

Assimilation Algorithms

Lecture 1: Basic Concepts

Sébastien Massart

ECMWF

22 May 2023

A Covid story



Forehead	37.5°C
Armpit L	36.0°C

A Covid story



Forehead	37.5°C
Armpit L	36.0°C
	36.3°C

A Covid story



Forehead	37.5°C
Armpit L	36.0°C
	36.3°C
Armpit R	36.2°C
	36.2°C

A Covid story



Forehead	37.5°C	A priori
Armpit L	36.0°C	Observation bias
	36.3°C	Observation error
Armpit R	36.2°C	Obs. operator error
	36.2°C	

Outline

- 1 History and Terminology
- 2 Elementary Statistics — The Scalar Analysis Problem
- 3 Extension to Multiple Dimensions
- 4 Optimal Interpolation
- 5 Summary

Outline

- 1 History and Terminology
- 2 Elementary Statistics — The Scalar Analysis Problem
- 3 Extension to Multiple Dimensions
- 4 Optimal Interpolation
- 5 Summary

Interpreting the weather situation

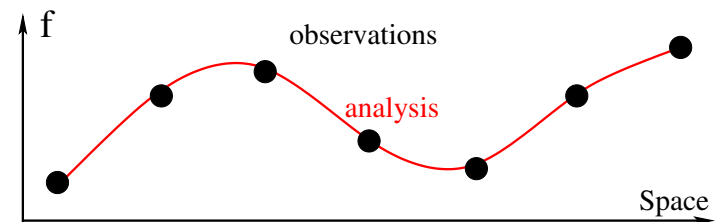
Definition

Analysis: The process of approximating the true state of a (geo-)physical system at a given time using the available knowledge.

- ✘ First hand analysis of synoptic observations in 1850 by LeVerrier and Fitzroy.



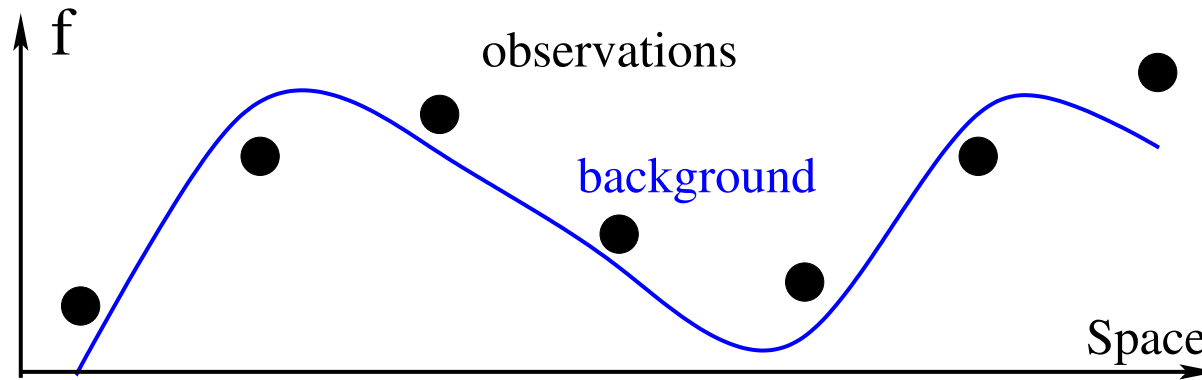
- ✘ Polynomial Interpolation in the 1950s by Panofsky with the developments of computers



The black dots denote the data points, while the red curve shows the polynomial interpolation.

Background

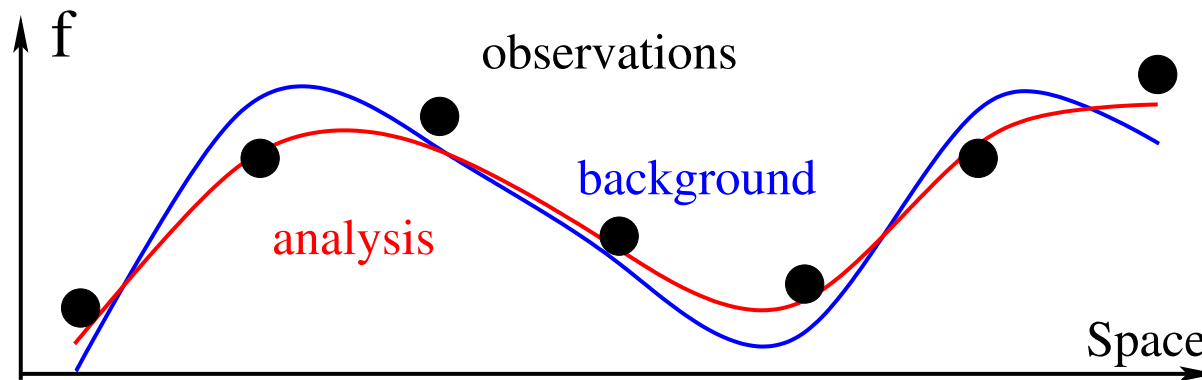
- ✘ An important step forward was made by Gilchrist and Cressman (1954), who introduced the idea of using a previous numerical forecast to provide a preliminary estimate of the analysis.



- ✘ This prior estimate was called the **background**.

Background

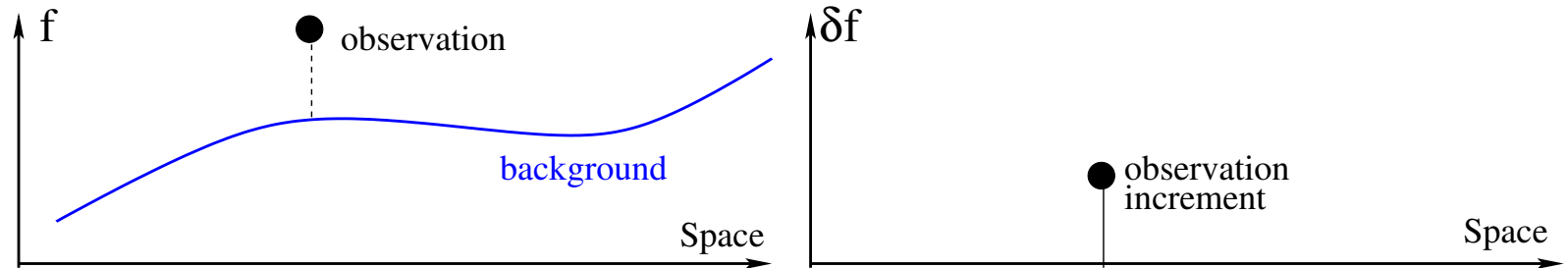
- ✘ An important step forward was made by Gilchrist and Cressman (1954), who introduced the idea of using a previous numerical forecast to provide a preliminary estimate of the analysis.



- ✘ This prior estimate was called the **background**.

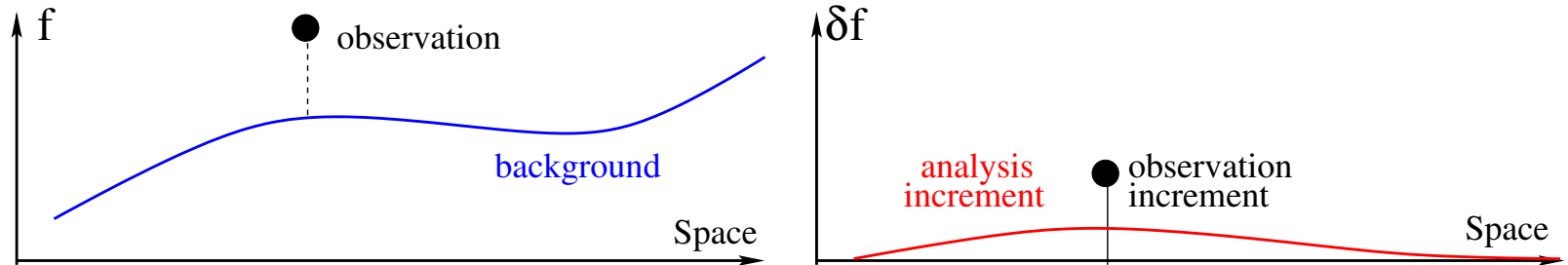
Optimal interpolation

- ✘ Bergthorsson and Döös (1955) took the idea of using a **background** field a step further by casting the analysis problem in terms of **increments** which were added to the background.



Optimal interpolation

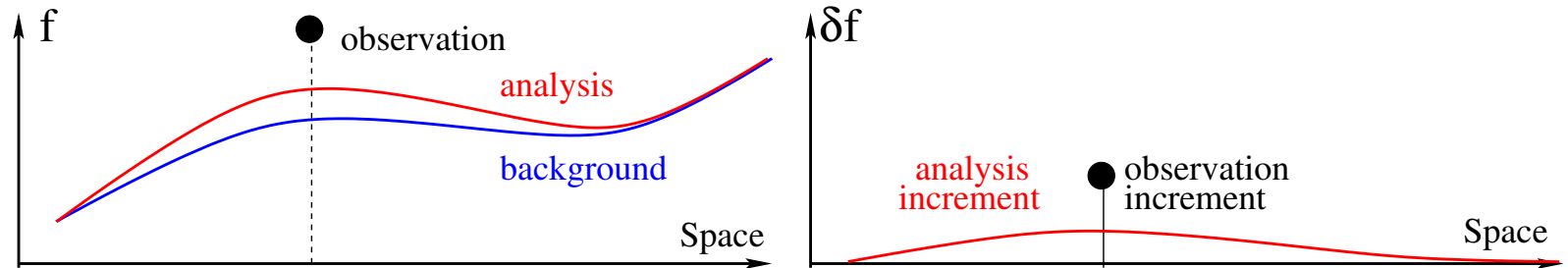
- ✘ Bergthorsson and Döös (1955) took the idea of using a **background** field a step further by casting the analysis problem in terms of **increments** which were added to the background.



- ✘ The increments were weighted linear combinations of nearby observation increments (observation minus background), with the weights determined statistically.

Optimal interpolation

- ✘ Bergthorsson and Döös (1955) took the idea of using a **background** field a step further by casting the analysis problem in terms of **increments** which were added to the background.



- ✘ The increments were weighted linear combinations of nearby observation increments (observation minus background), with the weights determined statistically.
- ✘ This idea of statistical combination of background and synoptic observations led ultimately to **Optimal Interpolation**.
- ✘ The use of statistics to merge model fields with observations is fundamental to all current methods of analysis.

Data Assimilation

- ✘ An important change of emphasis happened in the early 1970s with the introduction of primitive-equation models.
- ✘ Primitive equation models support inertia-gravity waves. This makes them much more fussy about their initial conditions than the filtered models that had been used hitherto.
- ✘ The analysis procedure became much more intimately linked with the model. The analysis had to produce an initial state that respected the model's dynamical balances.
- ✘ Unbalanced increments from the analysis procedure would be rejected as a result of geostrophic adjustment.
- ✘ Initialisation techniques (which suppress inertia-gravity waves) became important.



Data Assimilation

The idea that the analysis procedure must present observational information to the model in a way in which it can be absorbed (i.e. not rejected by geostrophic adjustment) led to the coining of the term **data assimilation**.

Wiktionary: Assimilate

1. To incorporate nutrients into the body, especially after digestion.

⇒ *Food is assimilated and converted into organic tissue.*

Data Assimilation

The idea that the analysis procedure must present observational information to the model in a way in which it can be absorbed (i.e. not rejected by geostrophic adjustment) led to the coining of the term **data assimilation**.

Wiktionary: Assimilate

1. To incorporate nutrients into the body, especially after digestion.

⇒ *Food is assimilated and converted into organic tissue.*

2. To incorporate or absorb knowledge into the mind.

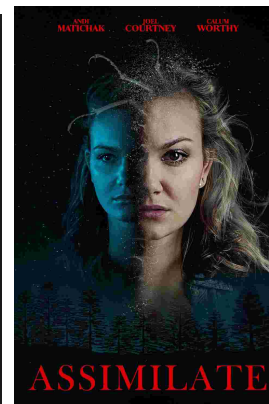
⇒ *The teacher paused in their lecture to allow the students to assimilate what they had said.*

Data Assimilation

The idea that the analysis procedure must present observational information to the model in a way in which it can be absorbed (i.e. not rejected by geostrophic adjustment) led to the coining of the term **data assimilation**.

Wiktionary: Assimilate

1. To incorporate nutrients into the body, especially after digestion.
 - ⇒ *Food is assimilated and converted into organic tissue.*
2. To incorporate or absorb knowledge into the mind.
 - ⇒ *The teacher paused in their lecture to allow the students to assimilate what they had said.*
3. To absorb a group of people into a community.
 - ⇒ *The aliens in the science-fiction film wanted to assimilate human beings into their own race.*



Data Assimilation

The idea that the analysis procedure must present observational information to the model in a way in which it can be absorbed (i.e. not rejected by geostrophic adjustment) led to the coining of the term **data assimilation**.

Wiktionary: Assimilate

1. To incorporate nutrients into the body, especially after digestion.

⇒ *Food is assimilated and converted into organic tissue.*

2. To incorporate or absorb knowledge into the mind.

⇒ *The teacher paused in their lecture to allow the students to assimilate what they had said.*

3. To absorb a group of people into a community.

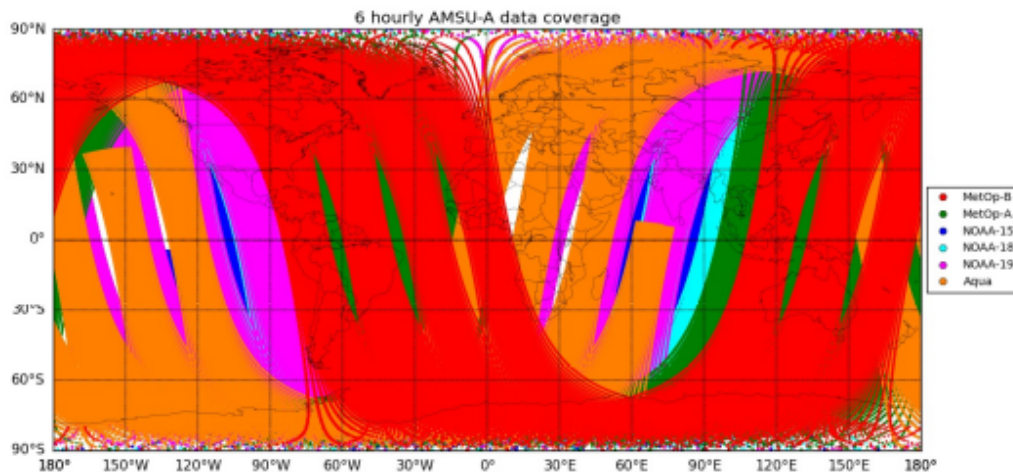
⇒ *The aliens in the science-fiction film wanted to assimilate human beings into their own race.*

Our definition

✘ The process of objectively adapting the model state to observations in a statistically optimal way taking into account model and observation errors

Data Assimilation

- ✘ A final impetus towards the modern concept of data assimilation came from the increasing availability of asynoptic observations from satellite instruments.
- ✘ It was no longer sufficient to think of the analysis purely in terms of spatial interpolation of contemporaneous observations.
- ✘ The time dimension became important, and the model dynamics assumed the role of propagating observational information in time to allow a synoptic view of the state of the system to be generated from asynoptic data.



- ✘ Example of satellite data coverage in 6 hours (AMSU-A data).

Outline

- 1 History and Terminology
- 2 Elementary Statistics — The Scalar Analysis Problem**
- 3 Extension to Multiple Dimensions
- 4 Optimal Interpolation
- 5 Summary

Elementary Statistics

Problem

Suppose we want to estimate the body temperature of a person, given:

- ✘ A prior estimate: T_b .
- ✘ A thermometer: T_o .
- ✘ The true (unknown) body temperature T_t .

Errors

- ✘ The errors in T_b and T_o are:

$$\varepsilon_b = T_b - T_t$$

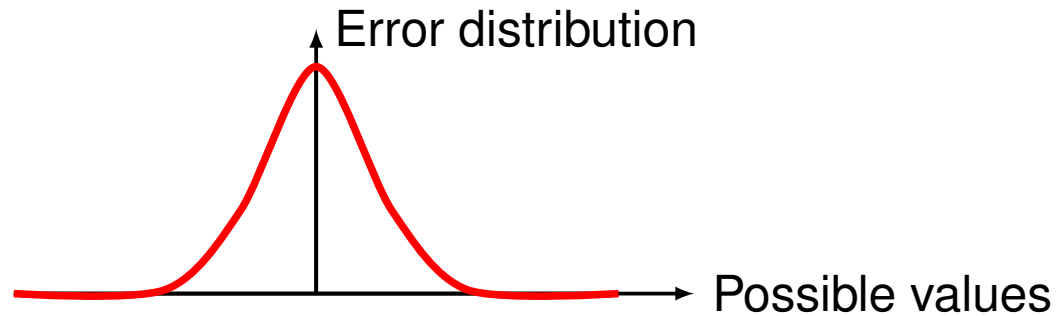
$$\varepsilon_o = T_o - T_t$$

- ✘ ε_b and ε_o are random variables (or stochastic variables)

Elementary Statistics

Hypotheses

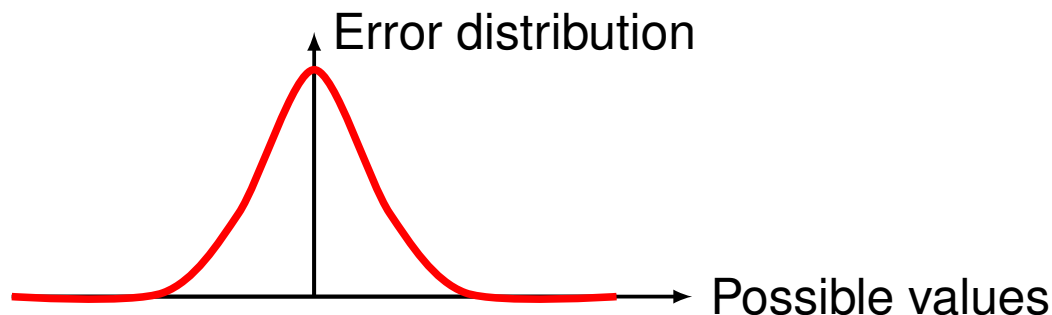
- ✘ We will assume that the error statistics of T_b and T_o are known.



Elementary Statistics

Hypotheses

- ✘ We will assume that the error statistics of T_b and T_o are known.



- ✘ We will assume that T_b and T_o have been adjusted (**bias corrected**) so that their mean errors are zero:

$$\overline{\varepsilon_b} = \overline{\varepsilon_o} = 0.$$

- ✘ There is usually no reason for ε_b and ε_o to be connected in any way:

$$\overline{\varepsilon_o \varepsilon_b} = 0.$$

- ✘ The quantity $\overline{\varepsilon_o \varepsilon_b}$ represents the **covariance** between the error of our prior estimate and the error of our thermometer measurement.

Elementary Statistics

- ✘ We estimate the body temperature as a **linear combination** of T_b and T_o :

$$T_a = \alpha T_o + \beta T_b + \gamma$$

Elementary Statistics

- ✘ We estimate the body temperature as a **linear combination** of T_b and T_o :

$$T_a = \alpha T_o + \beta T_b + \gamma$$

- ✘ Denote the error of our estimate as $\varepsilon_a = T_a - T_t$.
- ✘ We have:

$$T_a = T_t + \varepsilon_a = \alpha (T_t + \varepsilon_o) + \beta (T_t + \varepsilon_b) + \gamma$$

$$\text{or } \varepsilon_a = (\alpha + \beta - 1) T_t + \alpha \varepsilon_o + \beta \varepsilon_b + \gamma$$

Elementary Statistics

- ✘ We estimate the body temperature as a **linear combination** of T_b and T_o :

$$T_a = \alpha T_o + \beta T_b + \gamma$$

- ✘ Denote the error of our estimate as $\varepsilon_a = T_a - T_t$.

- ✘ We have:

$$T_a = T_t + \varepsilon_a = \alpha (T_t + \varepsilon_o) + \beta (T_t + \varepsilon_b) + \gamma$$

$$\text{or } \varepsilon_a = (\alpha + \beta - 1) T_t + \alpha \varepsilon_o + \beta \varepsilon_b + \gamma$$

- ✘ We want the estimate to be **unbiased**: $\overline{\varepsilon_a} = 0$:

$$\overline{\varepsilon_a} = (\alpha + \beta - 1) T_t + \gamma = 0$$

- ✘ Since this holds for any T_t , we must have

$$\Leftrightarrow \gamma = 0 \text{ and } \alpha + \beta - 1 = 0.$$

- ✘ Then

$$T_a = \alpha T_o + (1 - \alpha) T_b$$

Elementary Statistics

- ✘ The general **Linear Unbiased Estimate** is:

$$T_a = \alpha T_o + (1 - \alpha) T_b$$

- ✘ Now consider the error of this estimate.
- ✘ Subtracting T_t from both sides of the equation gives

$$\varepsilon_a = \alpha \varepsilon_o + (1 - \alpha) \varepsilon_b$$

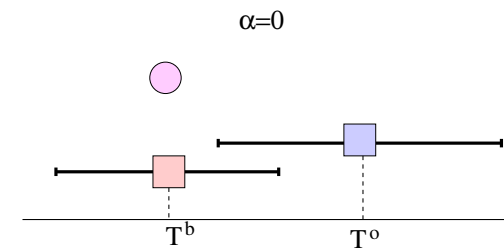
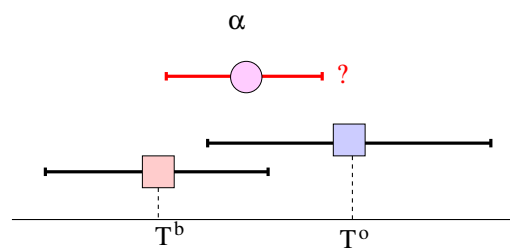
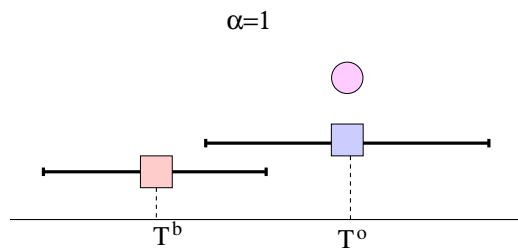
Elementary Statistics

- ✘ The general **Linear Unbiased Estimate** is:

$$T_a = \alpha T_o + (1 - \alpha) T_b$$

- ✘ Now consider the error of this estimate.
- ✘ Subtracting T_t from both sides of the equation gives

$$\epsilon_a = \alpha \epsilon_o + (1 - \alpha) \epsilon_b$$



Elementary Statistics

- ✘ The general **Linear Unbiased Estimate** is:

$$T_a = \alpha T_o + (1 - \alpha) T_b$$

- ✘ Now consider the error of this estimate.
- ✘ Subtracting T_t from both sides of the equation gives

$$\varepsilon_a = \alpha \varepsilon_o + (1 - \alpha) \varepsilon_b$$

- ✘ The variance of the estimate is:

$$\overline{\varepsilon_a^2} = \alpha^2 \overline{\varepsilon_o^2} + 2\alpha(1 - \alpha) \overline{\varepsilon_o \varepsilon_b} + (1 - \alpha)^2 \overline{\varepsilon_b^2}$$

- ✘ With the previous hypothesis $\overline{\varepsilon_o \varepsilon_b} = 0$:

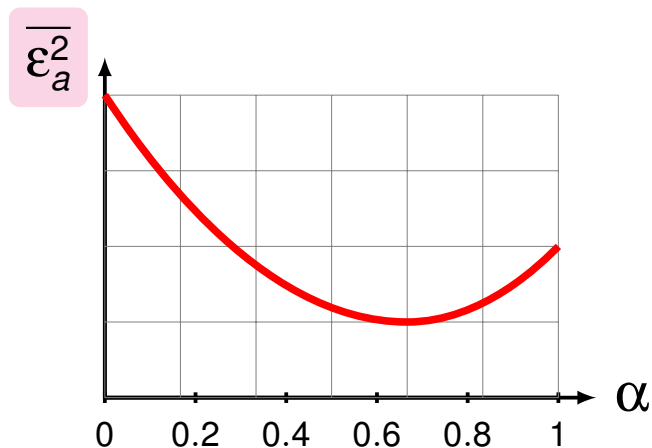
$$\overline{\varepsilon_a^2} = \alpha^2 \overline{\varepsilon_o^2} + (1 - \alpha)^2 \overline{\varepsilon_b^2}$$

Elementary Statistics

$$\overline{\varepsilon_a^2} = \alpha^2 \overline{\varepsilon_o^2} + (1 - \alpha)^2 \overline{\varepsilon_b^2}$$

We can easily derive some properties of our estimate:

- ✘ $\frac{d\overline{\varepsilon_a^2}}{d\alpha} = 2\alpha\overline{\varepsilon_o^2} - 2(1 - \alpha)\overline{\varepsilon_b^2}$
- ✘ For $\alpha = 0$, $\overline{\varepsilon_a^2} = \overline{\varepsilon_b^2}$ and $\frac{d\overline{\varepsilon_a^2}}{d\alpha} = -2\overline{\varepsilon_b^2} < 0$
- ✘ For $\alpha = 1$, $\overline{\varepsilon_a^2} = \overline{\varepsilon_o^2}$ and $\frac{d\overline{\varepsilon_a^2}}{d\alpha} = 2\overline{\varepsilon_o^2} > 0$



From this we can deduce:

- ✘ For $0 \leq \alpha \leq 1$, $\overline{\varepsilon_a^2} \leq \max(\overline{\varepsilon_b^2}, \overline{\varepsilon_o^2})$
- ✘ The minimum-variance estimate occurs for $\alpha \in (0, 1)$.
- ✘ The minimum-variance estimate satisfies $\overline{\varepsilon_a^2} < \min(\overline{\varepsilon_b^2}, \overline{\varepsilon_o^2})$, which means it is lower than the variance of each piece of information.

Elementary Statistics

The minimum-variance estimate occurs when

$$\frac{d \overline{\varepsilon_a^2}}{d\alpha} = 2\alpha \overline{\varepsilon_o^2} - 2(1 - \alpha) \overline{\varepsilon_b^2} = 0$$
$$\Rightarrow \alpha = \frac{\overline{\varepsilon_b^2}}{\overline{\varepsilon_b^2} + \overline{\varepsilon_o^2}}.$$

It is not difficult to show that the error variance of this **minimum-variance** estimate is:

$$\frac{1}{\overline{\varepsilon_a^2}} = \frac{1}{\overline{\varepsilon_b^2}} + \frac{1}{\overline{\varepsilon_o^2}},$$

and the analysis is:

$$\frac{T_a}{\overline{\varepsilon_a^2}} = \frac{T_b}{\overline{\varepsilon_b^2}} + \frac{T_o}{\overline{\varepsilon_o^2}}.$$

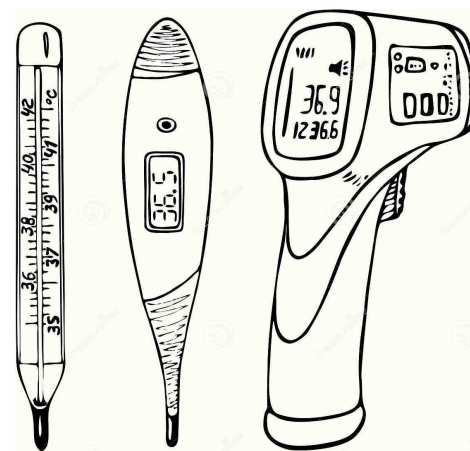
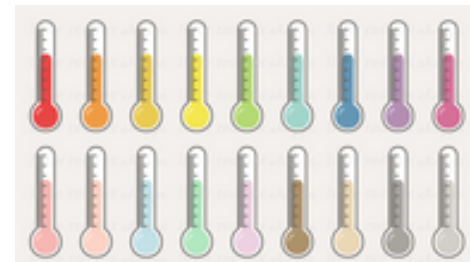
Outline

- 1 History and Terminology
- 2 Elementary Statistics — The Scalar Analysis Problem
- 3 Extension to Multiple Dimensions**
- 4 Optimal Interpolation
- 5 Summary

A Covid story - Part 2



Forehead	37.5°C
Armpit L - T ₁	36.0°C
	36.3°C
Armpit R - T ₁	36.2°C
	36.2°C
Armpit L - T ₂	36.1°C
	36.2°C
Armpit R - T ₂	35.9°C
	36.0°C
Armpit L - T ₃	...°C
	...°C
Armpit R - T ₃	...°C
	...°C



Extension to Multiple Dimensions

- ✘ Now, let's turn our attention to the **multi-dimensional** case.
- ✘ Instead of a scalar prior estimate T_b , we now consider **a vector \mathbf{x}_b** .
- ✘ We can think of \mathbf{x}_b as representing the entire state of **a numerical model** at some time.
- ✘ The elements of \mathbf{x}_b might be grid-point values, spherical harmonic coefficients, etc., and some elements may represent temperatures, humidity, others wind components, etc.
- ✘ We refer to \mathbf{x}_b as the background.
- ✘ Similarly, we generalise the observation to a vector \mathbf{y} .
- ✘ \mathbf{y} can contain a disparate collection of observations at different locations, and of different variables.

Extension to Multiple Dimensions

- ✘ The major difference between the simple scalar example and the multi-dimensional case is that there is no longer a one-to-one correspondence between the elements of the observation vector and those of the background vector.



- ✘ It is no longer trivial to compare observations and background.
- ✘ When the background is a state of a numerical model at some time
 - ⇒ Observations are not necessarily located at model gridpoints
 - ⇒ The observed variables (e.g. radiances) may not correspond directly with any of the variables of the model.
 - ⇒ To overcome this problem, we must assume that our model is a more-or-less complete representation of reality, so that we can always determine “model equivalents” of the observations.

Extension to Multiple Dimensions

- ✘ We formalise this by assuming the existence of an **observation operator**, \mathcal{H} .
- ✘ Given a model-space vector, \mathbf{x} , the vector $\mathcal{H}(\mathbf{x})$ can be compared directly with \mathbf{y} , and represents the “model equivalent” of \mathbf{y} .

$$\mathbf{x} \xrightarrow{\mathcal{H}(\cdot)} \mathcal{H}(\mathbf{x}) \rightarrow \text{Scales} \leftarrow \mathbf{y}$$

- ✘ For now, we will assume that \mathcal{H} is perfect. I.e. it does not introduce any error, so that:

$$\mathcal{H}(\mathbf{x}_t) = \mathbf{y}_t$$

where \mathbf{x}_t is the true state, and \mathbf{y}_t contains the true values of the observed quantities.

Extension to Multiple Dimensions

- ✘ As we did in the scalar case, we will look for an analysis that is a linear combination of the available information:

$$\mathbf{x}_a = \mathbf{F} \mathbf{x}_b + \mathbf{K} \mathbf{y} + \mathbf{c}$$

where \mathbf{F} and \mathbf{K} are matrices, and where \mathbf{c} is a vector.

- ✘ If \mathcal{H} is linear, we can proceed as in the scalar case and look for a **linear unbiased estimate**.
- ✘ In the more general case of nonlinear \mathcal{H} , we will require that error-free inputs ($\mathbf{x}_b = \mathbf{x}_t$ and $\mathbf{y} = \mathbf{y}_t$) produce an error-free analysis ($\mathbf{x}_a = \mathbf{x}_t$):

$$\mathbf{x}_t = \mathbf{F} \mathbf{x}_t + \mathbf{K} \mathcal{H}(\mathbf{x}_t) + \mathbf{c}$$

- ✘ Since this applies for any \mathbf{x}_t , we must have $\mathbf{c} = 0$ and

$$\mathbf{I} \equiv \mathbf{F} + \mathbf{K} \mathcal{H}(\cdot) \quad \text{or} \quad \mathbf{F} \equiv \mathbf{I} - \mathbf{K} \mathcal{H}(\cdot)$$

- ✘ Our analysis equation is thus:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K} (\mathbf{y} - \mathcal{H}(\mathbf{x}_b))$$

Extension to Multiple Dimensions

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathcal{H}(\mathbf{x}_b))$$

- ✘ Remember that in the scalar case, we had

$$\begin{aligned} T_a &= \alpha T_o + (1 - \alpha) T_b \\ &= T_b + \alpha(T_o - T_b) \end{aligned}$$

- ✘ We see that the matrix \mathbf{K} plays a role equivalent to that of the coefficient α .
- ✘ \mathbf{K} is called the **gain matrix**.
- ✘ It determines the weight given to the **innovation** $\mathbf{y} - \mathcal{H}(\mathbf{x}_b)$
- ✘ It handles the transformation of information defined in “observation space” to the space of model variables.

Extension to Multiple Dimensions

- ✘ The next step in deriving the analysis equation is to describe the statistical properties of the analysis errors.
- ✘ We define

$$\boldsymbol{\varepsilon}_a = \mathbf{x}_a - \mathbf{x}_t$$

$$\boldsymbol{\varepsilon}_b = \mathbf{x}_b - \mathbf{x}_t$$

$$\boldsymbol{\varepsilon}_o = \mathbf{y} - \mathbf{y}_t$$

- ✘ We will assume that the errors are small, so that

$$\mathcal{H}(\mathbf{x}_b) = \mathcal{H}(\mathbf{x}_t) + \mathbf{H} \boldsymbol{\varepsilon}_b + O(\boldsymbol{\varepsilon}_b^2)$$

where \mathbf{H} is the Jacobian of \mathcal{H} (if \mathbf{H} is nonlinear).

Extension to Multiple Dimensions

- ✘ Substituting the expressions for the errors into our analysis equation, and using $\mathcal{H}(\mathbf{x}_t) = \mathbf{y}_t$, gives (to first order):

$$\boldsymbol{\varepsilon}_a = \boldsymbol{\varepsilon}_b + \mathbf{K} (\boldsymbol{\varepsilon}_o - \mathbf{H} \boldsymbol{\varepsilon}_b)$$

- ✘ As in the scalar example, we will assume that the mean errors have been removed, so that $\bar{\boldsymbol{\varepsilon}}_b = \bar{\boldsymbol{\varepsilon}}_o = 0$. We see that this implies that $\bar{\boldsymbol{\varepsilon}}_a = 0$.
- ✘ In the scalar example, we derived the variance of the analysis error, and defined our optimal analysis to minimise this variance.
- ✘ In the multi-dimensional case, we must deal with **covariances**.

Covariance

- ✘ The **covariance** between two variables x_i and x_j is defined as

$$\text{cov}(x_i, x_j) = \overline{(x_i - \bar{x}_i)(x_j - \bar{x}_j)}$$

- ✘ Given a vector $\mathbf{x} = (x_1, x_2, \dots, x_N)^T$, we can arrange the covariances into a **covariance matrix**, \mathbf{C} , such that $C_{ij} = \text{cov}(x_i, x_j)$.

- ✘ Equivalently:

$$\mathbf{C} = \overline{(\mathbf{x} - \bar{\mathbf{x}})(\mathbf{x} - \bar{\mathbf{x}})^T}$$

- ✘ Covariance matrices are **symmetric** and **positive definite**
 - ⇒ symmetric: $\mathbf{C}^T = \mathbf{C}$
 - ⇒ positive definite: $\mathbf{z}^T \mathbf{C} \mathbf{z}$ is positive for every non-zero vector \mathbf{z}

Extension to Multiple Dimensions

✘ The analysis error is:

$$\begin{aligned}\epsilon_a &= \epsilon_b + \mathbf{K} (\epsilon_o - \mathbf{H} \epsilon_b) \\ &= (\mathbf{I} - \mathbf{KH}) \epsilon_b + \mathbf{K} \epsilon_o\end{aligned}$$

Extension to Multiple Dimensions

✘ The analysis error is:

$$\begin{aligned}\boldsymbol{\varepsilon}_a &= \boldsymbol{\varepsilon}_b + \mathbf{K} (\boldsymbol{\varepsilon}_o - \mathbf{H} \boldsymbol{\varepsilon}_b) \\ &= (\mathbf{I} - \mathbf{K}\mathbf{H}) \boldsymbol{\varepsilon}_b + \mathbf{K} \boldsymbol{\varepsilon}_o\end{aligned}$$

✘ Forming the **analysis error covariance matrix** gives:

$$\begin{aligned}\overline{\boldsymbol{\varepsilon}_a \boldsymbol{\varepsilon}_a^T} &= \overline{[(\mathbf{I} - \mathbf{K}\mathbf{H}) \boldsymbol{\varepsilon}_b + \mathbf{K} \boldsymbol{\varepsilon}_o] [(\mathbf{I} - \mathbf{K}\mathbf{H}) \boldsymbol{\varepsilon}_b + \mathbf{K} \boldsymbol{\varepsilon}_o]^T} \\ &= (\mathbf{I} - \mathbf{K}\mathbf{H}) \overline{\boldsymbol{\varepsilon}_b \boldsymbol{\varepsilon}_b^T} (\mathbf{I} - \mathbf{K}\mathbf{H})^T \\ &\quad + \mathbf{K} \overline{\boldsymbol{\varepsilon}_o \boldsymbol{\varepsilon}_o^T} \mathbf{K}^T \\ &\quad + \mathbf{K} \overline{\boldsymbol{\varepsilon}_o \boldsymbol{\varepsilon}_b^T} (\mathbf{I} - \mathbf{K}\mathbf{H})^T + (\mathbf{I} - \mathbf{K}\mathbf{H}) \overline{\boldsymbol{\varepsilon}_b \boldsymbol{\varepsilon}_o^T} \mathbf{K}^T\end{aligned}$$

✘ Assuming that the background and observation errors are uncorrelated (i.e. $\overline{\boldsymbol{\varepsilon}_o \boldsymbol{\varepsilon}_b^T} = \overline{\boldsymbol{\varepsilon}_b \boldsymbol{\varepsilon}_o^T} = 0$), we find:

$$\overline{\boldsymbol{\varepsilon}_a \boldsymbol{\varepsilon}_a^T} = (\mathbf{I} - \mathbf{K}\mathbf{H}) \overline{\boldsymbol{\varepsilon}_b \boldsymbol{\varepsilon}_b^T} (\mathbf{I} - \mathbf{K}\mathbf{H})^T + \mathbf{K} \overline{\boldsymbol{\varepsilon}_o \boldsymbol{\varepsilon}_o^T} \mathbf{K}^T$$

Extension to Multiple Dimensions

$$\overline{\boldsymbol{\varepsilon}_a \boldsymbol{\varepsilon}_a^T} = (\mathbf{I} - \mathbf{KH}) \overline{\boldsymbol{\varepsilon}_b \boldsymbol{\varepsilon}_b^T} (\mathbf{I} - \mathbf{KH})^T + \mathbf{K} \overline{\boldsymbol{\varepsilon}_o \boldsymbol{\varepsilon}_o^T} \mathbf{K}^T$$

- ✘ This expression is the equivalent of the expression we obtained for the error of the scalar analysis:

$$\begin{aligned} \overline{\varepsilon_a^2} &= (1 - \alpha)^2 \overline{\varepsilon_b^2} + \alpha^2 \overline{\varepsilon_o^2} \\ &= (1 - \alpha) \overline{\varepsilon_b^2} (1 - \alpha) + \alpha \overline{\varepsilon_o^2} \alpha \end{aligned}$$

- ✘ Again, we see that \mathbf{K} plays essentially the same role in the multi-dimensional analysis as α plays in the scalar case.
- ✘ In the scalar case, we chose α to minimise the variance of the analysis error.
- ✘ What do we mean by the **minimum-variance** analysis in the multi-dimensional case?

Extension to Multiple Dimensions

- ✘ Note that the diagonal elements of a covariance matrix are **variances**
 $C_{ij} = \text{cov}(x_i, x_j) = \overline{(x_i - \bar{x}_i)(x_j - \bar{x}_j)}$.
- ✘ Hence, we can define the minimum-variance analysis as the analysis that minimises the sum of the diagonal elements of the analysis error covariance matrix.
- ✘ The sum of the diagonal elements of a matrix is called the **trace**.
- ✘ In the scalar case, we found the minimum-variance analysis by setting $\frac{d\varepsilon_a^2}{d\alpha}$ to zero.
- ✘ In the multidimensional case, we are going to set

$$\frac{\partial \text{trace}(\overline{\varepsilon_a \varepsilon_a^T})}{\partial \mathbf{K}} = \mathbf{0}$$

- ✘ Note: $\frac{\partial \text{trace}(\overline{\varepsilon_a \varepsilon_a^T})}{\partial \mathbf{K}}$ is the matrix whose ij^{th} element is $\frac{\partial \text{trace}(\overline{\varepsilon_a \varepsilon_a^T})}{\partial K_{ij}}$.

Extension to Multiple Dimensions

✘ We have: $\overline{\boldsymbol{\varepsilon}_a \boldsymbol{\varepsilon}_a^T} = (\mathbf{I} - \mathbf{KH}) \overline{\boldsymbol{\varepsilon}_b \boldsymbol{\varepsilon}_b^T} (\mathbf{I} - \mathbf{KH})^T + \mathbf{K} \overline{\boldsymbol{\varepsilon}_o \boldsymbol{\varepsilon}_o^T} \mathbf{K}^T$.

✘ The following matrix identities come to our rescue:

$$\begin{aligned} \frac{\partial \text{trace}(\mathbf{KAK}^T)}{\partial \mathbf{K}} &= \mathbf{K}(\mathbf{A} + \mathbf{A}^T) \\ \frac{\partial \text{trace}(\mathbf{KA})}{\partial \mathbf{K}} &= \mathbf{A}^T \\ \frac{\partial \text{trace}(\mathbf{AK}^T)}{\partial \mathbf{K}} &= \mathbf{A} \end{aligned}$$

✘ Applying these to $\partial \text{trace}(\overline{\boldsymbol{\varepsilon}_a \boldsymbol{\varepsilon}_a^T}) / \partial \mathbf{K}$ gives:

$$\frac{\partial \text{trace}(\overline{\boldsymbol{\varepsilon}_a \boldsymbol{\varepsilon}_a^T})}{\partial \mathbf{K}} = 2\mathbf{K} \left[\mathbf{H} \overline{\boldsymbol{\varepsilon}_b \boldsymbol{\varepsilon}_b^T} \mathbf{H}^T + \overline{\boldsymbol{\varepsilon}_o \boldsymbol{\varepsilon}_o^T} \right] - 2 \overline{\boldsymbol{\varepsilon}_b \boldsymbol{\varepsilon}_b^T} \mathbf{H}^T = \mathbf{0}$$

✘ Hence: $\mathbf{K} = \overline{\boldsymbol{\varepsilon}_b \boldsymbol{\varepsilon}_b^T} \mathbf{H}^T \left[\mathbf{H} \overline{\boldsymbol{\varepsilon}_b \boldsymbol{\varepsilon}_b^T} \mathbf{H}^T + \overline{\boldsymbol{\varepsilon}_o \boldsymbol{\varepsilon}_o^T} \right]^{-1}$.

Extension to Multiple Dimensions

$$\mathbf{K} = \overline{\boldsymbol{\varepsilon}_b \boldsymbol{\varepsilon}_b^T} \mathbf{H}^T \left[\mathbf{H} \overline{\boldsymbol{\varepsilon}_b \boldsymbol{\varepsilon}_b^T} \mathbf{H}^T + \overline{\boldsymbol{\varepsilon}_o \boldsymbol{\varepsilon}_o^T} \right]^{-1}$$

- ✘ This optimal gain matrix is called the **Kalman Gain Matrix**.
- ✘ Note the similarity with the optimal gain we derived for the scalar analysis:

$$\alpha = \overline{\varepsilon_b^2} \left[\overline{\varepsilon_b^2} + \overline{\varepsilon_o^2} \right]^{-1}$$

- ✘ The variance of analysis error for the optimal scalar problem was:

$$\frac{1}{\overline{\varepsilon_a^2}} = \frac{1}{\overline{\varepsilon_b^2}} + \frac{1}{\overline{\varepsilon_o^2}}$$

- ✘ The equivalent expression for the multi-dimensional case is:

$$\left[\overline{\boldsymbol{\varepsilon}_a \boldsymbol{\varepsilon}_a^T} \right]^{-1} = \left[\overline{\boldsymbol{\varepsilon}_b \boldsymbol{\varepsilon}_b^T} \right]^{-1} + \mathbf{H}^T \left[\overline{\boldsymbol{\varepsilon}_o \boldsymbol{\varepsilon}_o^T} \right]^{-1} \mathbf{H}$$

Notation

- ✘ The notation we have used for covariance matrices can get a bit cumbersome.
- ✘ The standard notation is:

$$\mathbf{P}^a \equiv \overline{\boldsymbol{\varepsilon}_a \boldsymbol{\varepsilon}_a^T}$$
$$\mathbf{P}^b \equiv \overline{\boldsymbol{\varepsilon}_b \boldsymbol{\varepsilon}_b^T}$$
$$\mathbf{R} \equiv \overline{\boldsymbol{\varepsilon}_o \boldsymbol{\varepsilon}_o^T}$$

- ✘ In many analysis schemes, the true covariance matrix of background error, \mathbf{P}^b , is not known, or is too large to be used.
- ✘ In this case, we use an approximate background error covariance matrix. This approximate matrix is denoted by \mathbf{B} .

Alternative Expression for the Kalman Gain

Finally, we derive an alternative expression for the Kalman gain:

$$\mathbf{K} = \mathbf{P}^b \mathbf{H}^T [\mathbf{H} \mathbf{P}^b \mathbf{H}^T + \mathbf{R}]^{-1}$$

Multiplying both sides by $[\mathbf{P}^{b-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}]$ gives:

$$\begin{aligned} [\mathbf{P}^{b-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}] \mathbf{K} &= [\mathbf{H}^T + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H} \mathbf{P}^b \mathbf{H}^T] [\mathbf{H} \mathbf{P}^b \mathbf{H}^T + \mathbf{R}]^{-1} \\ &= \mathbf{H}^T \mathbf{R}^{-1} [\mathbf{R} + \mathbf{H} \mathbf{P}^b \mathbf{H}^T] [\mathbf{H} \mathbf{P}^b \mathbf{H}^T + \mathbf{R}]^{-1} \\ &= \mathbf{H}^T \mathbf{R}^{-1} \end{aligned}$$

Hence:

$$\mathbf{K} = [\mathbf{P}^{b-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H}]^{-1} \mathbf{H}^T \mathbf{R}^{-1}$$

- ✘ Expression 1: need the inverse of a matrix of dimension size(\mathbf{R})
- ✘ Expression 2: need the inverse of a matrix of dimension size(\mathbf{P}^b)
- ✘ Remember that $\mathbf{x}_a = \mathbf{x}_b + \mathbf{K}(\mathbf{y} - \mathcal{H}(\mathbf{x}_b))$

Outline

- 1 History and Terminology
- 2 Elementary Statistics — The Scalar Analysis Problem
- 3 Extension to Multiple Dimensions
- 4 Optimal Interpolation**
- 5 Summary

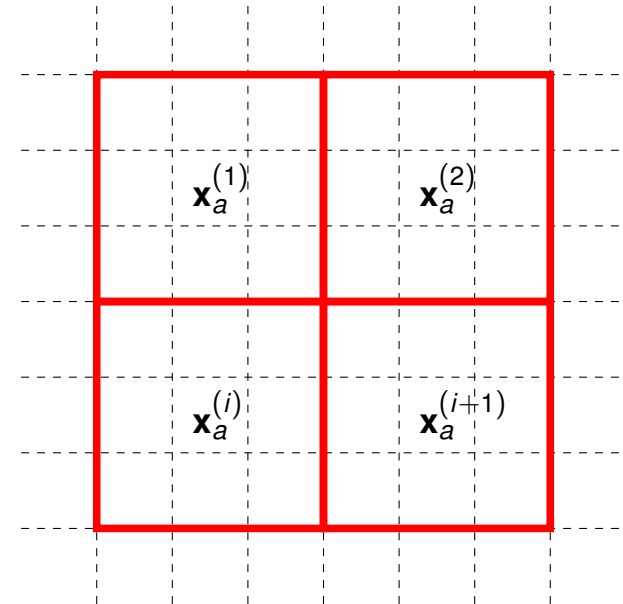
Optimal Interpolation

- ✘ **Optimal Interpolation** is a statistical data assimilation method based on the multi-dimensional analysis equations we have just derived.
- ✘ The Kalman gain \mathbf{K} can not be computed because of the size of \mathbf{P}^b and \mathbf{R}
- ✘ The basic idea is to split the global analysis into a number of boxes which can be analysed independently:

$$\mathbf{x}_a^{(i)} = \mathbf{x}_b^{(i)} + \mathbf{K}^{(i)} \left[\mathbf{y}^{(i)} - \mathcal{H}^{(i)}(\mathbf{x}_b) \right]$$

where

$$\mathbf{x}_a = \begin{pmatrix} \mathbf{x}_a^{(1)} \\ \mathbf{x}_a^{(2)} \\ \vdots \\ \mathbf{x}_a^{(M)} \end{pmatrix} \quad \mathbf{x}_b = \begin{pmatrix} \mathbf{x}_b^{(1)} \\ \mathbf{x}_b^{(2)} \\ \vdots \\ \mathbf{x}_b^{(M)} \end{pmatrix} \quad \mathbf{K} = \begin{pmatrix} \mathbf{K}^{(1)} \\ \mathbf{K}^{(2)} \\ \vdots \\ \mathbf{K}^{(M)} \end{pmatrix}$$

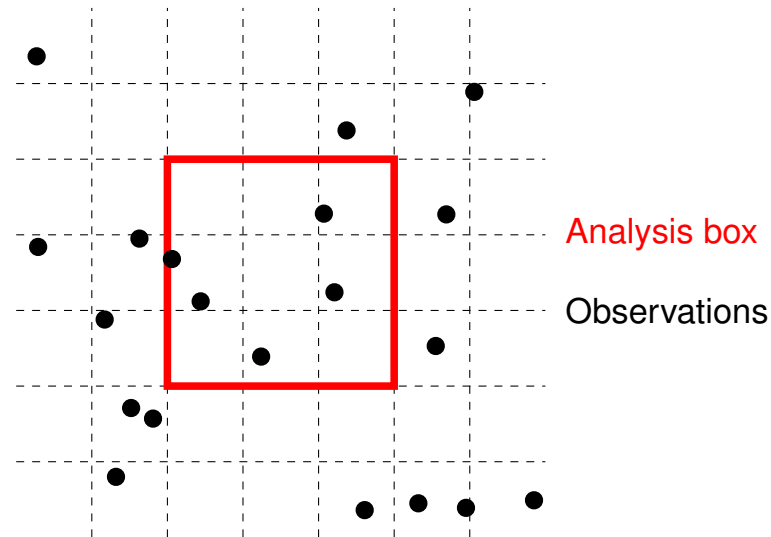


- ✘ The method was used operationally at ECMWF from 1979 until 1996, when it was replaced by 3D-Var.

Optimal Interpolation

$$\mathbf{x}_a^{(i)} = \mathbf{x}_b^{(i)} + \mathbf{K}^{(i)} \left(\mathbf{y}^{(i)} - \mathcal{H}^{(i)}(\mathbf{x}_b) \right)$$

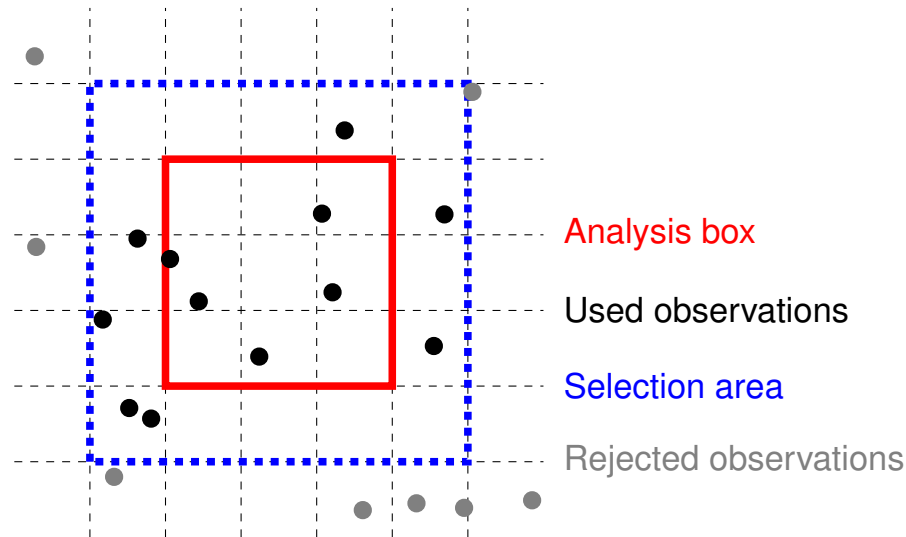
- ✘ In principle, we should use *all* available observations to calculate the analysis for each box. However, this might be too expensive.
- ✘ To produce a computationally-feasible algorithm, Optimal Interpolation (OI) restricts the observations used for each box to those observations which lie in a surrounding selection area:



Optimal Interpolation

$$\mathbf{x}_a^{(i)} = \mathbf{x}_b^{(i)} + \mathbf{K}^{(i)} \left(\mathbf{y}^{(i)} - \mathcal{H}^{(i)}(\mathbf{x}_b) \right)$$

- ✘ In principle, we should use *all* available observations to calculate the analysis for each box. However, this might be too expensive.
- ✘ To produce a computationally-feasible algorithm, Optimal Interpolation (OI) restricts the observations used for each box to those observations which lie in a surrounding selection area:



Optimal Interpolation

- ✘ The gain matrix used for each box is:

$$\mathbf{K}^{(i)} = (\mathbf{P}^b \mathbf{H}^T)^{(i)} \left[(\mathbf{H} \mathbf{P}^b \mathbf{H}^T)^{(i)} + \mathbf{R}^{(i)} \right]^{-1}$$

- ✘ Now, the dimension of the matrix $\left[(\mathbf{H} \mathbf{P}^b \mathbf{H}^T)^{(i)} + \mathbf{R}^{(i)} \right]$ is equal to the number of observations in the selection box.
- ✘ Selecting observations reduces the size of this matrix, making it feasible to use **direct solution methods** to invert it.
- ✘ Note that to implement Optimal Interpolation, we have to specify $(\mathbf{P}^b \mathbf{H}^T)^{(i)}$ and $(\mathbf{H} \mathbf{P}^b \mathbf{H}^T)^{(i)}$. This effectively limits us to very simple observation operators, corresponding to simple interpolations.
- ✘ This, together with the artifacts introduced by observation selection, was one of the main reasons for abandoning Optimal Interpolation in favour of 3D-Var.

Outline

- 1 History and Terminology
- 2 Elementary Statistics — The Scalar Analysis Problem
- 3 Extension to Multiple Dimensions
- 4 Optimal Interpolation
- 5 Summary**

Summary

- ✘ We derived the linear analysis equation for a simple scalar example.
- ✘ We showed that a particular choice of the weight α given to the observation resulted in an optimal **minimum-variance** analysis.
- ✘ We repeated the derivation for the multi-dimensional case. This required the introduction of the **observation operator**.
- ✘ The derivation for the multi-dimensional case closely paralleled the scalar derivation.
- ✘ The expressions for the gain matrix and analysis error covariance matrix were recognisably similar to the corresponding scalar expressions.
- ✘ Finally, we considered the practical implementation of the analysis equation, in an **Optimal Interpolation** data assimilation scheme.