Bayesian Theory, Cost Function, Minimization, Incremental Optimization, Observation System Simulation Experiment

Vinu Valsala

IITM

Bayes Theorem

- Bayes Theorem states that
  - **P (A | B) = P (B | A) * [P (A) / P(B)]**

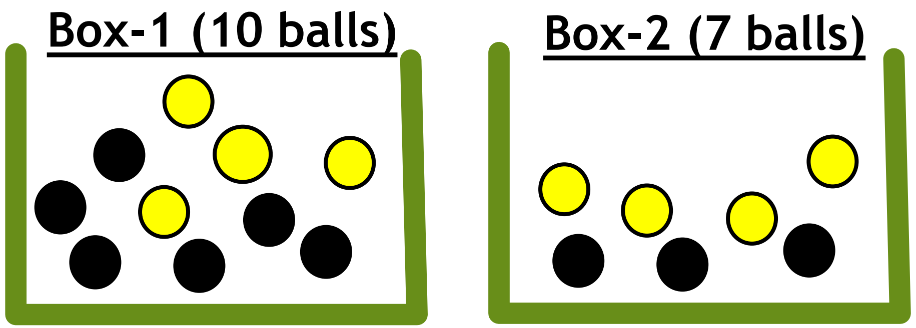
P (A | B) is the Probability of A given B

P (B | A) is the Probability of B given A

P (A) is the single probability of A

P (B) is the single probability of B

Example:



Suppose 'one' ball is chosen from one of the boxes, and that happens to be a 'black' ball, then what is the probability that it is drawn from Box-1?

Let A = Black Ball; E = Event of drawing

P (E1) = P (E2) = ½

P (A | E1) = 6/10 = 3/5

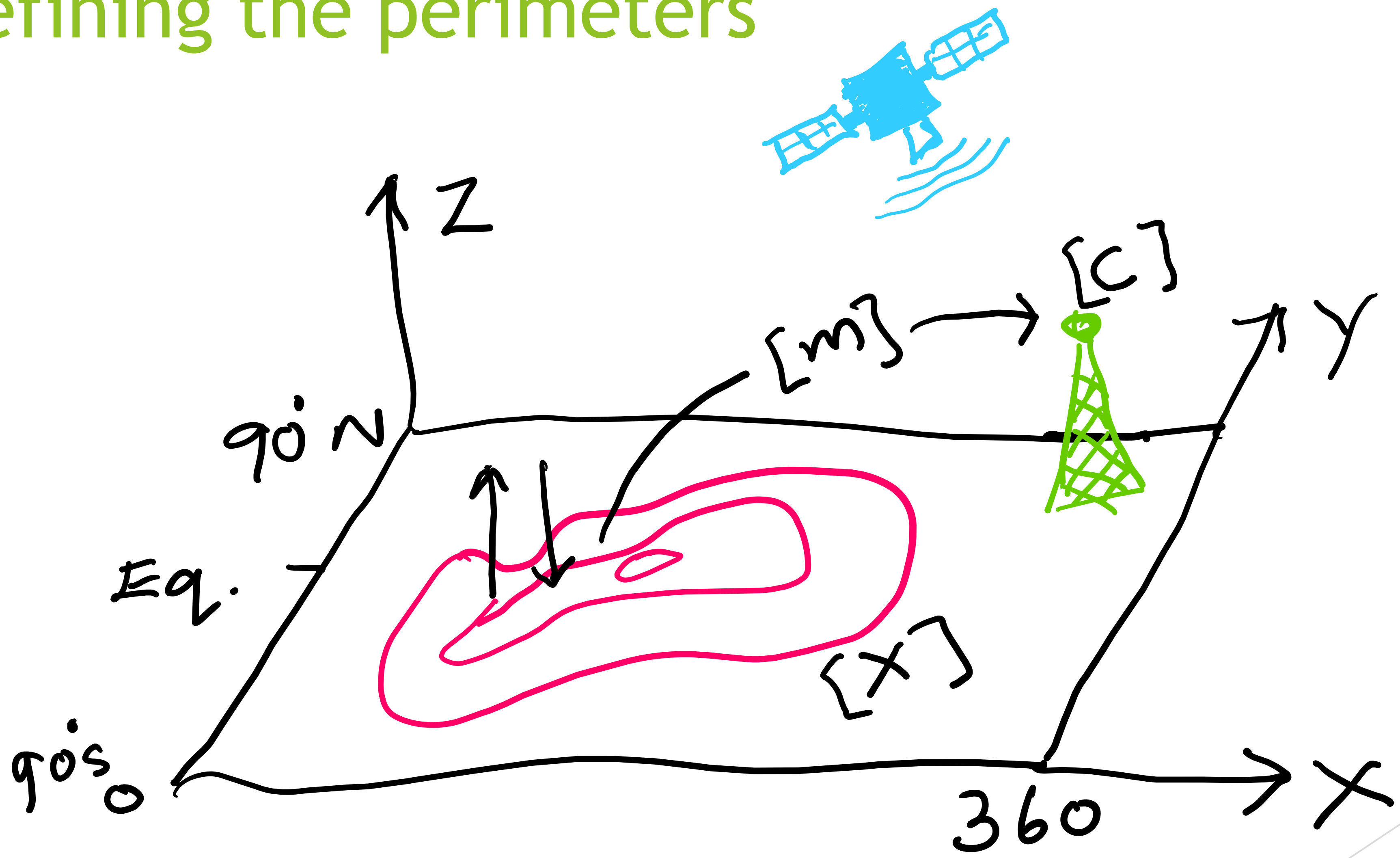
P (A | E2) = 3/7

P (E1 | A) = P (A | E1) *[P (E1) / P (A)]

P (A) = ½ * (6/10 + 3/7)

P (E1 | A) = 7/12 = 0.58 = 58%

Defining the perimeters



[c] is the observations of an atmospheric variable
[x] is a parameter (or a source) causing the [c]
[m] is a theoretical relation exist between the [x] and [c]

Theoretical relation between [x] and [c]

$$\boxed{\frac{\partial [c]}{\partial t} + \nabla \cdot [U c] - \nabla \cdot A \cdot \nabla [c] = X}$$

Here the above theoretical relation can be interpreted as a mathematical model, or a numerical model.

(What is ‘A’ and ‘U’ in the above model ?)

Application of Bayes theorem in our perimeter

P (x | c) is the probability of a source [x] given [c] is the observations caused by that [x]

P_m (c | x) is the probability of [c] given [x] derived via the theoretical relation (i.e. the model)

The Bayes theorem states that

$$P(x \mid c) = P_m(c \mid x) * P(x) \,/\, P(c)$$

Or it can be sufficiently replaced as;

$$P_{Estimate}(x \mid C) = \frac{P_{model}(c \mid x_{prior}) \cdot P(x)_{prior}}{P(C)}$$

Application of Bayes theorem in our perimeter

$$P_{Estimate}(x|c) = \frac{\boxed{P_{model}(c|x_{prior})} \cdot \boxed{P(x)_{prior}}}{\boxed{P(c)}}$$

This is the 'Optimal' estimate of [x]

This is generally known as 'Posterior'.

This we can model, if we know some distribution of [x]

This is known as the observations

This is some known distribution of [x]

This is generally known as 'Prior'

P (c) represent the entirety of the problem such as $P(c) = \int P(c \mid x_{prior})\, P(x_{prior})\, dx$, why ?

Constructing a 'G' from our model

$$\boxed{\frac{\partial [c]}{\partial t} + \nabla \cdot [U c] - \nabla \cdot A \cdot \nabla [c] = X}$$

$$\boxed{G X = C}$$

G stands for 'Greens' function.
Green's function is a 'response function' which generates a change in one variable due to an 'impulse' occurring elsewhere.

(Can you tell an example of a Green's function scenario in real life ?)

Weighted Least-Square Estimates

- The *weighted least square estimate* is defined as the vector that minimizes the objective function
  - $J(x) = (c - Gx)^T X^{-1} (c - Gx)$
  - where $\mathbf{X}^{-1}$ is the covariance matrix of $p(\mathbf{c}|\mathbf{m})$.
  - **The estimate (minima of J(x)) is $\mathbf{x}_{WLS} = [\mathbf{G}^T\mathbf{X}^{-1}\mathbf{G}]^{-1}[\mathbf{G}^T\mathbf{X}^{-1}\mathbf{c}]$**

Bayesian Least-Square Estimates

- The *Bayesian least square estimate* is defined as the vector that minimizes the objective function
  - $J(x) = (c - Gx)^T X^{-1} (c - Gx) + (x - x_0)^T W^{-1} (x - x_0)$
  - where $\mathbf{X}^{-1}$ is the covariance matrix of $p(\mathbf{c}|\mathbf{m})$.
  - **The estimate (minima of J(x)) is $\mathbf{x}_{BLS} = [\mathbf{G}^T\mathbf{X}^{-1}\mathbf{G} + \mathbf{W}^{-1}]^{-1}[\mathbf{G}^T\mathbf{X}^{-1}\mathbf{c} + \mathbf{W}^{-1}\mathbf{x}_0]$**

Bayesian Least-Square Estimates

- In the previous Estimate, do you think the Estimate is the absolute and having no errors?
  - Answer is No. All Estimators comes with residual error.
- In the previous Estimate, do you think the Model is the absolute reality and having no errors?
  - Answer is No. All models are *in-complete* and so does the predictions of relation between [x] and [c].
- So what are the chances of the total error in your Estimate in a least-square fashion?
  - $J(x) = (c - m)^2 + (x_{prior} - x_{posterior})^2$
  - Why we need to square the errors (?)

$J(x) = (c - m)^2 + (x_{prior} - x_{posterior})^2$

- The above function consists of [c], [m], [x_{prior}] and [$x_{posterior}$] which all are vectors.
- In a vector space, $[x]^2 = [x]^T[x]$
  - Example:

$$[x] = \begin{bmatrix} 2 \\ 3 \\ 4 \\ 0 \end{bmatrix} \qquad [x]^T \cdot [x] = \begin{bmatrix} 2 & 3 & 4 & 0 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \\ 4 \\ 0 \end{bmatrix} = [29]$$

- The above function J(x) is called the 'Penalty Function' or 'Cost Function' of this system of [c], [m], and [x].

Attaching Uncertainties to the Cost function

- Is the [c] is perfect observation? Is [m] a perfect model?
  - Answer is No.
- Therefore, we need to attach few 'Uncertainties' in your 'Penalty function' so that the entirety of the 'Penalty' is not accountable to just our estimates alone.

- $J(x) = (c - m)^T R^{-1} (c - m) + (x_{prior} - x_{posterior})^T B^{-1}(x_{prior} - x_{posterior})$

- Or in other words, we can express this as
  - $J(x) = (c - Gx)^T R^{-1} (c - Gx) + (x - x_0)^T B^{-1}(x - x_0)$

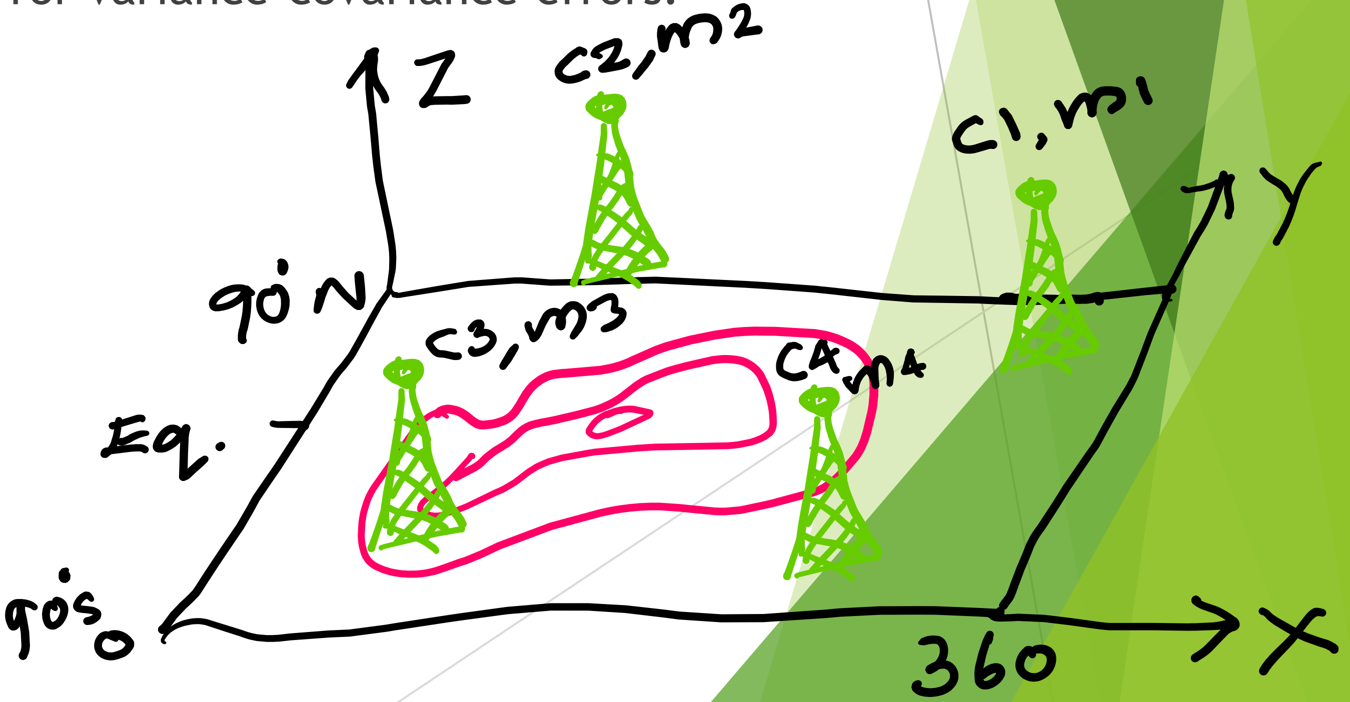
  (where $x = x_{posterior}$ and $x_0 = x_{prior}$ for convenience)

  (where R is the model and data error variance-covariance matrix, and B is the error variance-covariance in the assumed X)

Error variance-covariance matrices

- R represent the covariances of P (c | m).
  - If the simulated outcome '[m]' of observation '[c]' are equal, the probability of 'c' given 'm' is one.
  - In the above case R = [I] ; and the identity matrix
  - In reality, however, model has errors, and observations have errors; therefore, R represents the model-data error variance-covariance errors.

$$
R = \begin{matrix} \sigma_1 & \text{Cov} & \text{Cov} & \text{Cov} \\ \text{Cov} & \sigma_2 & \text{Cov} & \text{Cov} \\ \text{Cov} & \text{Cov} & \sigma_3 & \text{Cov} \\ \text{Cov} & \text{Cov} & \text{Cov} & \sigma_4 \end{matrix}
$$



The Cost Function

- $J(x) = (c - Gx)^T R^{-1} (c - Gx) + (x - x_0)^T B^{-1}(x - x_0)$
  - It refers to the total errors in our estimate
  - It contain two parts (not generic, but in this specific case)
  - The first part represent for the model inabilities to simulate an observation [c]
  - The second part contains the remaining errors in the estimate of [x]
  - In the totality the J(x) is the total error in the system
  - If J(x) is the total error, how can we find ‘minima’ of the error?
- Lets find $\partial J(x)/\partial x$ and put that equal to zero

The minimization

$$J(x) = (c - Gx)^T R^{-1} (c - Gx) + (x - x_0)^T B^{-1} (x - x_0)$$

$$= c^T R^{-1} c + x^T G^T R^{-1} G x - x^T G^T R^{-1} c - c^T R^{-1} G x$$

$$+ x^T B^{-1} x + x_0^T B^{-1} x_0 - x_0^T B^{-1} x - x^T B^{-1} x_0$$

Taking the identity

$$\boxed{c^T R^{-1} G x = x^T G^T R^{-1} c}$$

$$\boxed{x^T B^{-1} x_0 = x_0^T B^{-1} x}$$

The minimization

$$= c^T R^{-1} c + x^T G^T R^{-1} G x - 2 c^T R^{-1} G x$$

$$+ x^T B^{-1} x + x_0^T B^{-1} x_0 - 2 x_0^T B^{-1} x$$

Taking derivative of $J(x)$ w.r.t. x

and identity

$$\frac{\partial J(x)}{\partial x} = 0.$$

$$\boxed{\nabla(x^T A x) = 2Ax}$$

$$\boxed{\nabla(B^T x) = B}$$

$$= 0 + 2G^T R^{-1} G x - 2G^T R^{-1} c + 2B^{-1} x + 0$$

$$- 2B^{-1} x_0$$

$$\frac{\partial J}{\partial x} = 0$$

All "2" cancel.

$$(G^T R^{-1} G + B^{-1}) x = (G^T R^{-1} c + B^{-1} x_0)$$

The minimization

$$x = (G^T R^{-1} G + B^{-1})^{-1} (G^T R^{-1} c + B^{-1} x_0)$$

$$\boxed{x = x_0 + (G^T R^{-1} G + B^{-1})^{-1} G^T R^{-1} (c - G x_0)}$$

B is the Prior Error variance-covariance matrix.

$(G^T R^{-1} G + B^{-1})^{-1}$ is the posterior Error variance-covariance matrix.

$$VR = \frac{\text{Trace}(B) - \text{Trace}(G^T B^{-1} G + B^{-1})^{-1}}{\text{Trace}(B)}$$

The minimization

- Therefore, the minima of the cost function is obtained at a condition

$$\boxed{X = X_0 + \left(G^T R^{-1} G + B^{-1}\right)^{-1} G^T R^{-1} \left(C - G X_0\right)}$$

$$\boxed{X_a = X_b + [K]^{-1} (\text{data} - \text{model})}$$

The minimization

- The Uncertainty Reduction

  (i.e. Trace (B) – Trace (posterior_uncertainty))/Trace(B)

has no dependency on the observations [c]

This becomes the basis of the Observation System Simulation Experiment (OSSE)

G-Matrix for S-regions and T-towers

$$G = \begin{array}{cc} & \begin{array}{ccccccc} S1 & S2 & S3 & S4 & S5 & .. & SN \end{array} \\ \begin{array}{c} T1 \\ T2 \\ T3 \\ T4 \\ .. \\ TM \end{array} & \left[\begin{array}{ccccccc} \circ & \circ & & & & & \circ \\ \circ & \circ & & & & & \circ \\ \circ & \circ & & \left(\frac{\partial c}{\partial x} \right) & & & \circ \\ \circ & \circ & & & & & \circ \\ \vdots & \vdots & & & & & \vdots \\ \circ & \circ & & & & & \circ \end{array} \right]_{M \times N} \end{array}$$

R-Matrix for T-towers

$$
\mathbf{R} = \begin{array}{cc}
 & \begin{array}{ccccccc} T1 & T2 & T3 & T4 & T5 & .. & TM \end{array} \\
\begin{array}{c} T1 \\ T2 \\ T3 \\ T4 \\ T5 \\ .. \\ TM \end{array} &
\begin{bmatrix}
\sigma_1 & & & & & & \\
 & \sigma_2 & & & & & \\
 & & \sigma_3 & & & \text{covar} & \\
 & & & \sigma_4 & & & \\
 & & & & \sigma_5 & & \\
 & & \text{covar} & & & \ddots & \\
 & & & & & & \sigma_m
\end{bmatrix}_{M \times M}
\end{array}
$$

B-Matrix for the Background

$$\mathbf{B} = \begin{array}{cc} & \begin{array}{ccccccc} S1 & S2 & S3 & S4 & S5 & .. & SN \end{array} \\ \begin{array}{c} S1 \\ S2 \\ S3 \\ S4 \\ S5 \\ .. \\ SN \end{array} & \begin{bmatrix} \sigma_1 & & & & & & \\ & \sigma_2 & & & & & \text{covar} \\ & & \sigma_3 & & & \text{covar} & \\ & & & \sigma_4 & \nearrow & & \\ & & & \swarrow & \sigma_5 & \cdots & \\ & & \text{covar} & & & \ddots & \\ & \text{covar} & & & & & \sigma_N \end{bmatrix}_{N \times N} \end{array}$$

Inversion of Posterior Uncertainty

$$A = \left(G^T R^{-1} G + B^{-1}\right)^{-1}$$

$G = M \times N$

$R = M \times M$

$B = N \times N$

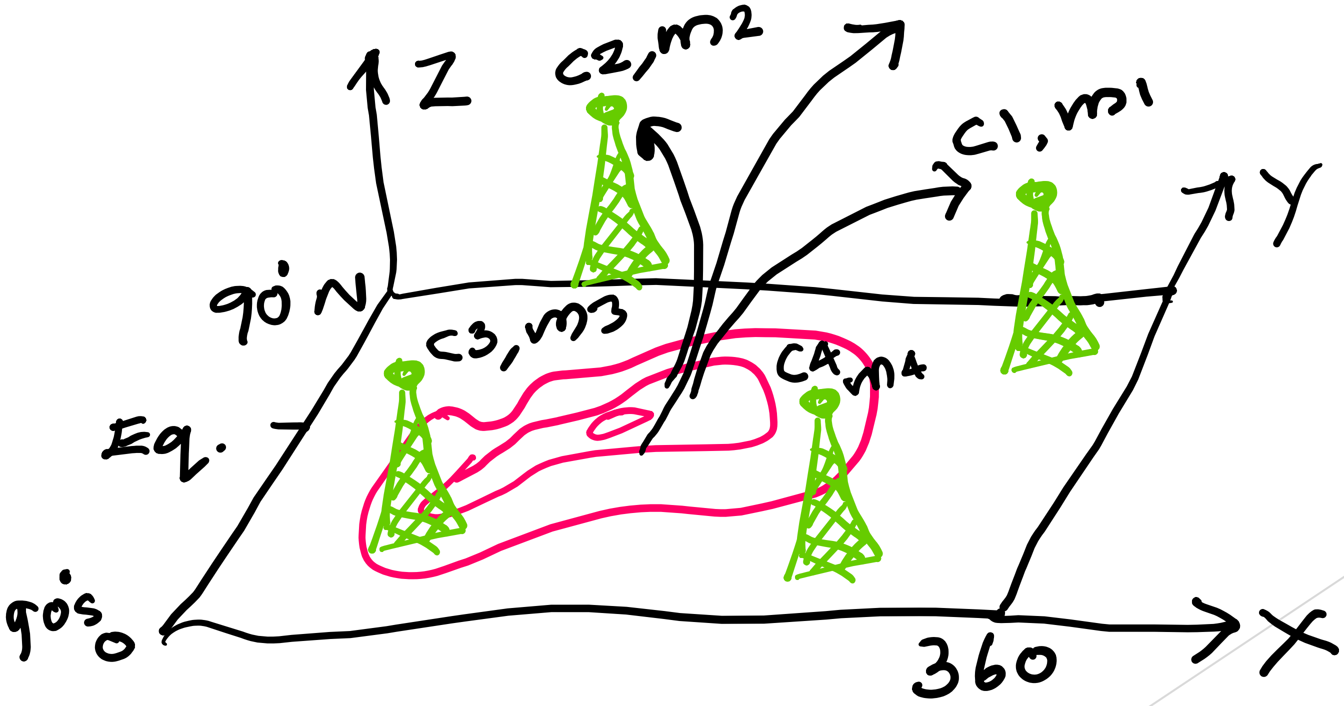
$$(N \times M)(M \times M) + N \times N$$

$$= N \times N$$

$$\boxed{UR = \frac{B - A}{B}}$$

Observation System Simulation Experiment

- Questions: Suppose you want to study a system, or want to do a Data Assimilation of a system, where are all the observations you need to take?

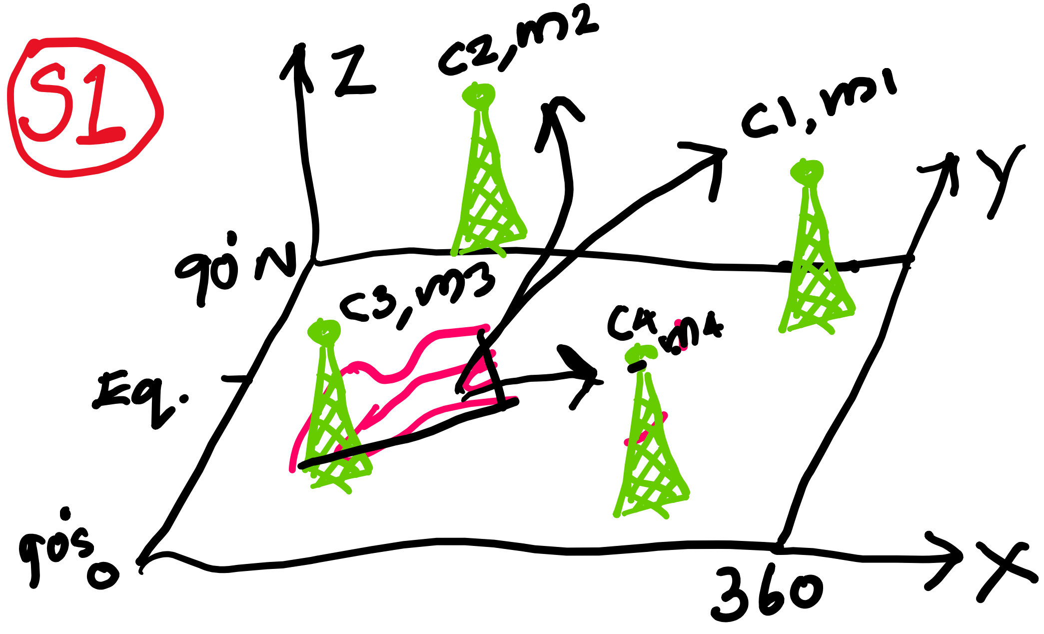


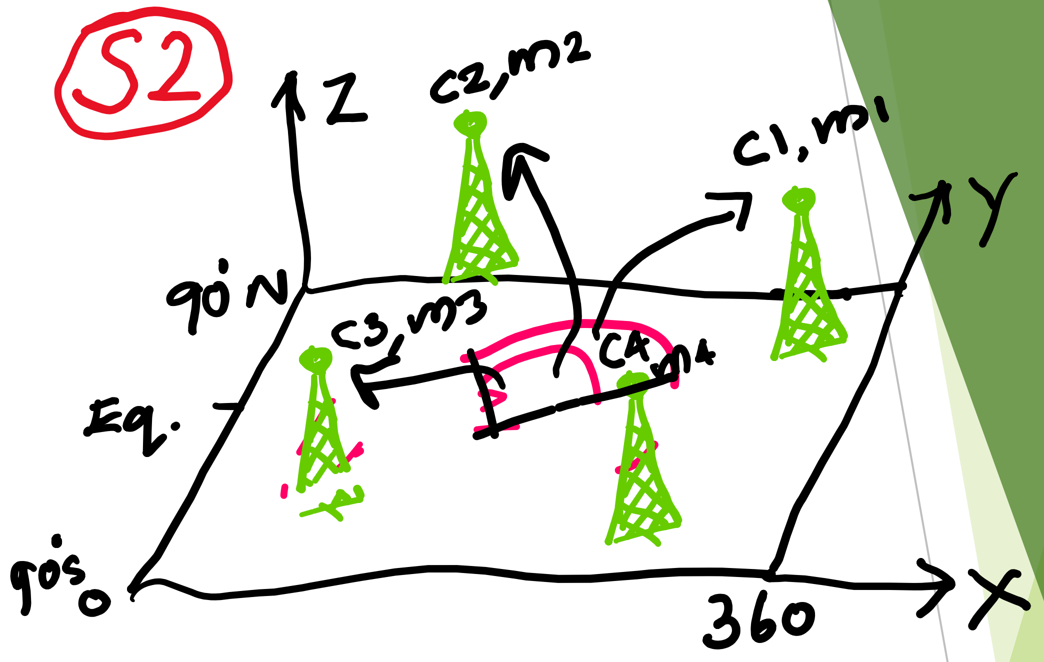
Observation System Simulation Experiment

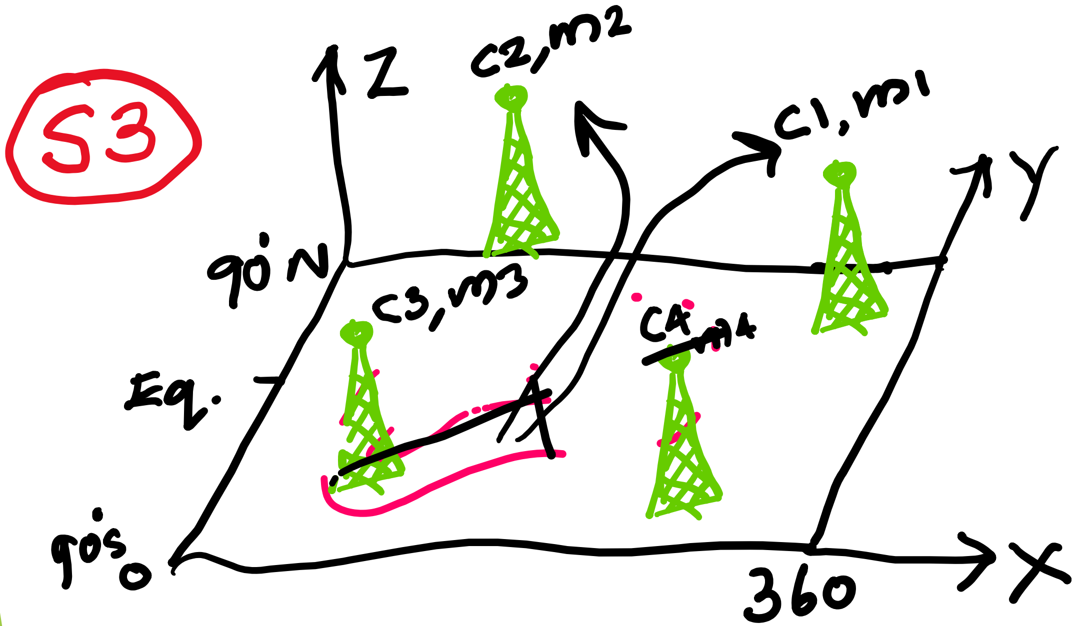
- One way of finding it is, by looking at, what is the contribution of an observation [c] to in reducing the posterior uncertainty in the data assimilation system.

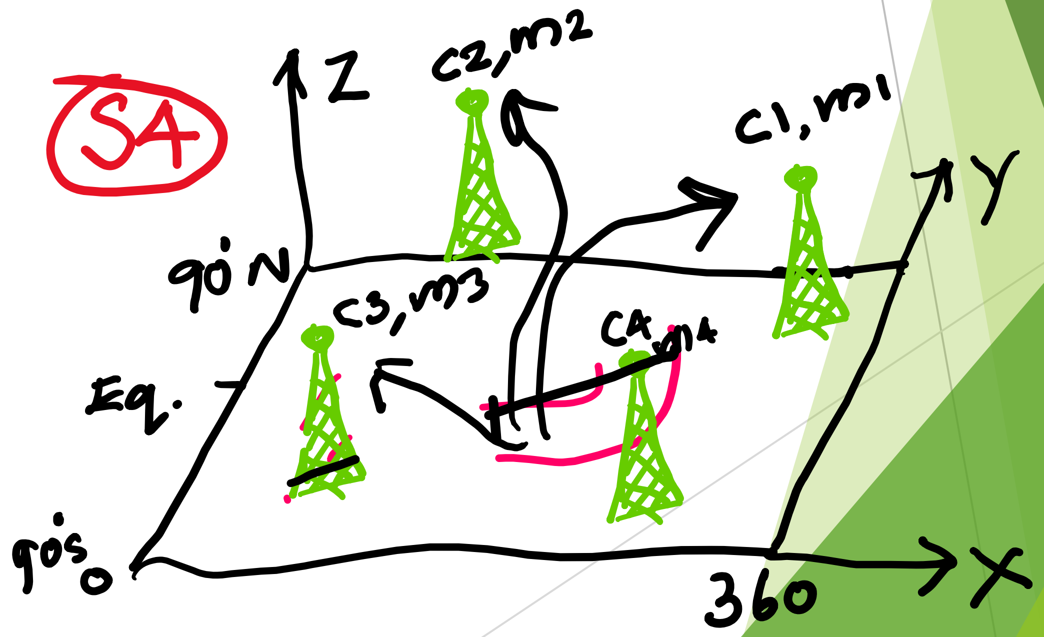
$$U.R. = \frac{\mathrm{Trace}(B) - \mathrm{Trace}\left(G^T R^{-1} G + B^{-1}\right)^{-1}}{\mathrm{Trace}(B)}$$

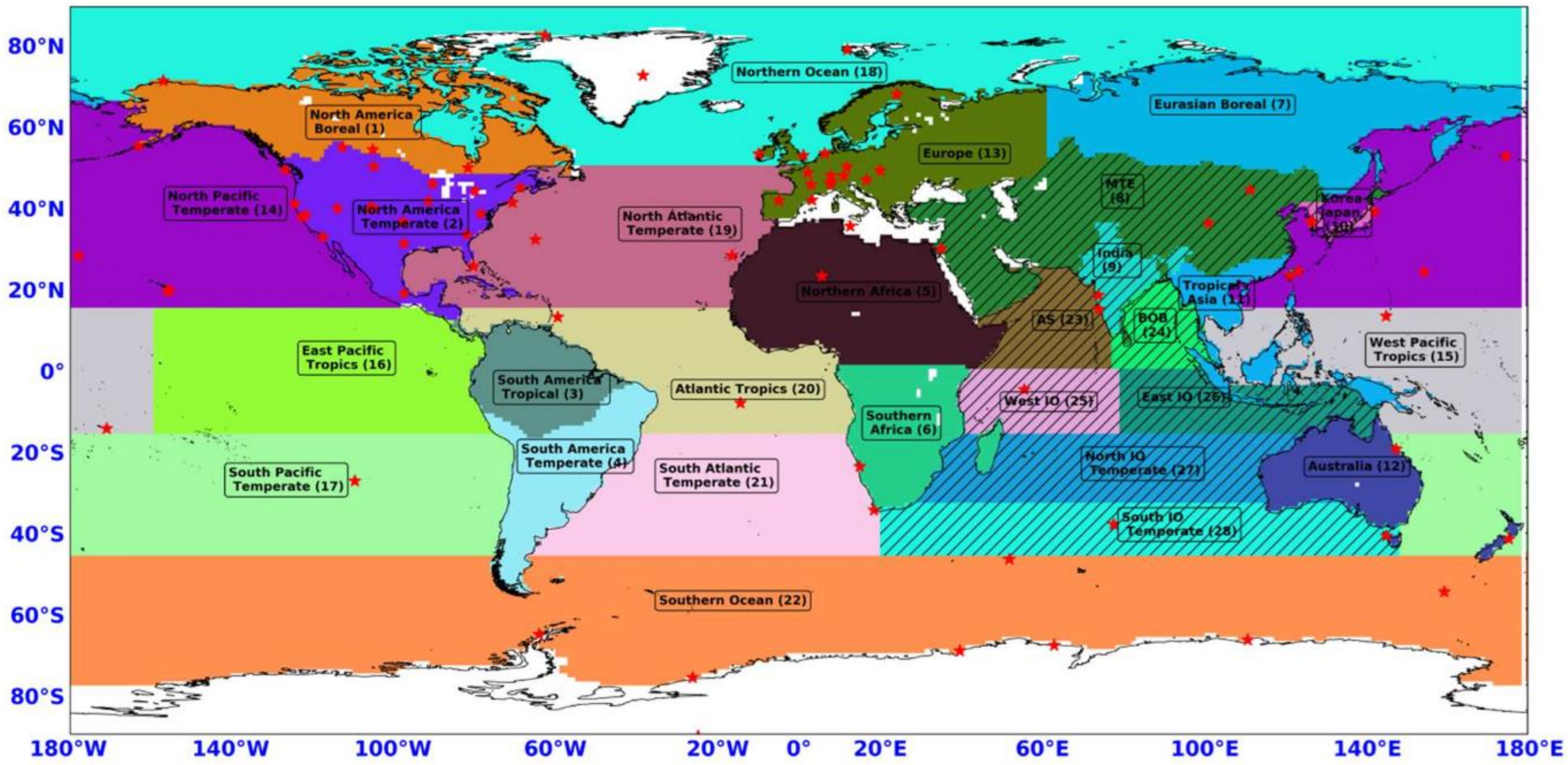
- To find this, you need [G], [B], and [R].
- How to construct these in the real world problems.


S1
Z
C2, m2
C1, m1
Y
90°N
C3, m3
C4, m4
Eq.
90°S
0
360
X


S2
Z
C2, m2
C1, m1
Y
90°N
C3, m3
C4, m4
Eq.
90°S
0
360
X

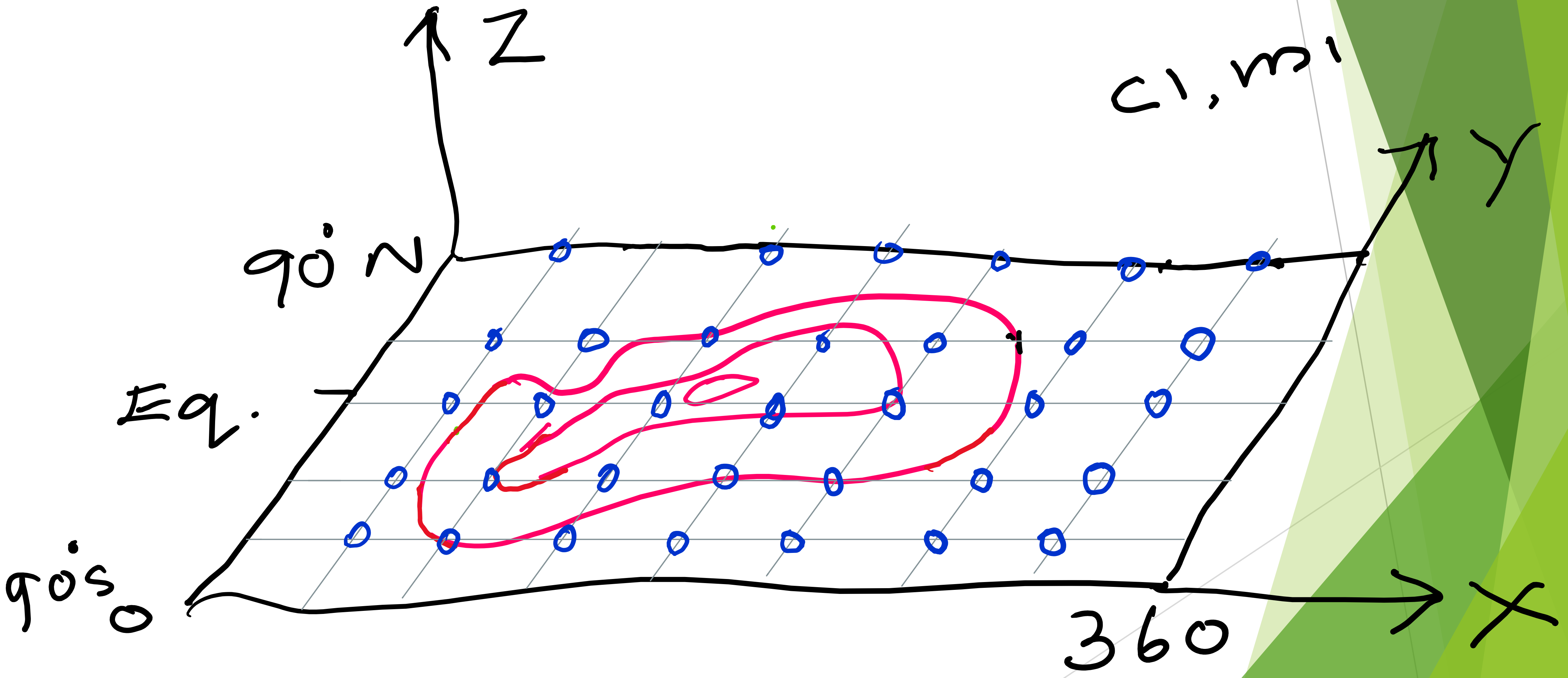

S3
Z
C2, m2
C1, m1
Y
90°N
C3, m3
C4, m4
Eq.
90°S
0
360
X

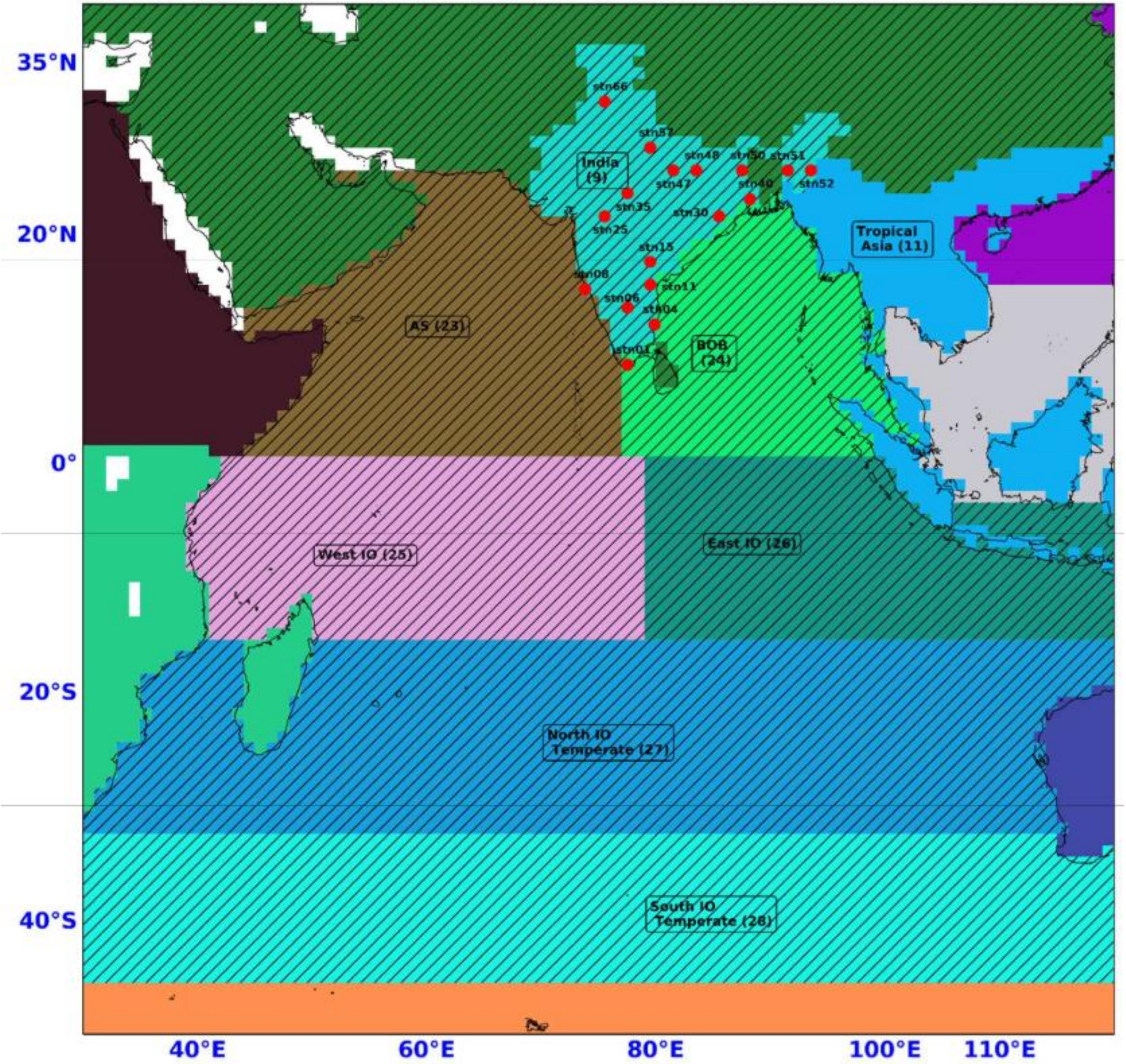

S4
Z
C2, m2
C1, m1
Y
90°N
C3, m3
C4, m4
Eq.
90°S
0
360
X


80°N
60°N
40°N
20°N
0°
20°S
40°S
60°S
80°S
180°W
140°W
100°W
60°W
20°W
0°
20°E
60°E
100°E
140°E
180°E
Northern Ocean (18)
North America Boreal (1)
Eurasian Boreal (7)
Europe (13)
North Pacific Temperate (14)
North America Temperate (2)
North Atlantic Temperate (19)
MTE (8)
Korea Japan (10)
India (9)
Tropical Asia (11)
Northern Africa (5)
AS (23)
BOB (24)
West Pacific Tropics (15)
East Pacific Tropics (16)
South America Tropical (3)
Atlantic Tropics (20)
Southern Africa (6)
West IO (25)
East IO (26)
South America Temperate (4)
North IO Temperate (27)
Australia (12)
South Pacific Temperate (17)
South Atlantic Temperate (21)
South IO Temperate (28)
Southern Ocean (22)

Incremental Optimization

o = Potential observation





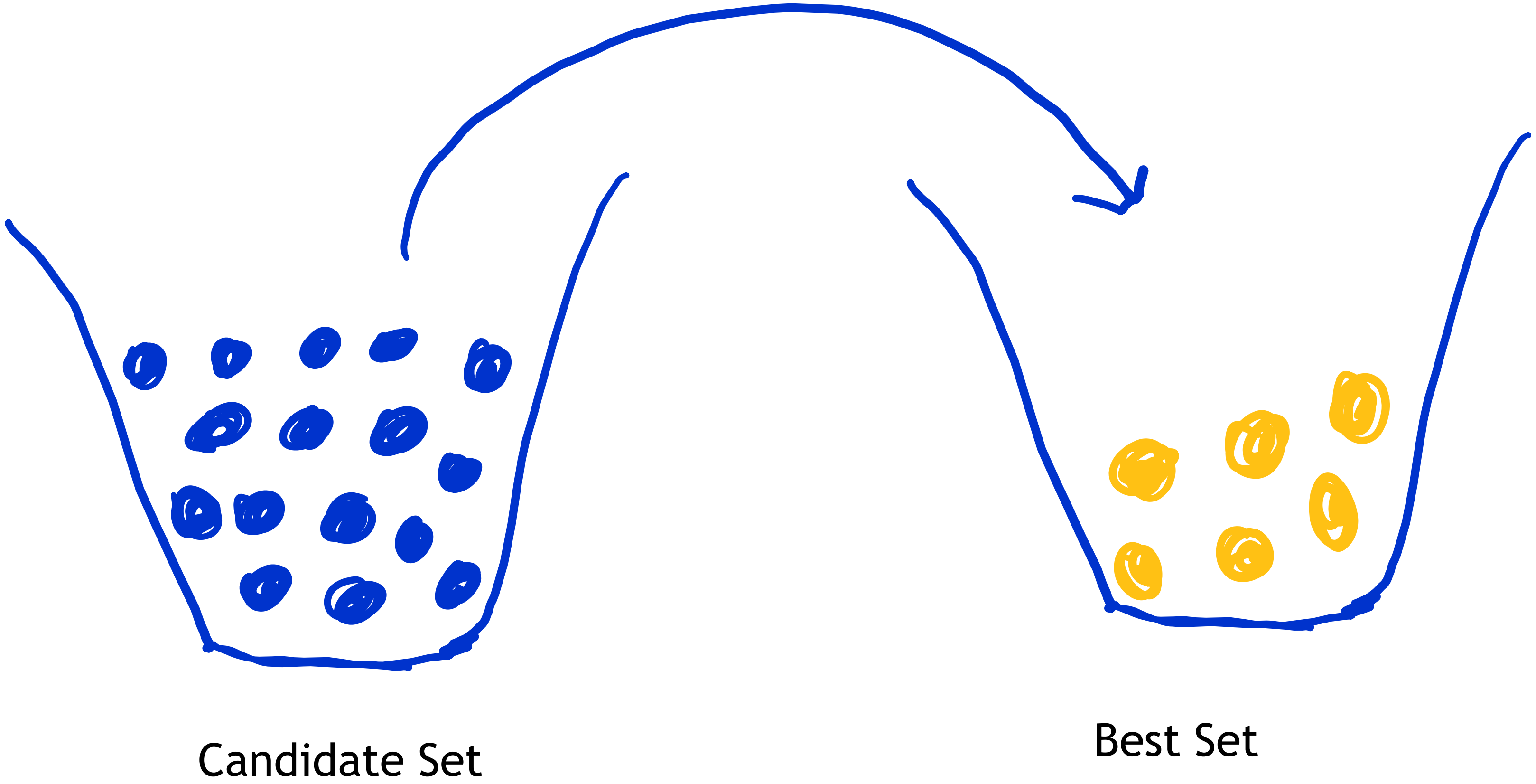
35°N
20°N
0°
20°S
40°S
40°E
60°E
80°E
100°E
110°E
India (9)
Tropical Asia (11)
AS (23)
BOB (24)
West IO (25)
East IO (26)
North IO Temperate (27)
South IO Temperate (28)
stn66
stn57
stn48
stn50
stn51
stn47
stn40
stn52
stn35
stn30
stn25
stn15
stn08
stn11
stn06
stn04
stn01

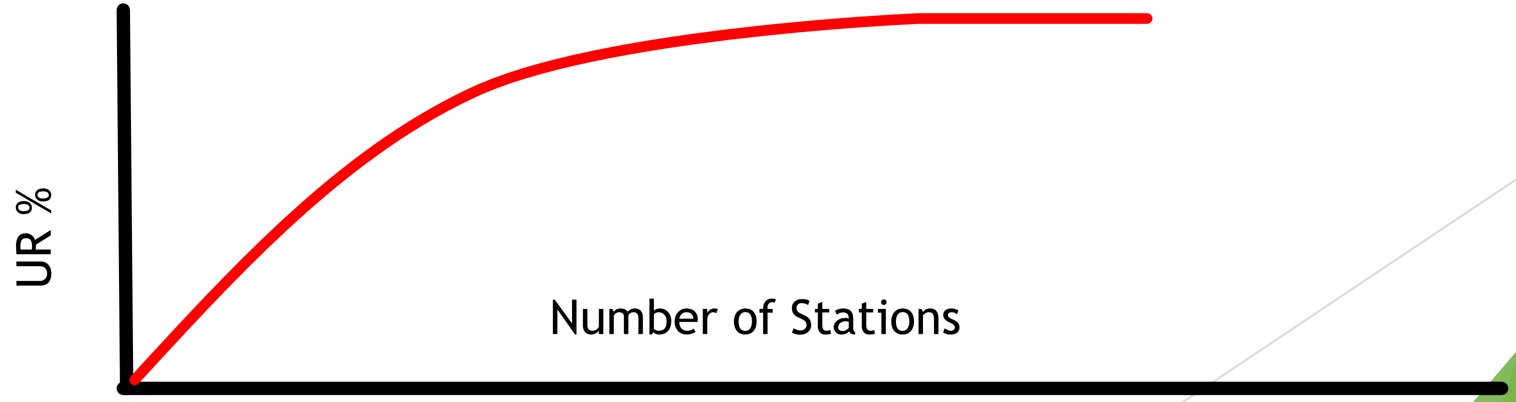
Incremental Optimization for OSSE

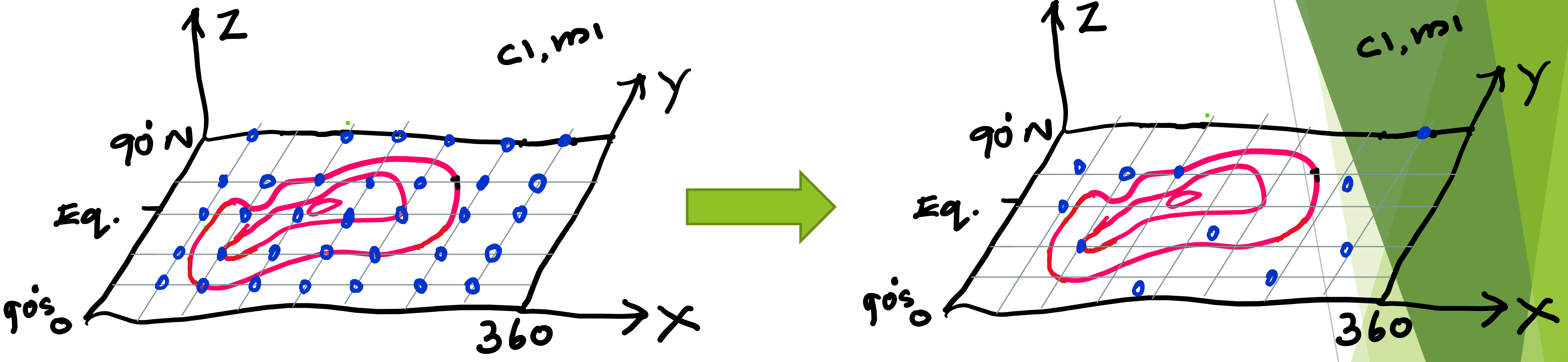
- For each observation location, calculate posterior-uncertainty (A) with
  - G = (1 x N) , R = (1 x 1) , B = (N x N) and find out which potential location of observation has maximum Uncertainty Reduction

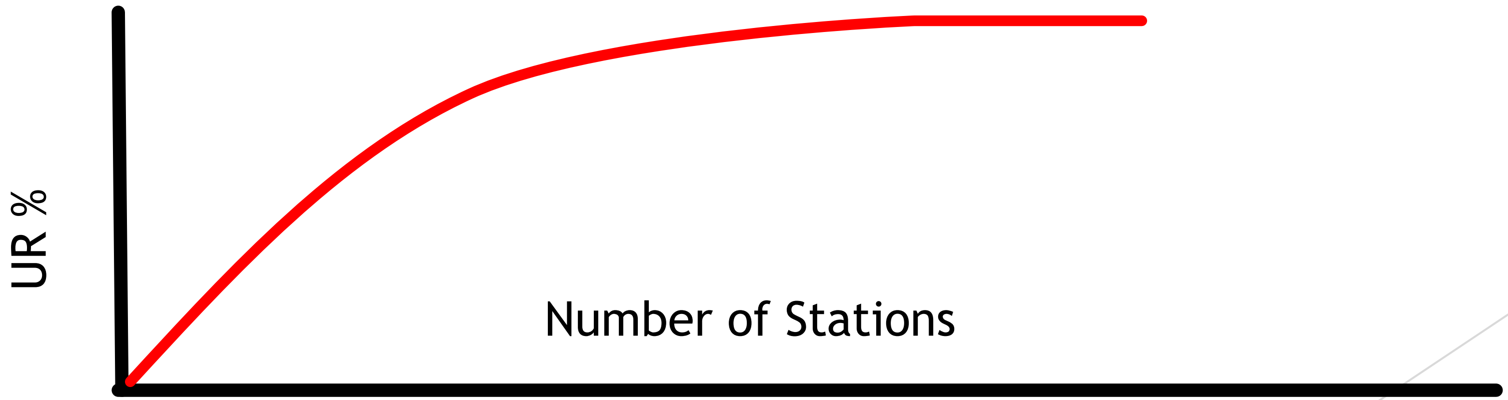
    (UR = Trace (B) – Trace (A))/Trace (B) and find out the Best Observation Location

- Repeat the processes with a combination of any other location with the above location in step 1 and find the following
  - G = (2 x N) , R = (2 x 2) , B = (N x N) and find out which combination of potential location of observation, together with the observation location from the first iteration, gives maximum UR and mark it as the best two observation locations.
- Repeat this processes until a combination of observation locations which results in a maximum UR.


Candidate Set
Best Set


UR %
Number of Stations


Z
cl,m
Y
90 N
Eq.
90 S
0
360
X


UR %
Number of Stations

Understanding on Variational Assimilation and Bayesian inversion and their equavalences

Cost Function.

$$J(x) = (\hat{x} - x_o)^2 C_S^{-1} (\hat{x} - x_o) + (G x_o - D) C_d^{-1} (G x_o - D)$$

The Bayesian inversion suggests that

$$\hat{x} = x_o + \{G^T C_D^{-1} G + C_S^{-1}\}^{-1} [G^T C_D^{-1}] \{D - G x_o]$$

Where G is response Function or Green's Function

C_D is observational Variances or Errors

C_S is prior variance-covarience Matrix.

D is the observation.

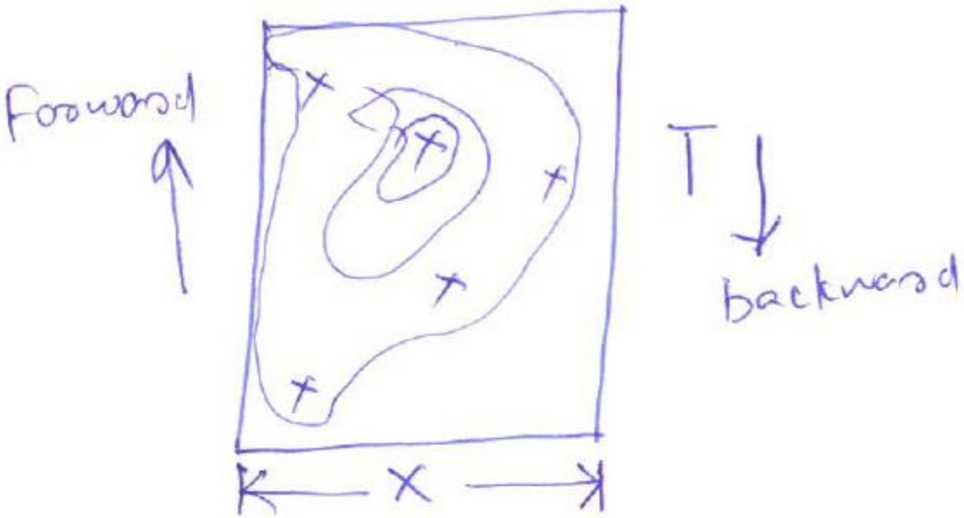
$G x_o$ is the model.

4D-Var.

$$J(x) = (\hat{x} - x_0)^2 C_s^{-1} (\hat{x} - x_0) + (H(x_{0t}) - D) C_d^{-1} (H(x_{0t}) - D).$$

$$\hat{x} = x_0 + \left[G^T C_D^{-1} G + C_s^{-1}\right]^{-1} \left[G^T C_D^{-1}\right] \lambda(x, T)$$

Where "λ" is the adjoint operator run backword in time with actual mode-data missfit and transfering data from observational space to model space.



(1)

G is a representer matrix which propagate impulse of observation in model space.

In a linear Advection Model, the 'G' contains elements of α_m in model space.

$$-\frac{\partial \alpha_m}{\partial t} - c\frac{\partial \alpha_m}{\partial x} = \delta(x-x_m)\,\delta(t-t_m)$$

with b.c. $\alpha_m(t,L) = 0$

with I.C. $\alpha_m(x,T) = 0$.

$\therefore \; G = \alpha_m(x,t)$.

In 4D-var assimilation, the optimized field

$$\hat{X} = X_0 + W_i^{-1} \lambda(x,t)$$

Therefore W_i^{-1} is equivalent to

$$\left[G^T C_D^{-1} G + C_s^{-1}\right]^{-1} \left[G^T C_D^{-1}\right] \quad \text{—— (1)}$$

The W_i and (1) are unitless. It represent a positive weight or number.

C_S is prior uncertainty

$\left[G^T C_D^{-1} G + C_s^{-1}\right]^{-1}$ is posterior uncertainty.

Summary

- Topics covered
  - Bayesian Theory,
  - Cost Function,
  - Minimization,
  - Incremental Optimization,
  - Observation System Simulation Experiment

- Further Reading:
  - Andrew Bennet, Inverse Problems on Ocean Modelling, Cambridge University Press
  - I .G. Enting, Inverse Problems in Atmospheric Constituent Transport, Cambridge University Press
  - Valsala et al., (2021), Observational System Simulation Experiment for Indian Ocean pCO2 measurements, Progress in Oceanography