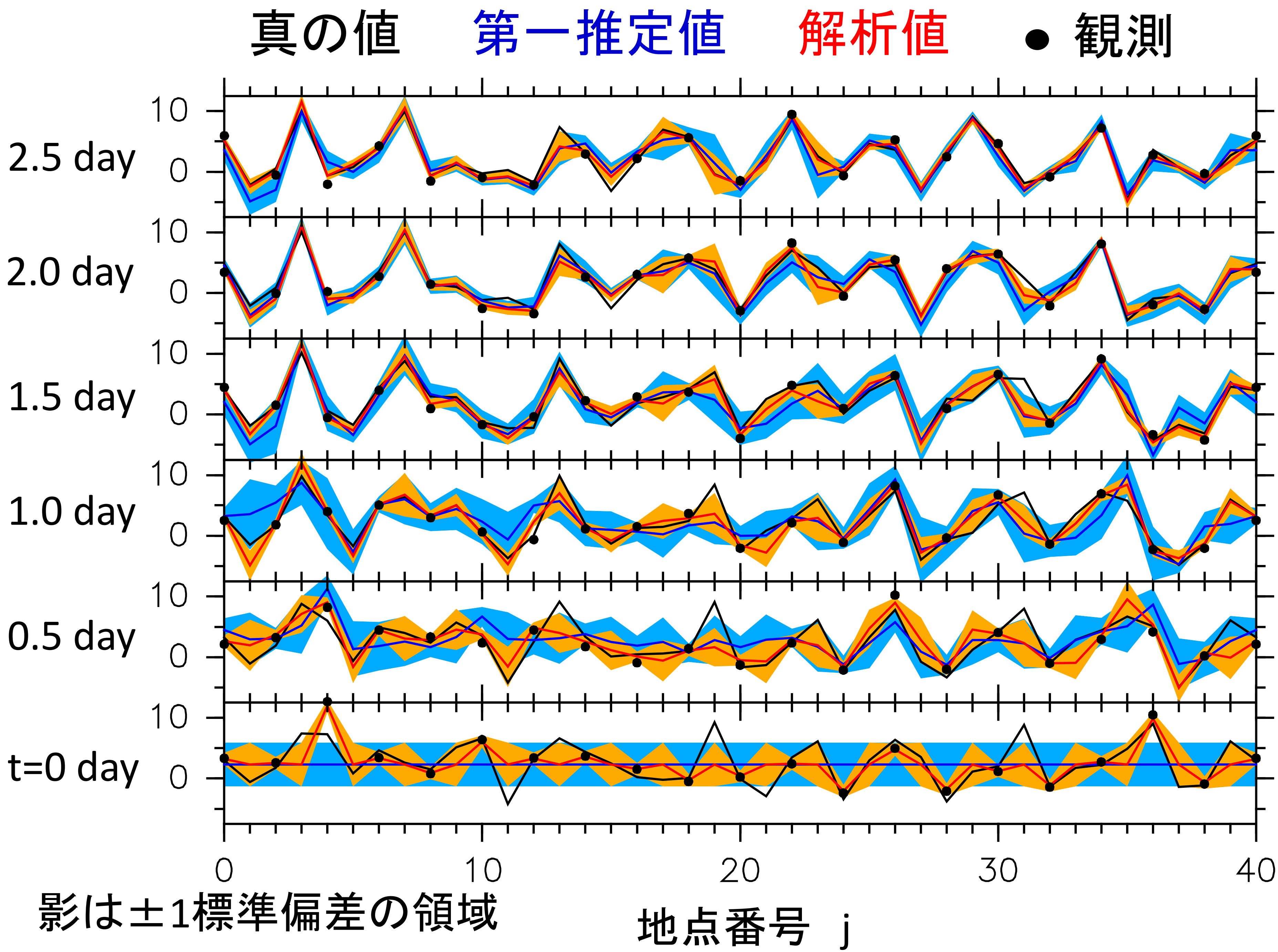


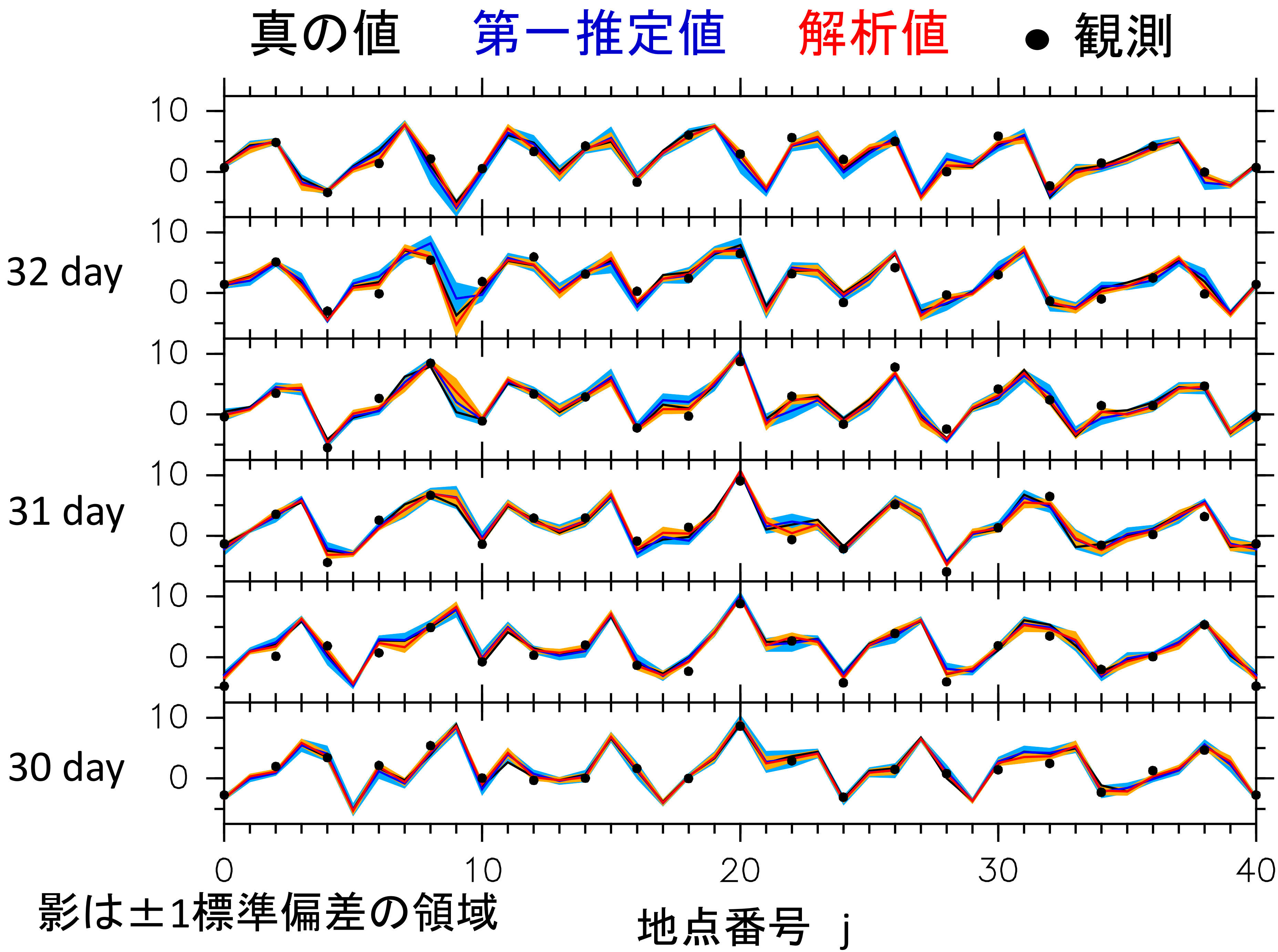
実験2：　データ同化（カルマン・フィルタ）

最初の第一推定値　・・・　気候学的平均 $X_j = 2.3$

最初の誤差共分散行列　・・・　分散 3.6^2 の対角行列

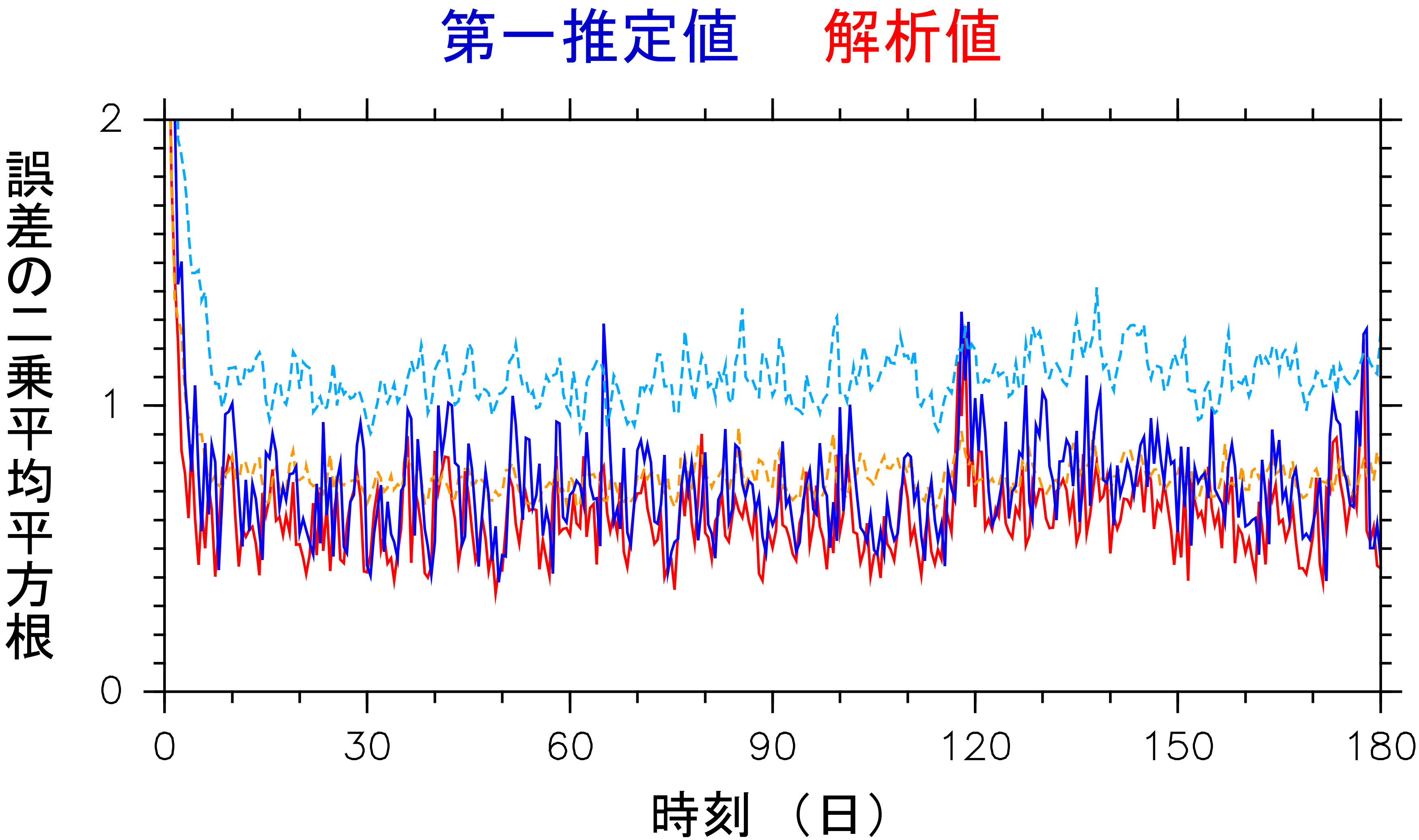
インフレーション： $\alpha = 0.5$


真の値
第一推定値
解析値
観測
2.5 day
2.0 day
1.5 day
1.0 day
0.5 day
t=0 day
10
0
0
10
20
30
40
影は±1標準偏差の領域
地点番号 j



真の値
第一推定値
解析値
● 観測
32 day
31 day
30 day
10
0
0
10
20
30
40
影は±1標準偏差の領域
地点番号 j

誤差(二乗平均平方根 RMSE)の時間変化



第一推定値の誤差共分散行列の時間変化