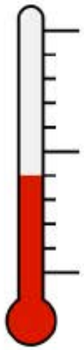
Introduction to sea ice data assimilation with DART/CICE

Yongfei Zhang and Cecilia Bitz[1]
Jeffrey Anderson, Nancy Collins, Jonathan Hendricks, Tim Hoar, and Kevin Raeder[2]
[1]University of Washington, Seattle, WA
[2]National Center for Atmospheric Sciences, Boulder, CO

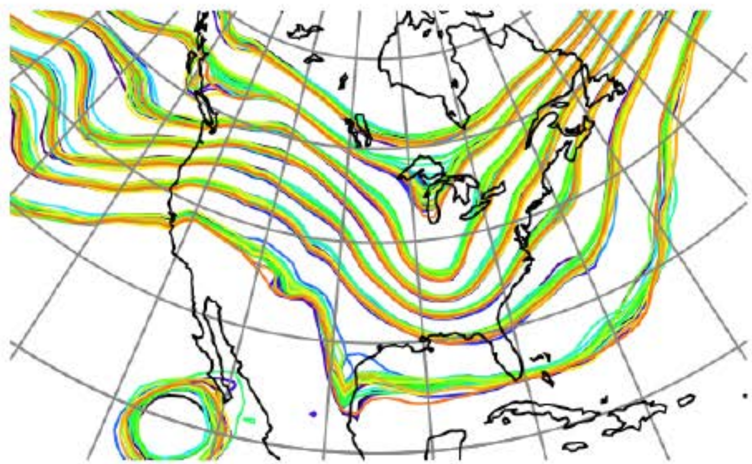
What is Data Assimilation?

Observations combined with a Model forecast…



+

- Reconstruct states of climate in the past
- Provide more accurate initial conditions for prediction

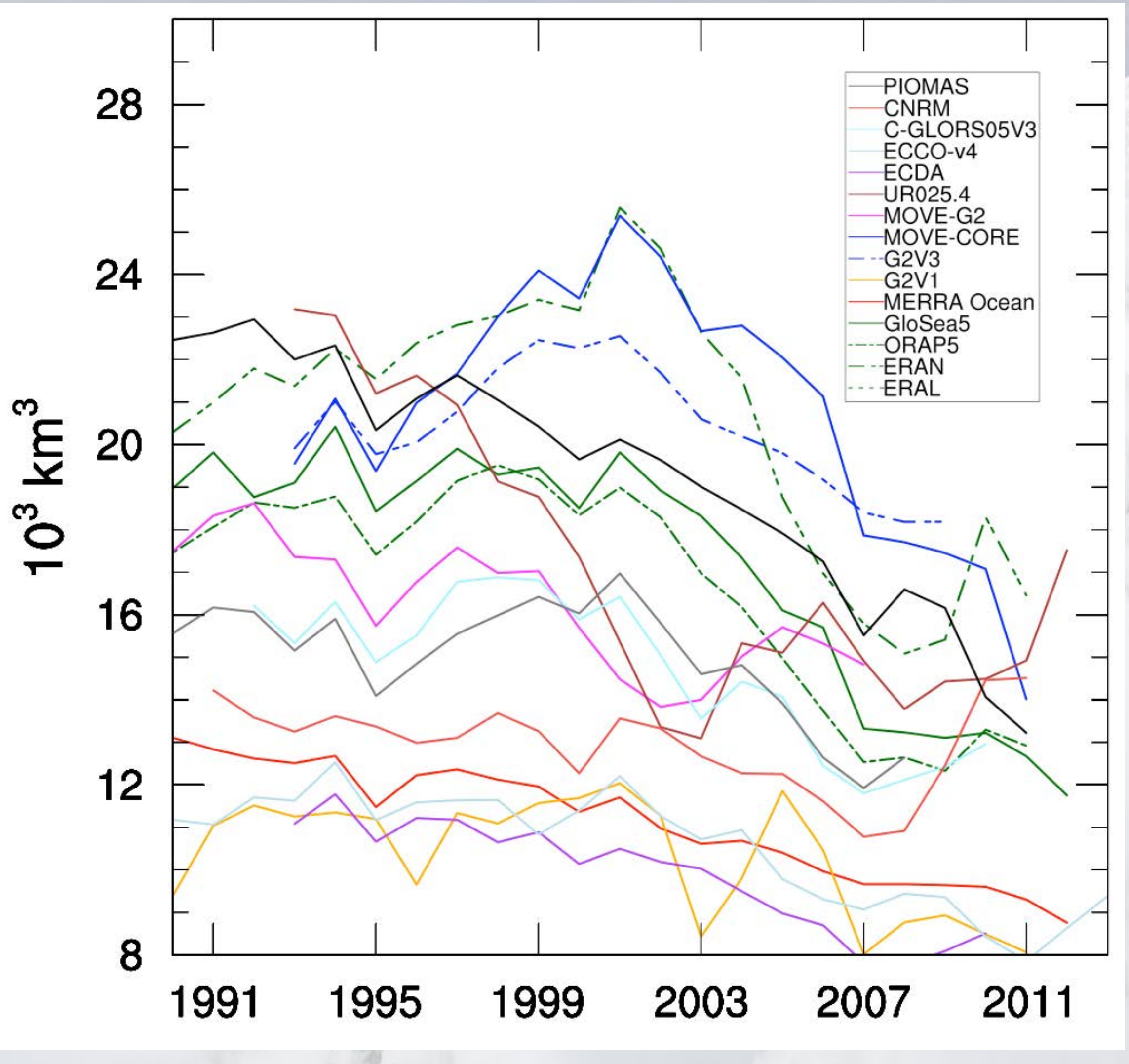


…to produce an analysis
(best possible estimate).

Borrowed from Anderson et al., 2013

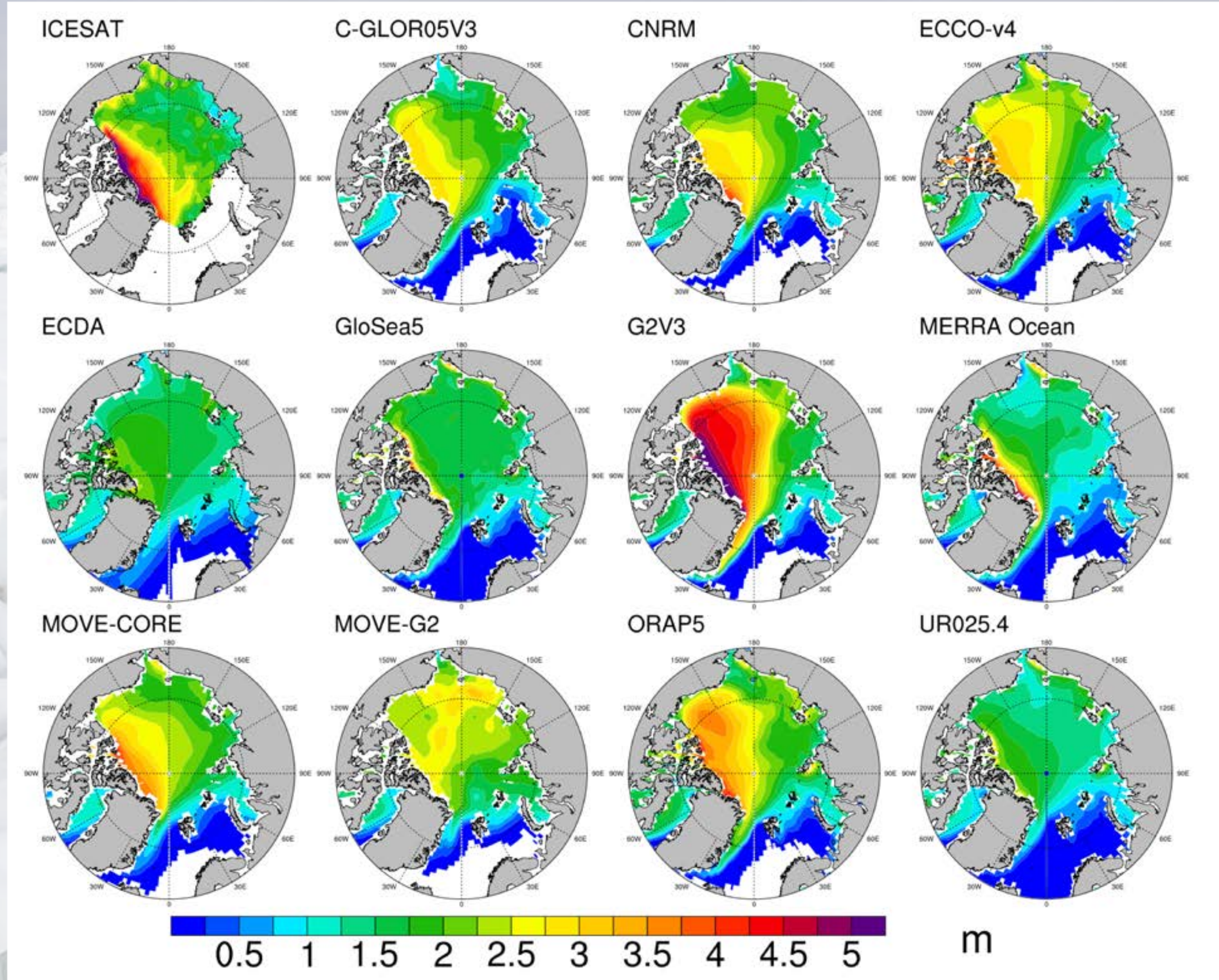
Uncertainty in sea ice reanalysis/reconstruction products (from which initial conditions are taken)

Annual volume of sea ice



Chevallier et al (2016)

Errors in sea ice reanalysis/reconstruction (from which initial conditions are taken)



Mean March 2003-2007 Sea Ice Thickness (m) in global ocean-sea ice reanalyses with assimilation of sea ice concentration

Chevallier et al (2016)

Basic theory of the ensemble Kalman filter (EnKF)

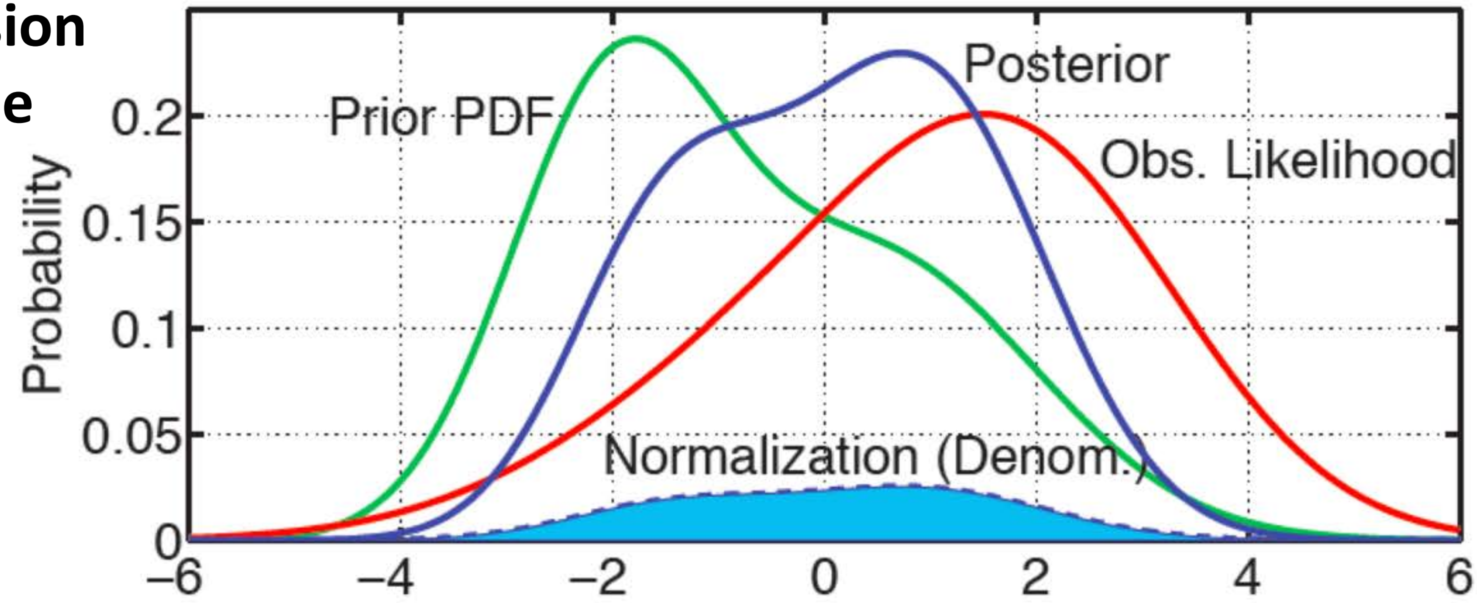
Check out the DART tutorial

https://www.image.ucar.edu/DAReS/DART/Manhattan/documentation/tutorial/

Bayes' Rule

$$p(A|BC) = \frac{p(B|AC)p(A|C)}{p(B|C)} = \frac{p(B|AC)p(A|C)}{\int p(B|x)p(x|C)dx}$$

A one-dimension example



A : Prior Estimate based on all previous information, C.

B : An additional observation.

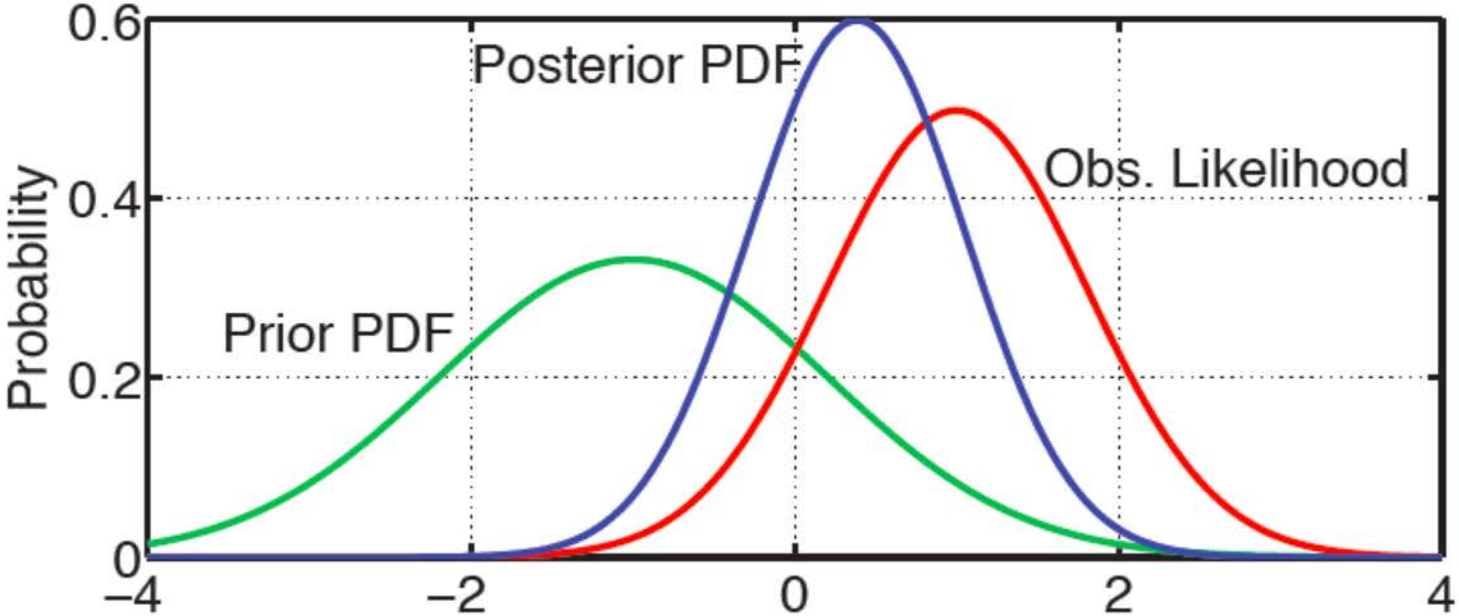
$p(A|BC)$: Posterior (updated estimate) based on C and B.

Product of Two Gaussians

$$p(A \mid BC) = \frac{p(B \mid AC)\,p(A \mid C)}{p(B \mid C)} = \frac{p(B \mid AC)\,p(A \mid C)}{\int p(B \mid x)\,p(x \mid C)\,dx}$$

A one-dimension example

This product is closed for Gaussian distributions.



Covariance: $\Sigma = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}$

Mean: $\mu = (\Sigma_1^{-1} + \Sigma_2^{-1})^{-1}(\Sigma_1^{-1}\mu_1 + \Sigma_2^{-1}\mu_2)$

The EnKF

$$\mathbf{x}_{\mathrm{t}}^{\mathrm{a}} = \mathbf{x}_{\mathrm{t}}^{\mathrm{b}} + \mathrm{K}\left(\mathbf{y}_{\mathrm{t}} - \mathcal{H}\left(\mathbf{x}_{\mathrm{t}}^{\mathrm{b}}\right)\right)$$

$$\mathbf{K} = \mathbf{P}_{\mathrm{t}}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{P}_{\mathrm{t}}^{\mathrm{b}}\mathbf{H}^{\mathrm{T}} + \mathrm{R}\right)^{-1}$$

$$\mathbf{P}_{\mathrm{t}}^{\mathrm{a}} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{P}_{\mathrm{t}}^{\mathrm{b}}$$

- In KF, error grows linearly
- It requires tremendous computational cost to forecast the background error covariance

Approximate the pdf by a finite sample

$$\mathbf{X}^{\mathrm{b}} = \left(\mathbf{x}_1^{\mathrm{b}}, \ldots, \mathbf{x}_m^{\mathrm{b}}\right)$$

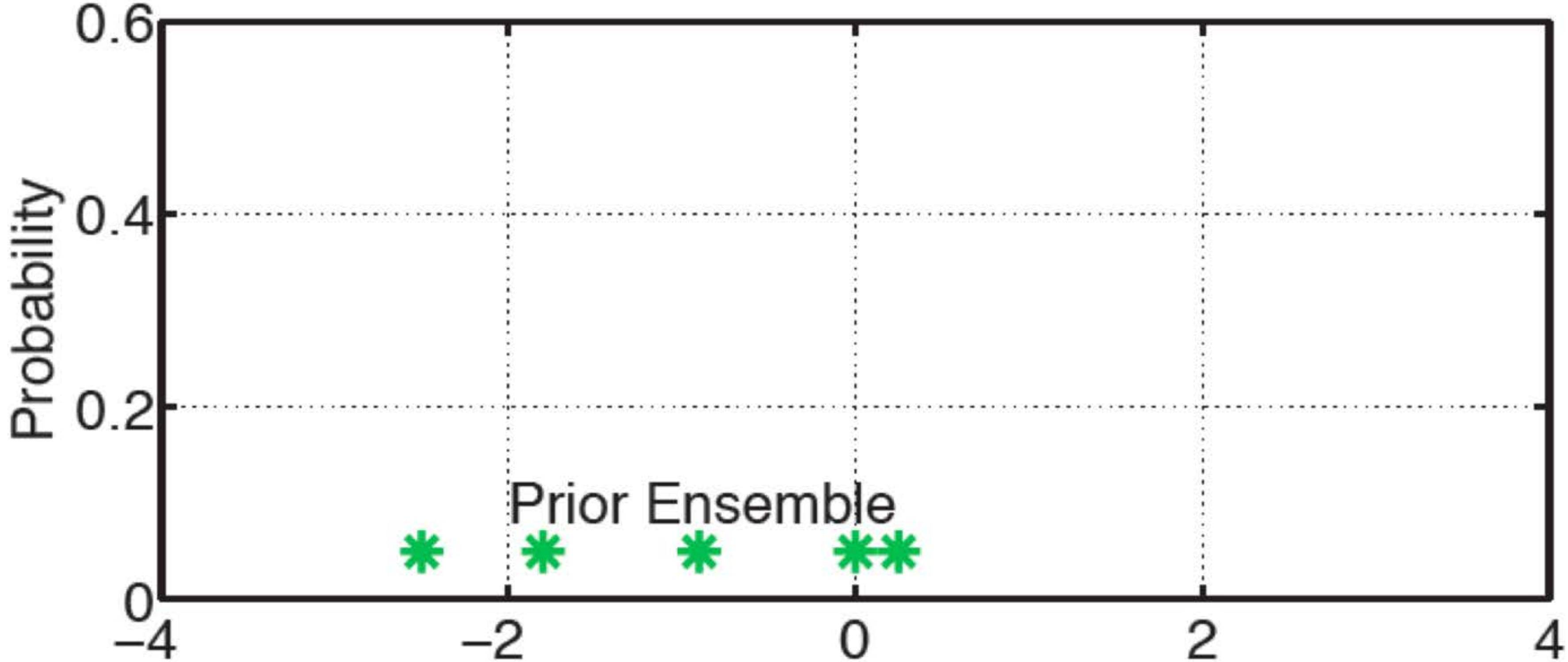
$$\overline{\mathbf{x}}^{\mathrm{b}} = \frac{1}{m}\sum_{i=1}^{m}\mathbf{x}_i^{\mathrm{b}}$$

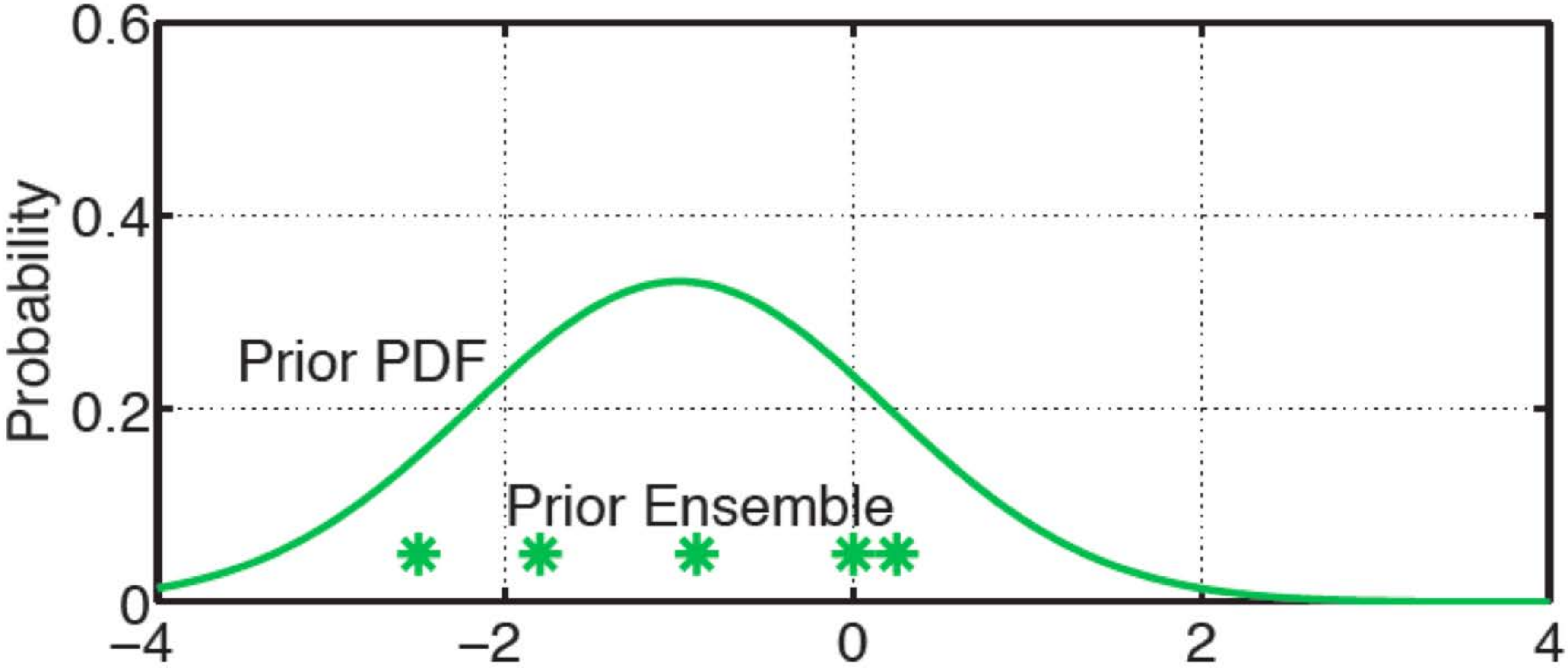
$$\mathbf{X}'^{\mathrm{b}} = \left(\mathbf{x}_1'^{\mathrm{b}}, \ldots, \mathbf{x}_m'^{\mathrm{b}}\right)$$

$$\hat{\mathbf{P}}^{\mathrm{b}} = \frac{1}{m-1}\mathbf{X}'^{\mathrm{b}}\mathbf{X}'^{\mathrm{b}^{\mathrm{T}}}$$

- Background error covariances are modeled using the ensemble of non-linear forecasts
- Reduces computational cost

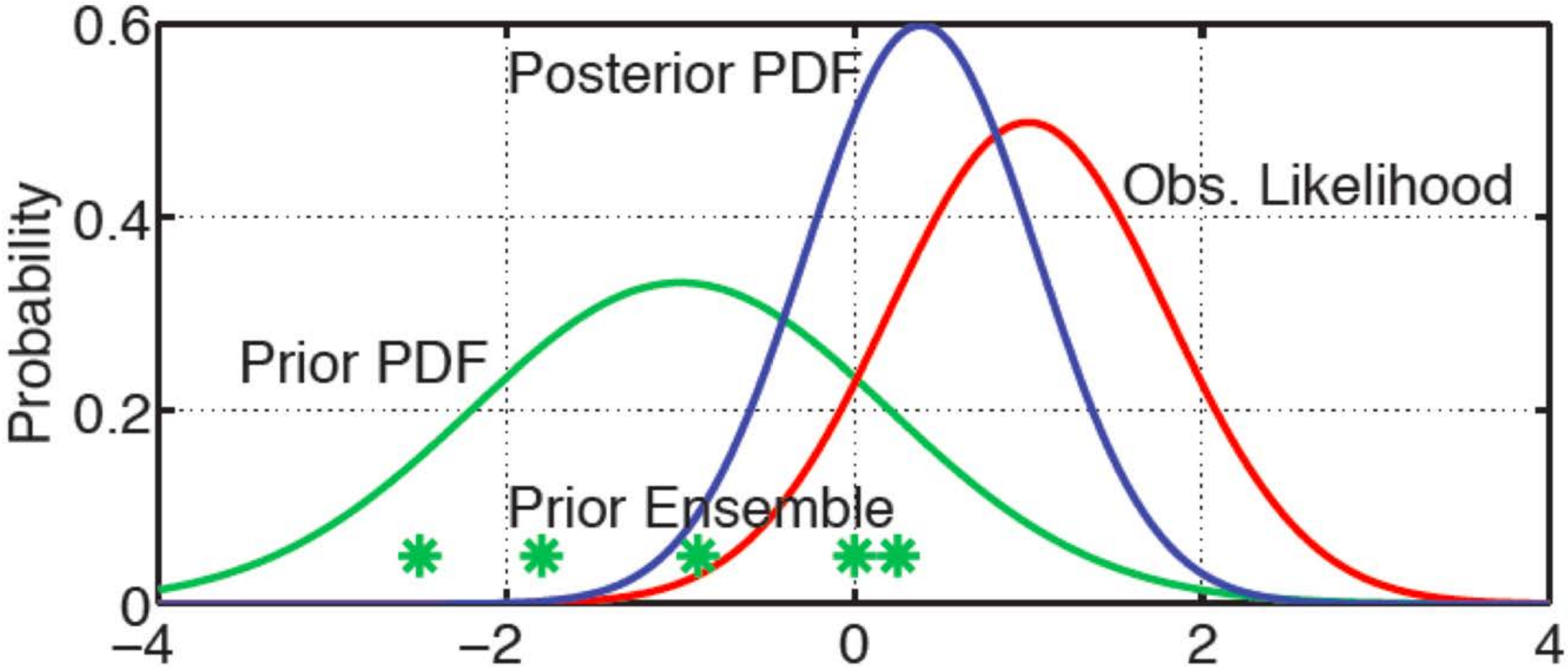
Ensemble filters: Prior is available as finite sample.





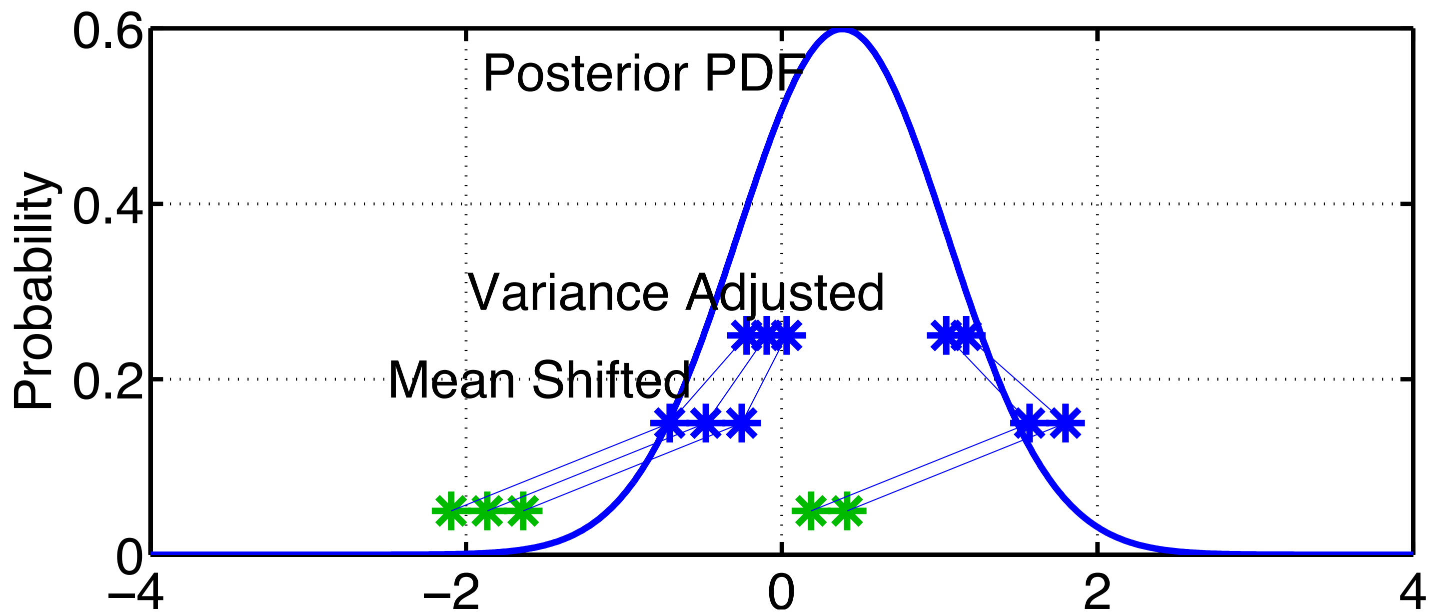
Fit a continuous (Gaussian for now) distribution to sample.

Product of prior Gaussian fit and Obs. likelihood is Gaussian.



Computing continuous posterior is simple.
BUT, need to have a SAMPLE of this PDF.

Ensemble Adjustment (Kalman) Filter



$$x_i^u = \left(x_i^p - \overline{x}^p\right) \bullet \left(\sigma^u / \sigma^p\right) + \overline{x}^u$$

i = 1,..., ensemble size.

p is prior,

u is update (posterior),

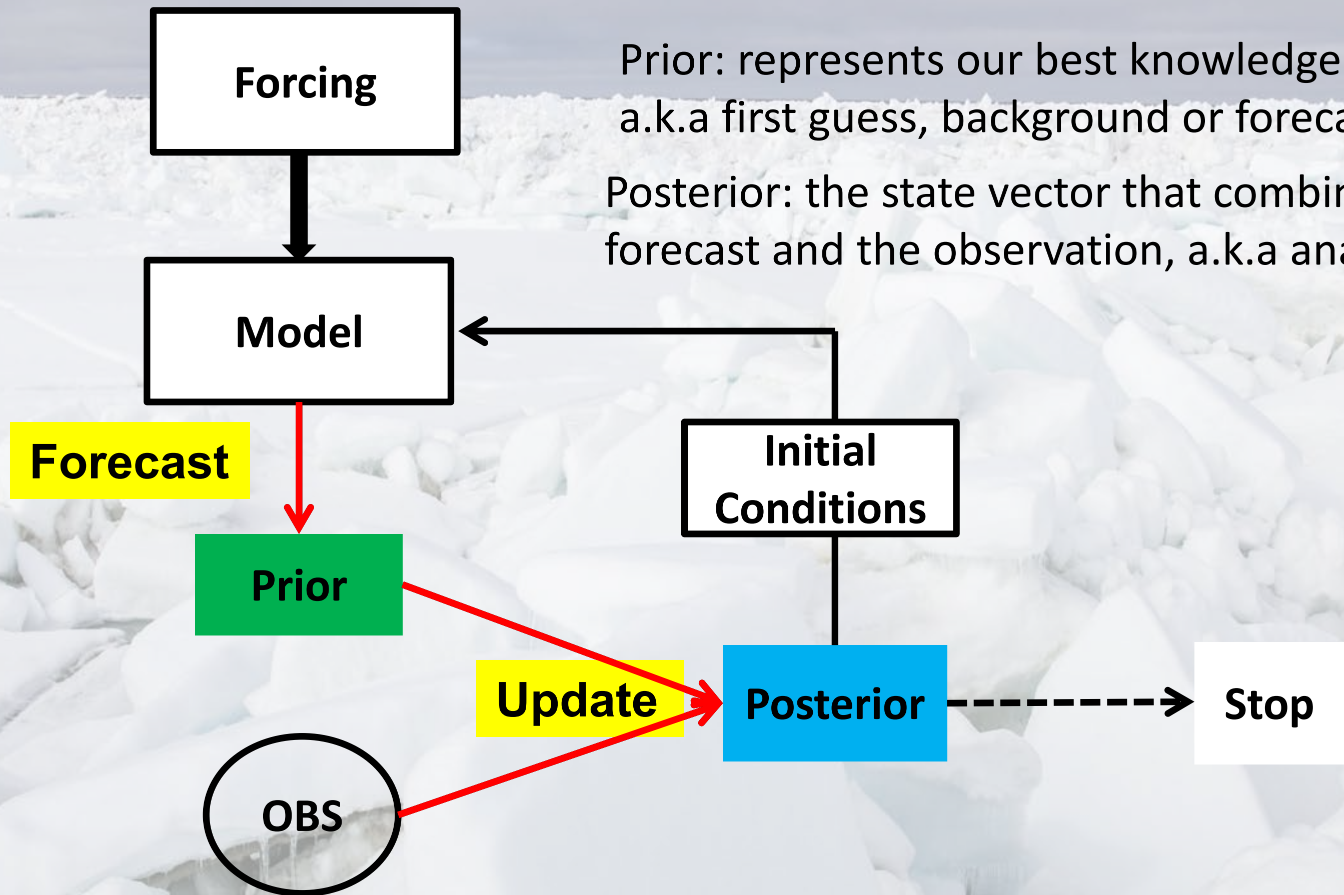
σ is standard deviation,

overbar is ensemble mean.

Data assimilation: to combine physical model simulations with observations

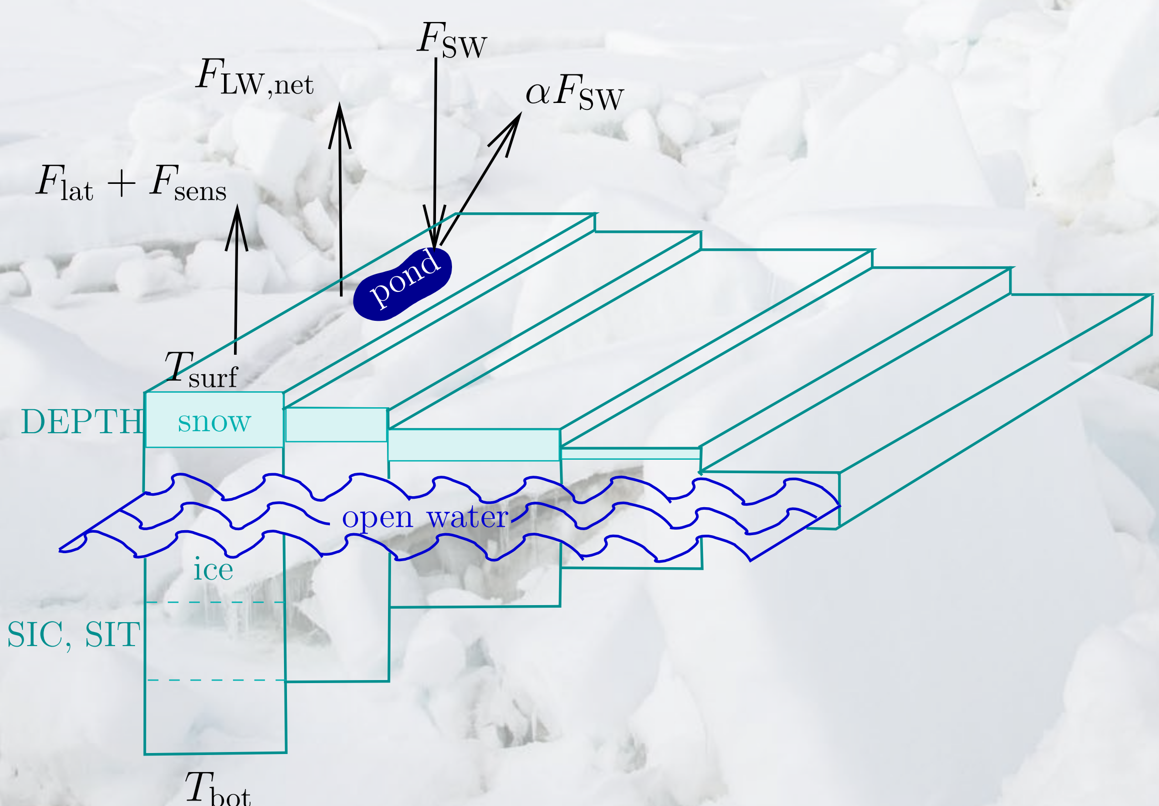
Prior: represents our best knowledge of the system, a.k.a first guess, background or forecast.

Posterior: the state vector that combines model's forecast and the observation, a.k.a analysis.



The Sea Ice Model

- The Los Alamos sea ice model version 5 (CICE5)
  - Contains a thermodynamic model and a dynamic model
  - Multiple ice thickness categories



Sea ice concentration (SIC)
Sea ice thickness (SIT)

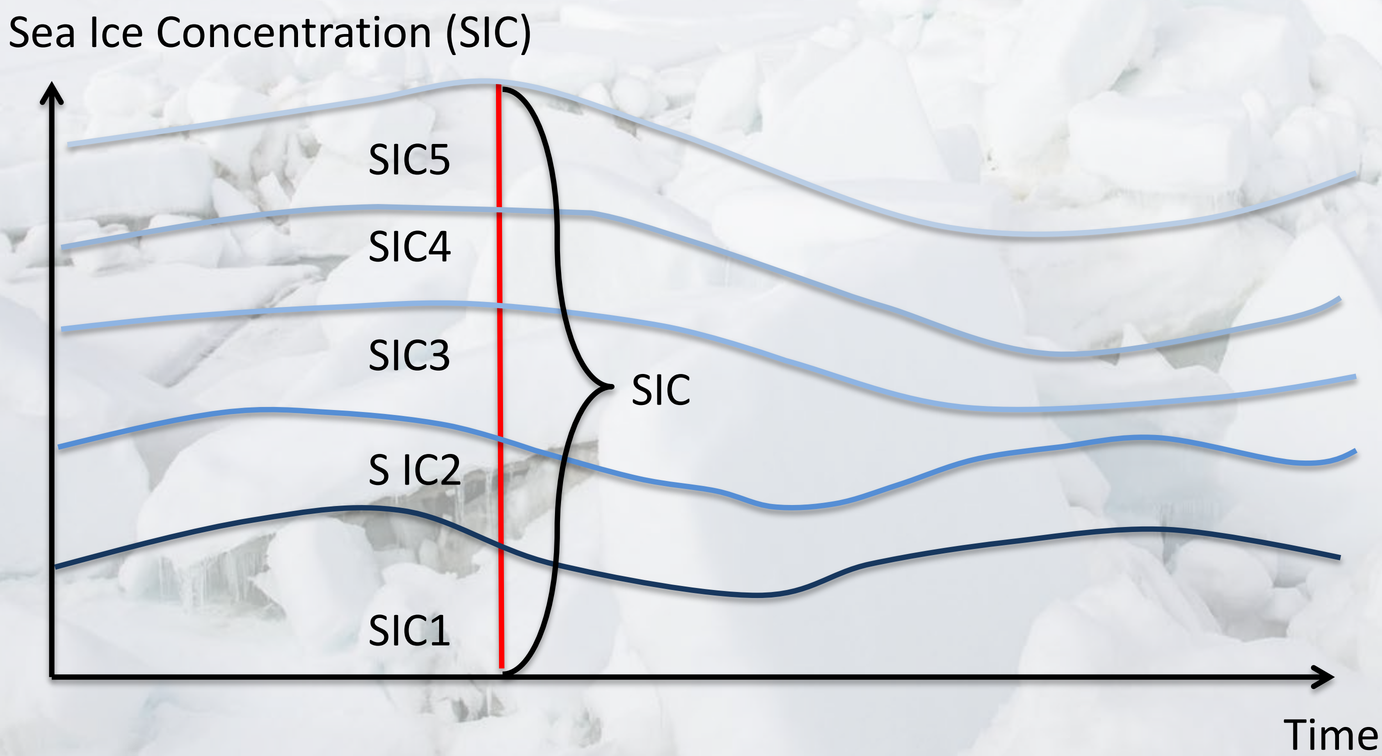
$$SIC = \sum_{i=1}^{N} SIC_i$$

$$SIT = \frac{\sum_{i=1}^{N} SIC_i \cdot SIT_i}{\sum_{i=1}^{N} SIC_i}$$

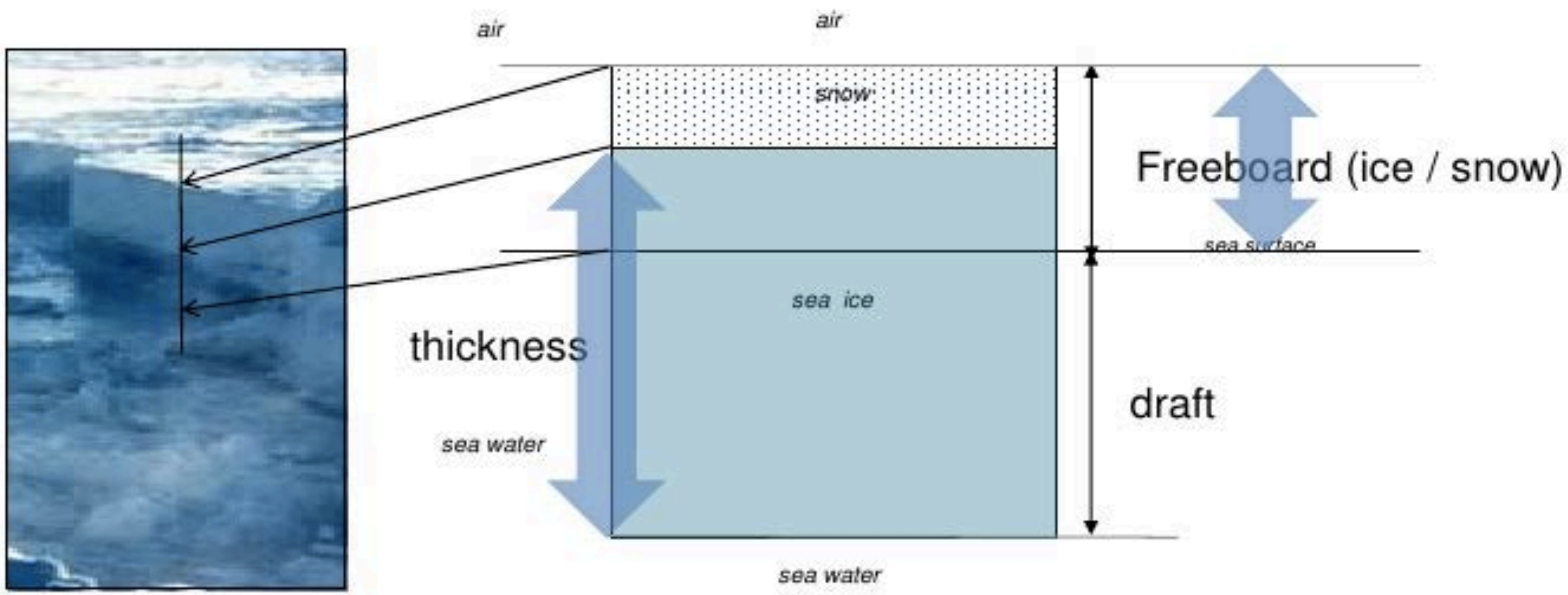
Observations

- In-situ observations are usually point-wise

- A satellite gives an “aggregate” estimate for all ice types at its resolution



Freeboard



Credit: Ron Kwok, NASA/JPL

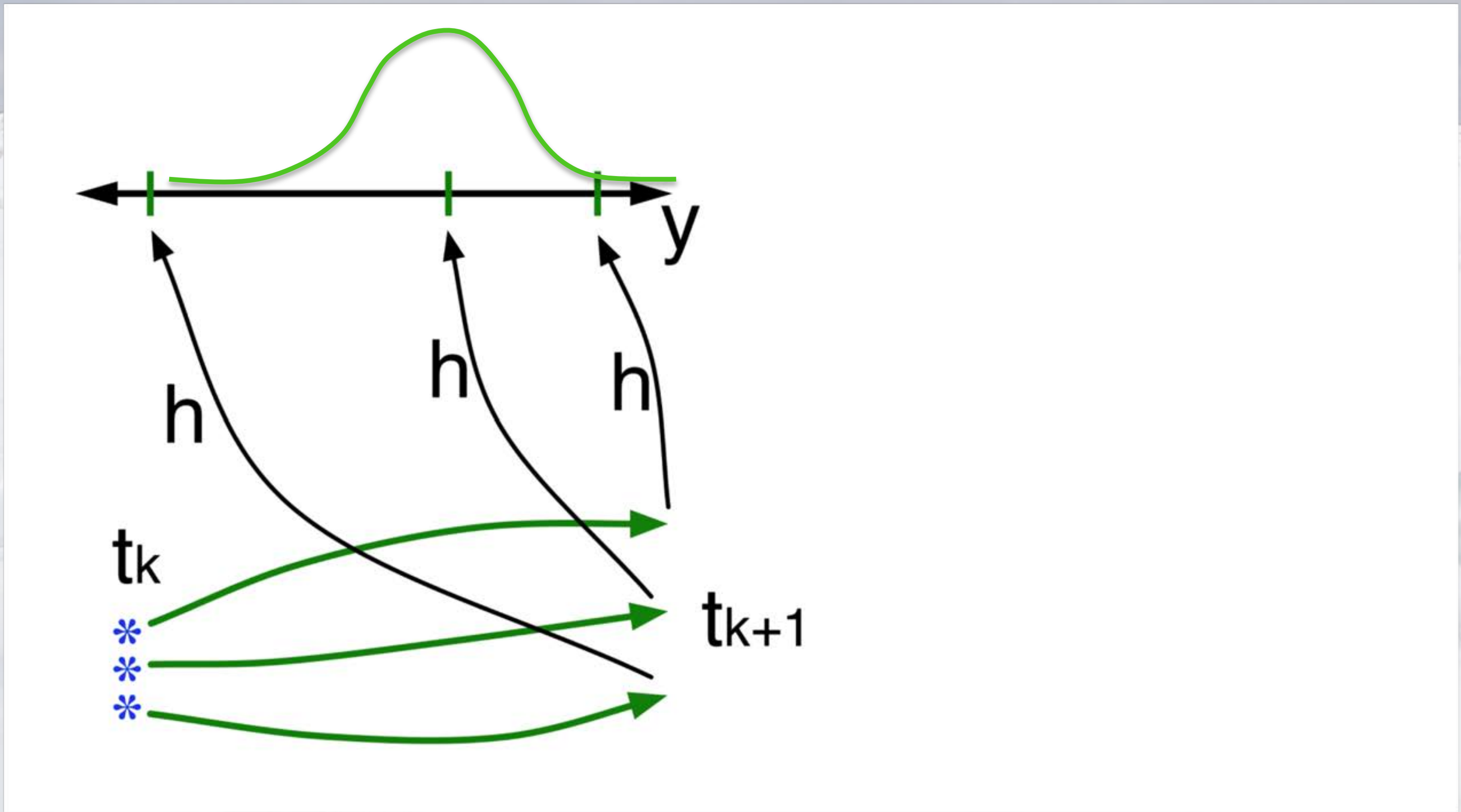
Sea ice thickness (SIT) can be retrieved from freeboard given snow/ice thicknesses and densities.

Flowchart of the Ensemble Kalman Filter

The model advances to the time step at which the observation is available

http://www.image.ucar.edu/DAReS/DART/

Flowchart of the Ensemble Kalman Filter

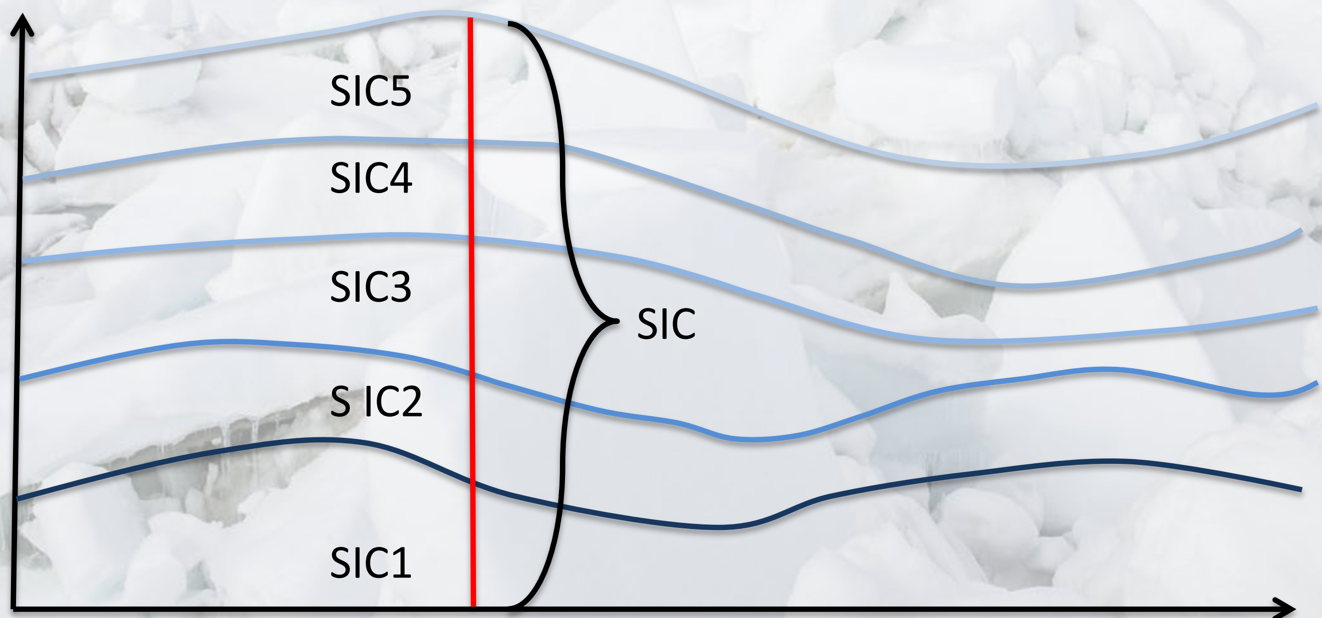


Get prior ensemble sample of observation y=h(x), by applying forward operator h to each ensemble member

http://www.image.ucar.edu/DAReS/DART/

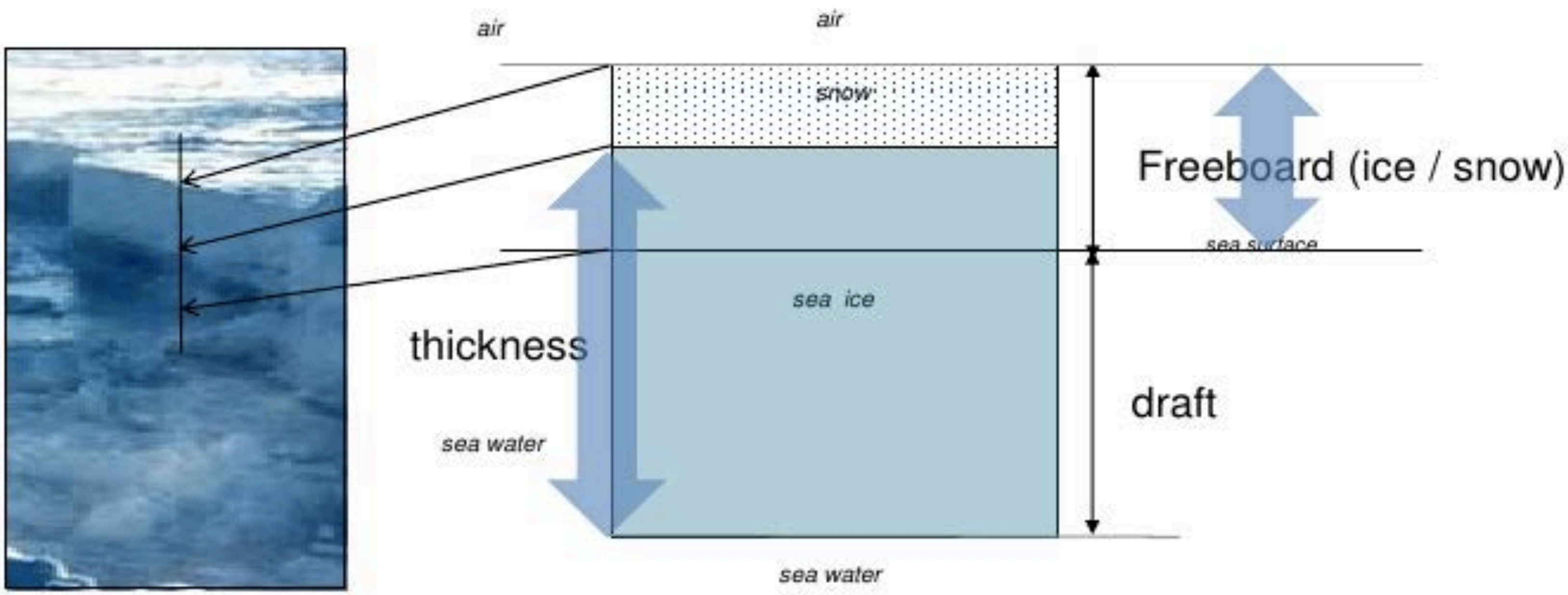
Observation operator

$$SIC = \sum_{i=1}^{N} SIC_i$$

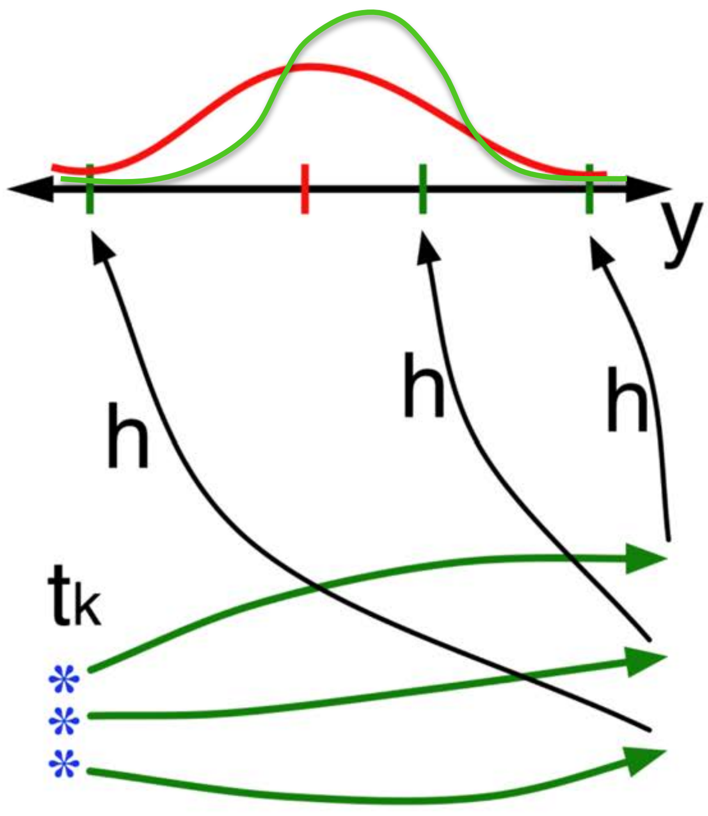


Observation operator

$$Fb = H_i \cdot \left(1 - \frac{\rho_i}{\rho_w}\right) - \frac{\rho_{sn}}{\rho_w} \cdot H_{sn}$$



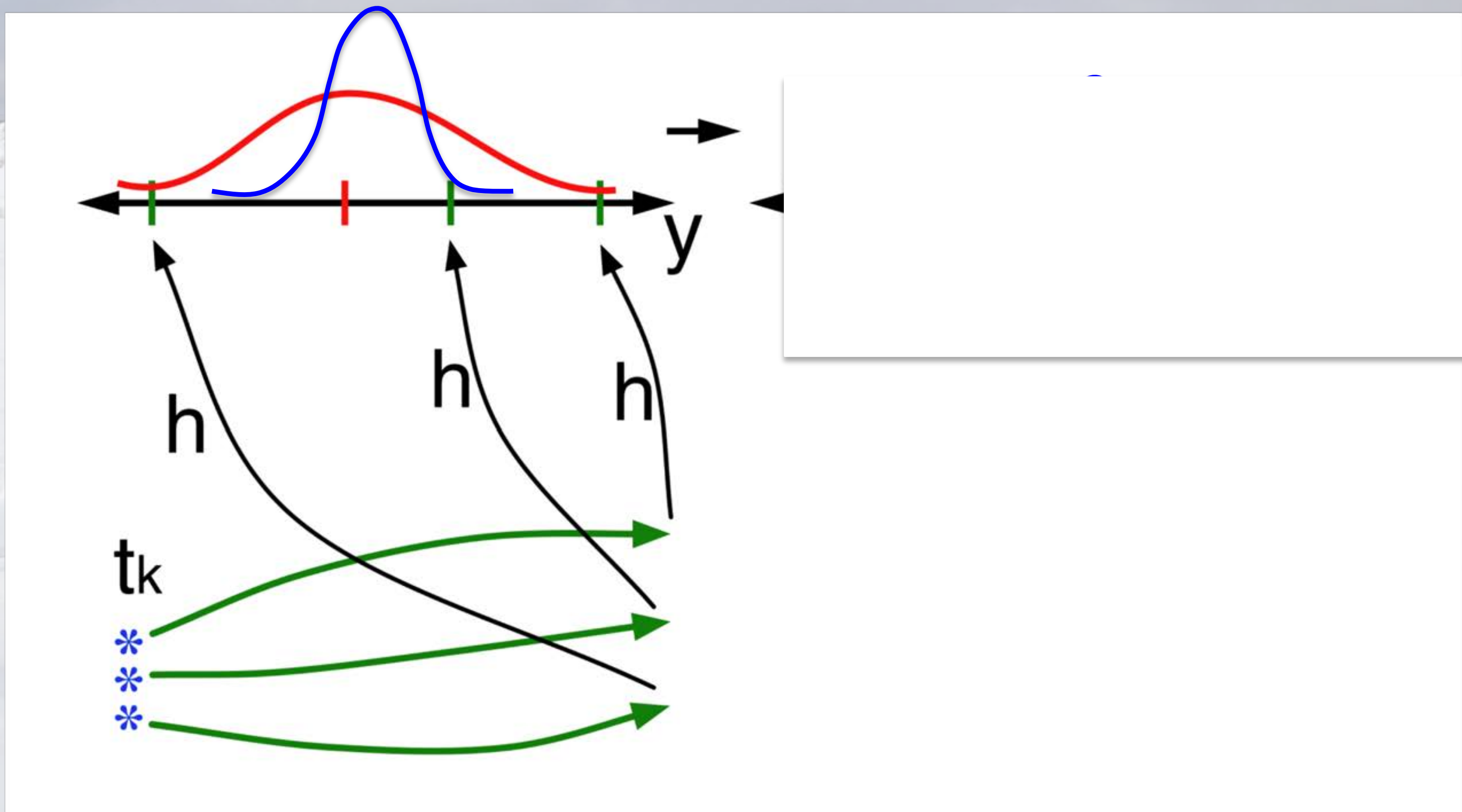
Flowchart of the Ensemble Kalman Filter



Get observed value and observational error distribution from the observation system

http://www.image.ucar.edu/DAReS/DART/

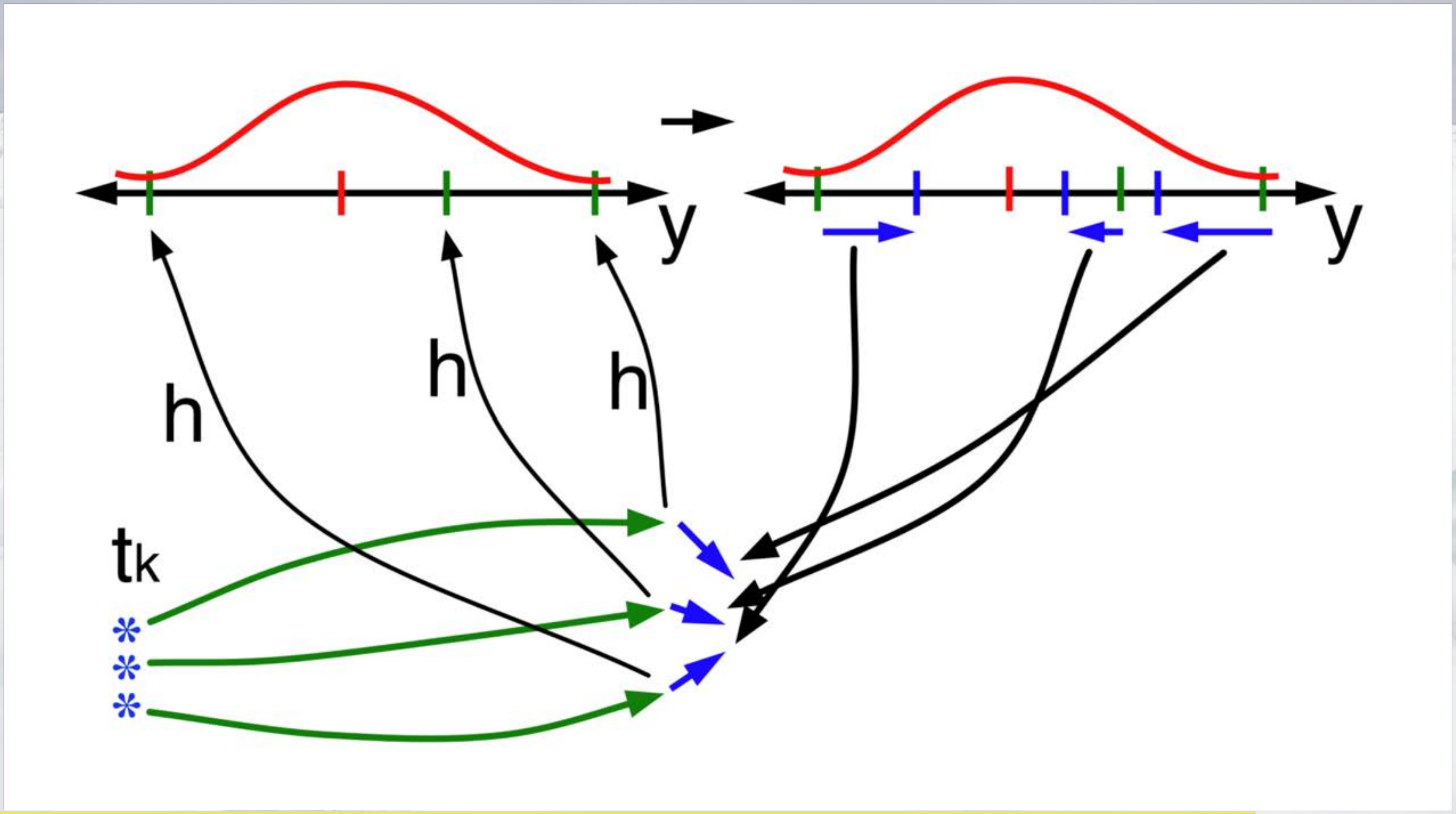
Flowchart of the Ensemble Kalman Filter



Get the posterior PDF, draw a sample from the posterior PDF, and find increment for each prior observation ensemble

http://www.image.ucar.edu/DAReS/DART/

Flowchart of the Ensemble Kalman Filter



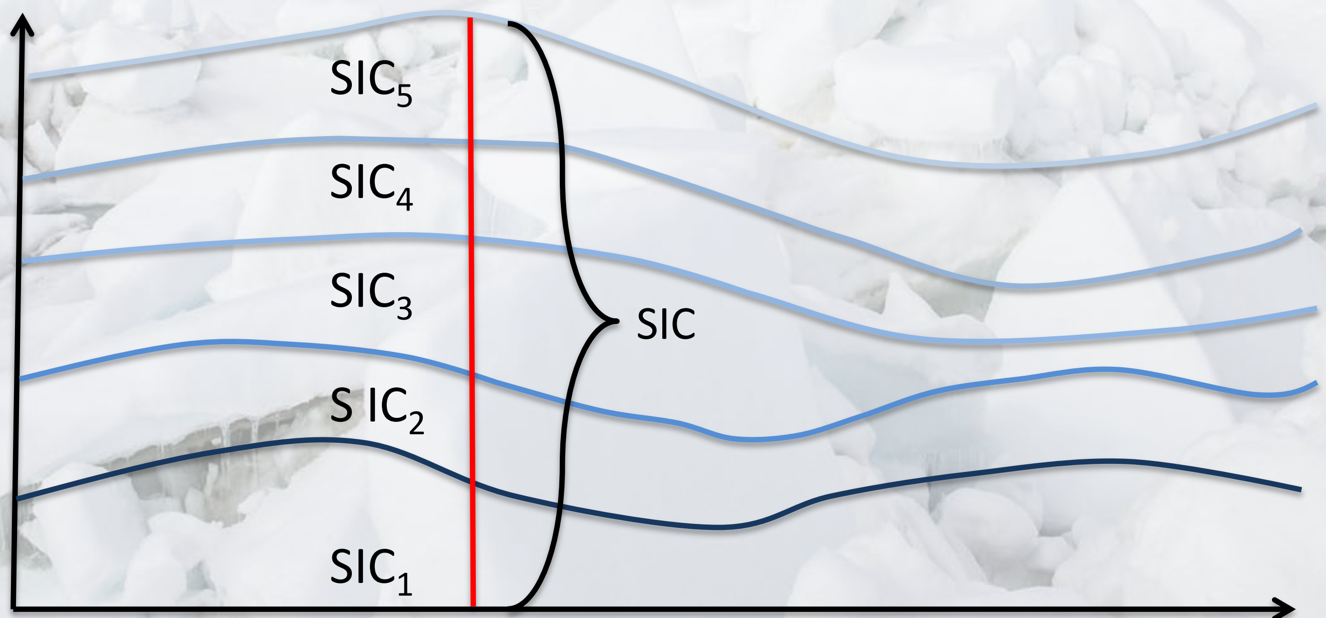
Use ensemble samples of y and each state variable to linearly regress observation increments onto state variable increments

http://www.image.ucar.edu/DAReS/DART/

Incrementing on unobserved variables

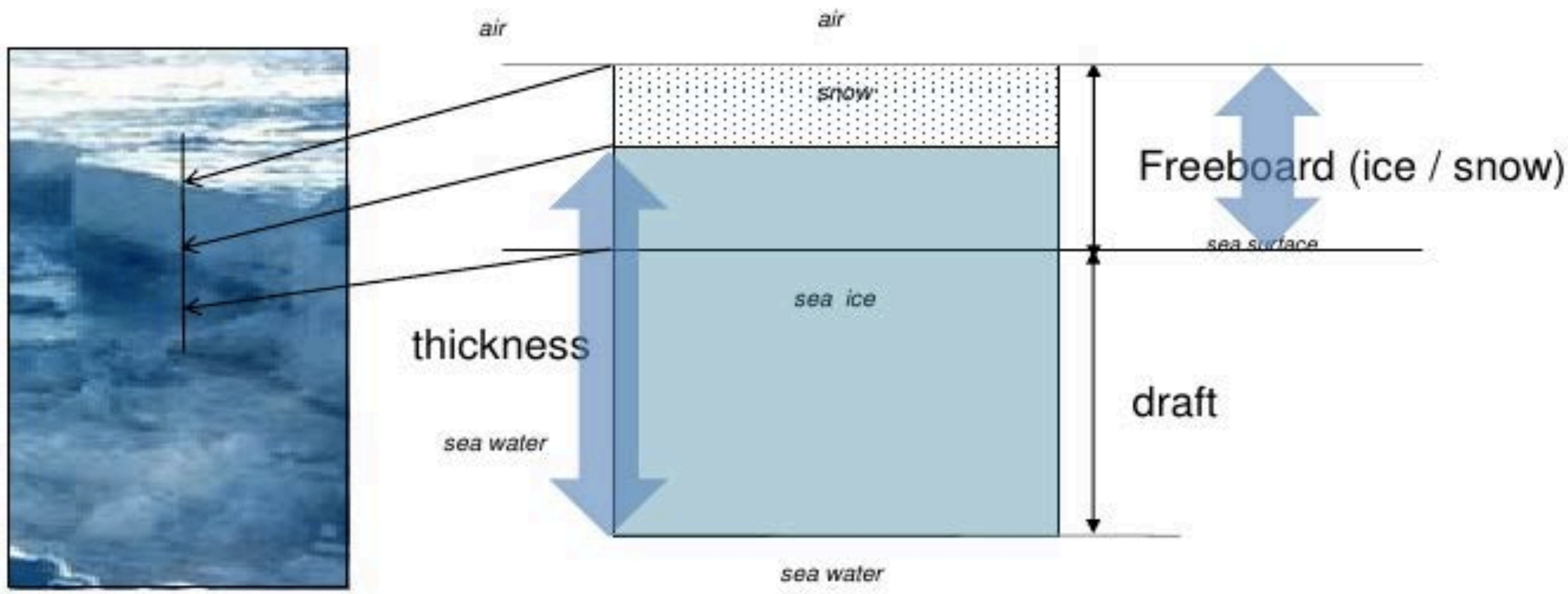
$$SIC = \sum_{i=1}^{N} SIC_i$$

SIC_i is an "unobserved" variable
SIC is the "observed" variable

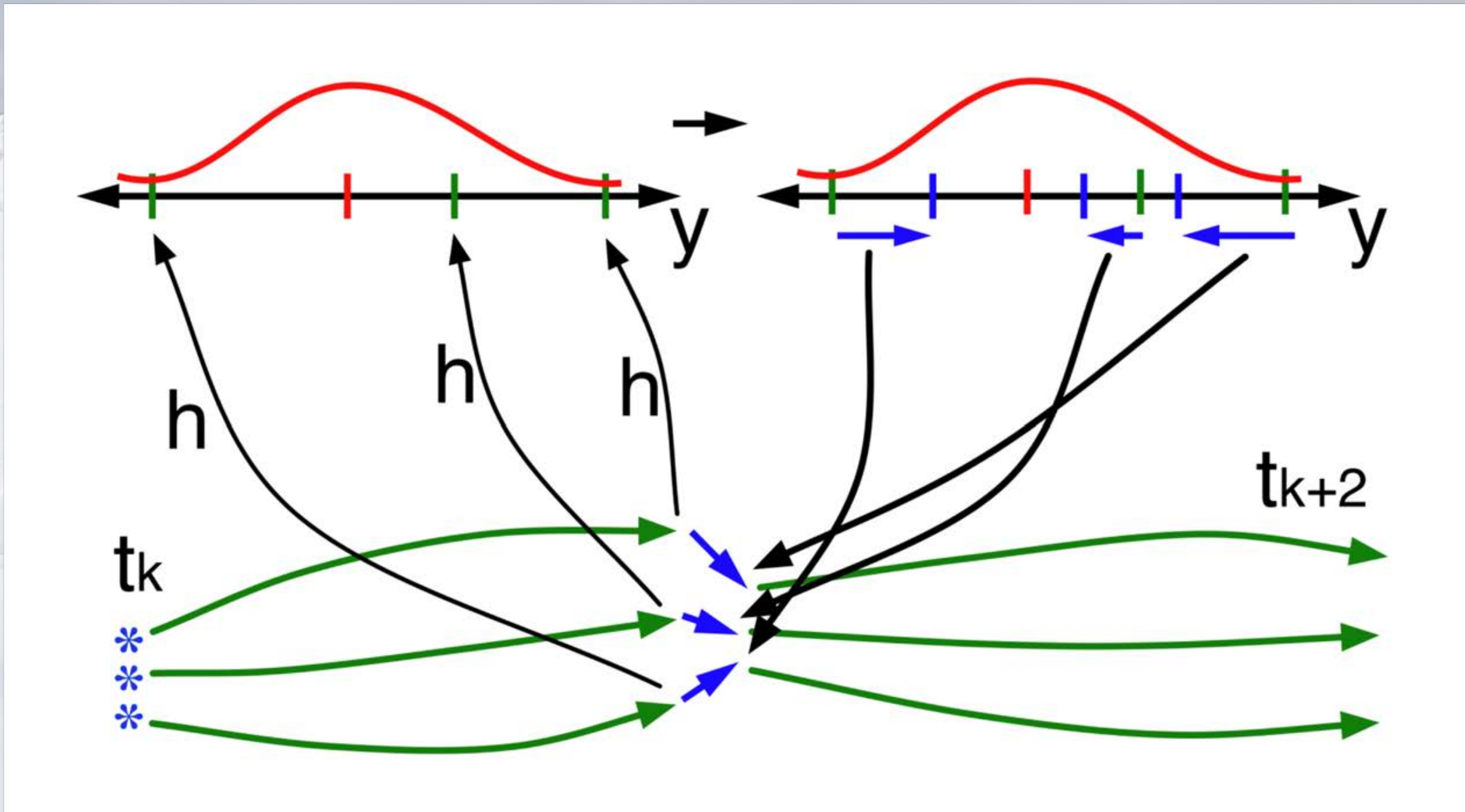


Fb is the observed variable
Hi and *Hsn* are unobserved variables

$$Fb = H_i \cdot \left(1 - \frac{\rho_i}{\rho_w}\right) - \frac{\rho_{sn}}{\rho_w} \cdot H_{sn}$$



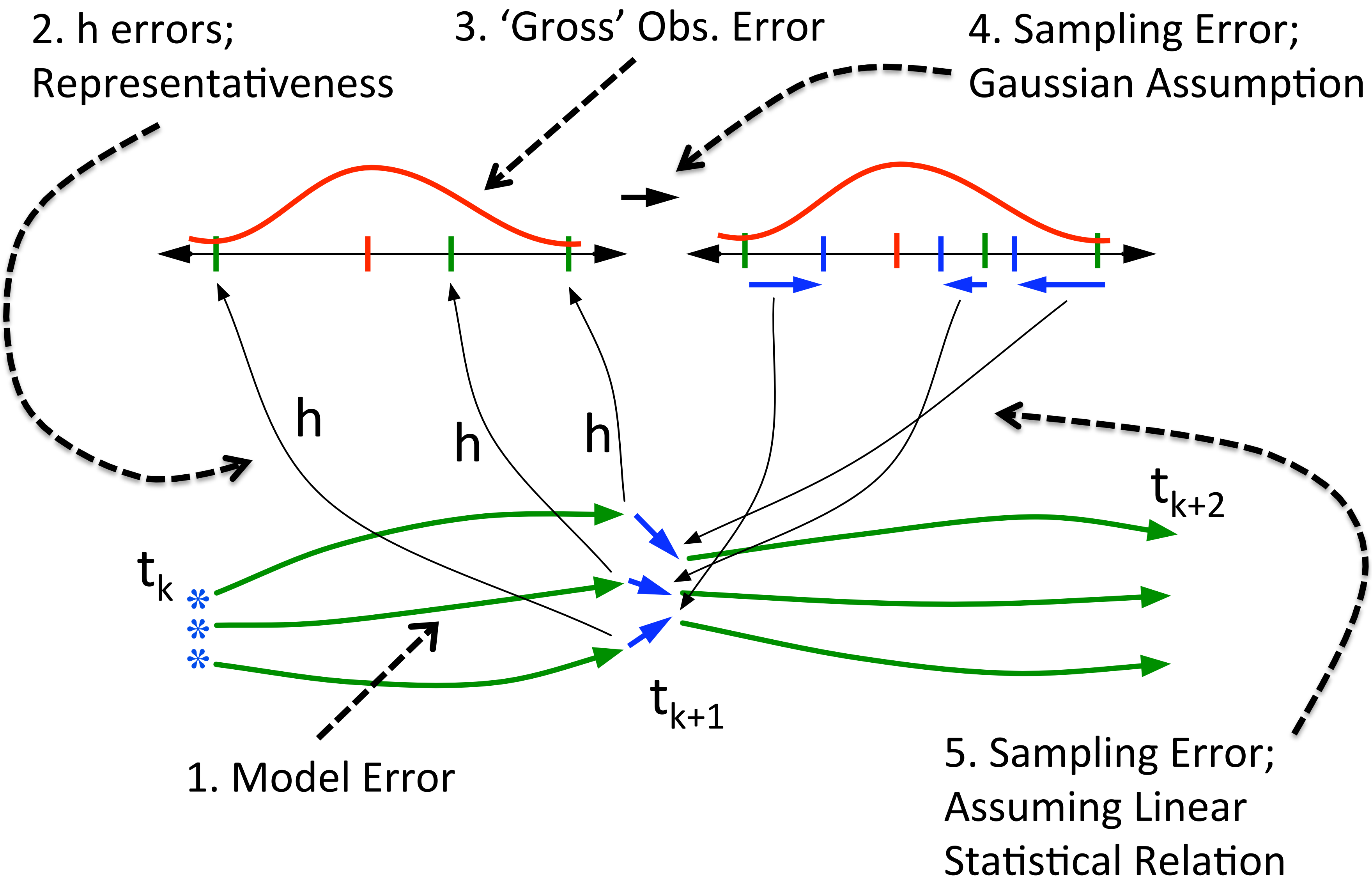
Flowchart of the Ensemble Kalman Filter



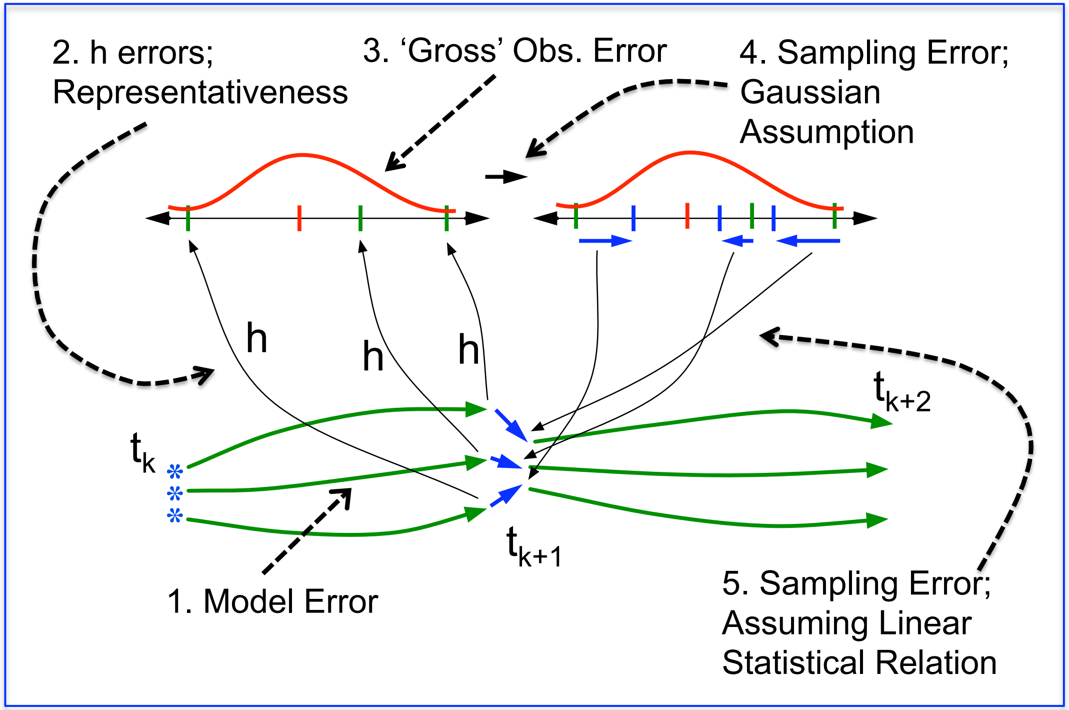
Model proceeds to the next time step with the new initial conditions

http://www.image.ucar.edu/DAReS/DART/

Some Error Sources in Ensemble Filters



Dealing with Ensemble Filter Errors



Fix 1, 2, 3 independently, HARD but ongoing.

Often, ensemble filters...

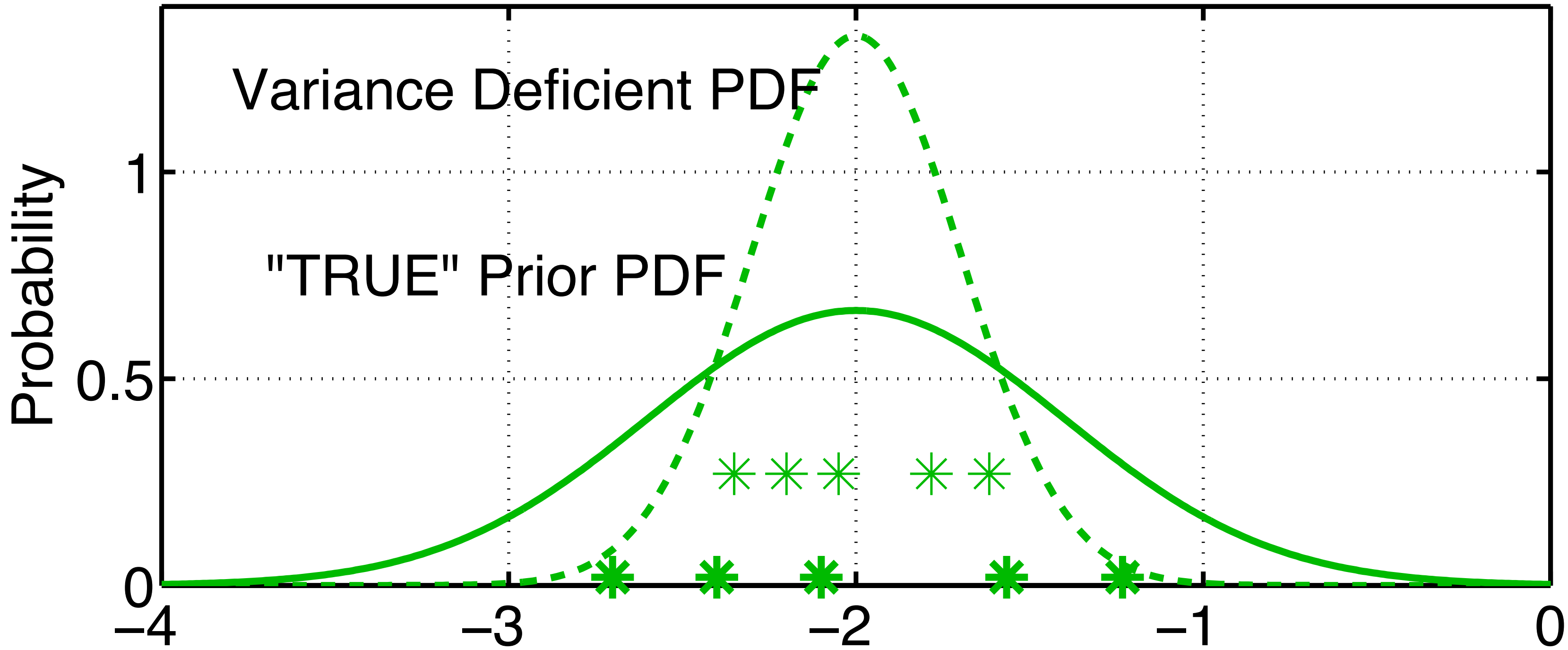
1-4: Variance inflation, Increase prior uncertainty to give obs more impact.

5. 'Localization': only let obs. impact a set of 'nearby' state variables.

Often smoothly decrease impact to 0 as function of distance.

Model/Filter Error: Filter Divergence and Variance Inflation

1. History of observations and physical system => ‘true’ distribution.
2. Sampling error, some model errors lead to insufficient prior variance.
3. Can lead to ‘filter divergence’: prior is too confident, obs. Ignored.

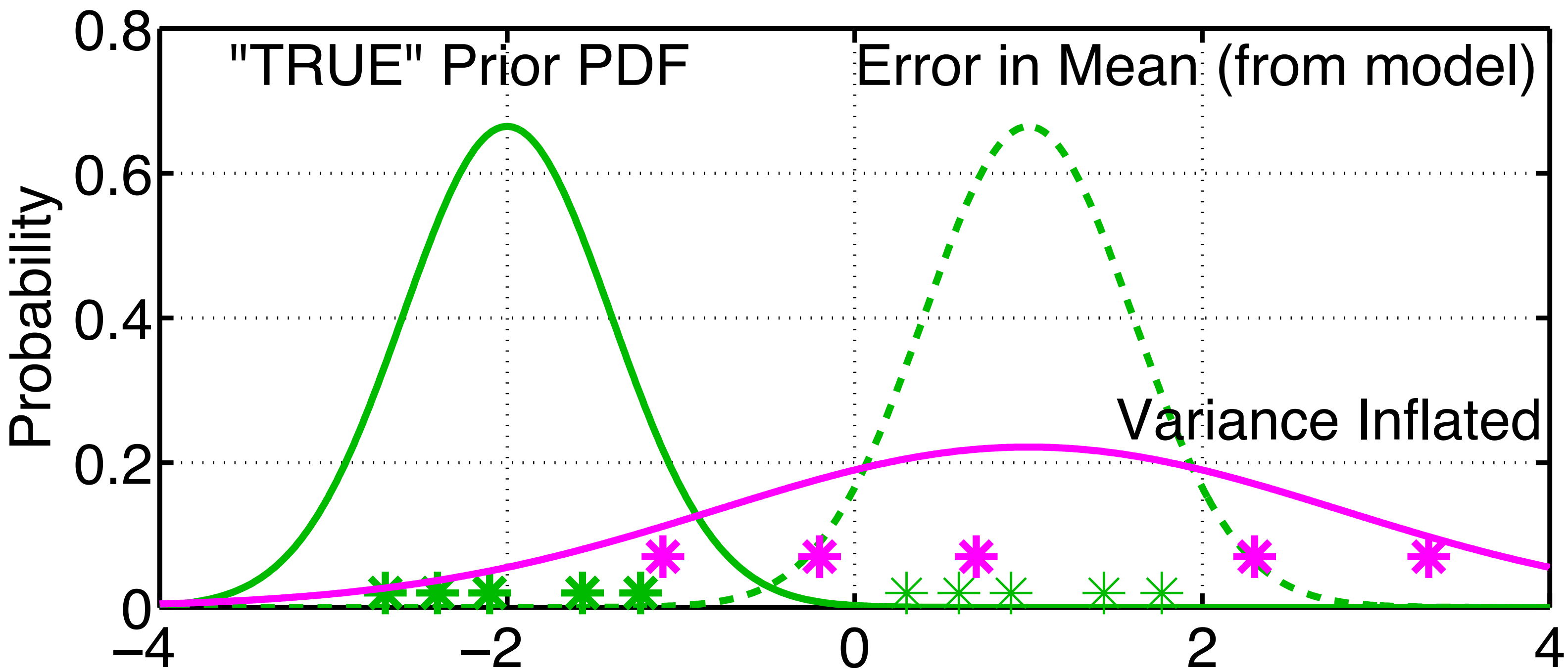


Naïve solution is variance inflation: just increase spread of prior.

For ensemble member i, $inflate(x_i) = \sqrt{\lambda}\,(x_i - \overline{x}) + \overline{x}$

Model/Filter Error: Filter Divergence and Variance Inflation

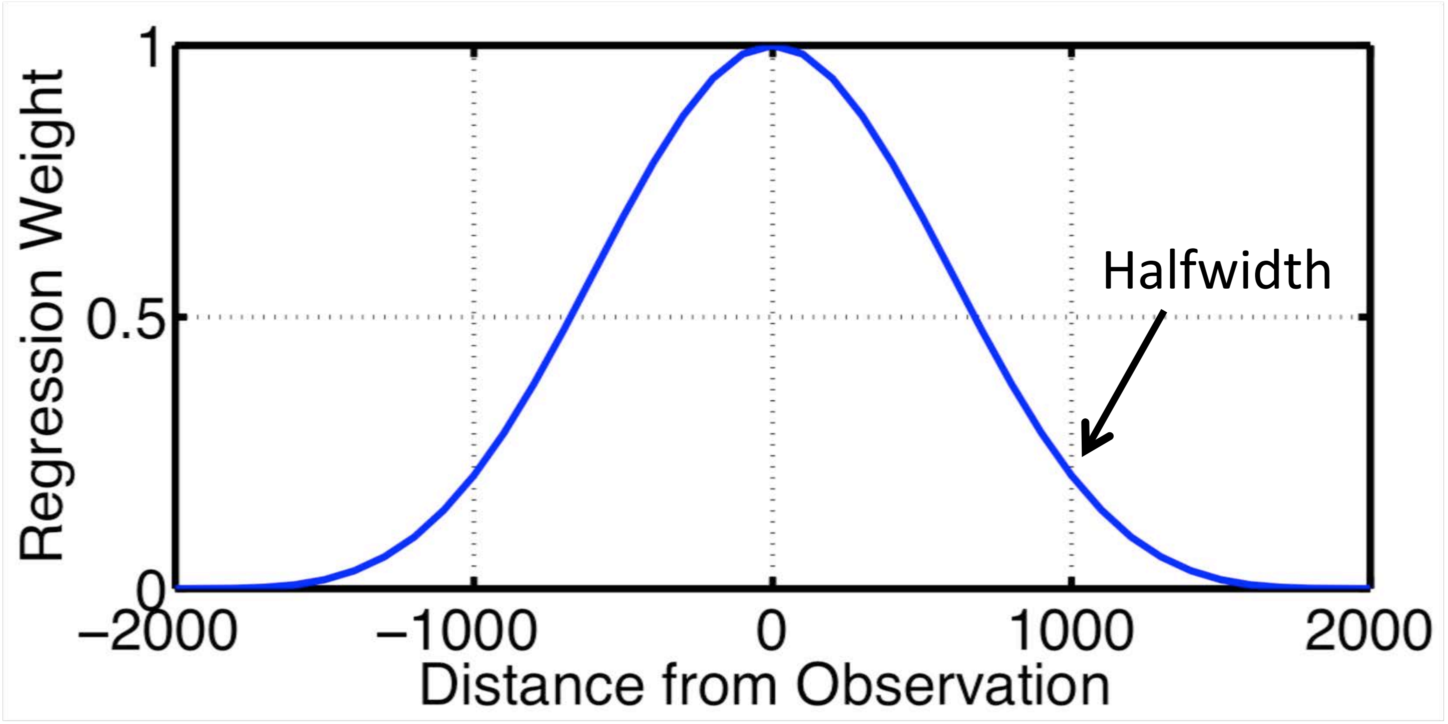
1. History of observations and physical system => ‘true’ distribution.
2. Most model errors also lead to erroneous shift in entire distribution.
3. Again, prior can be viewed as being TOO CERTAIN.



Inflating can ameliorate this.
Obviously, if we knew E(error), we’d correct for it directly.

Dealing with Regression Sampling Error

3. Use additional a priori information about relation between observations and state variables.



Can use other functions to weight regression.

Unclear what distance means for some obs./state variable pairs.

Referred to as LOCALIZATION.

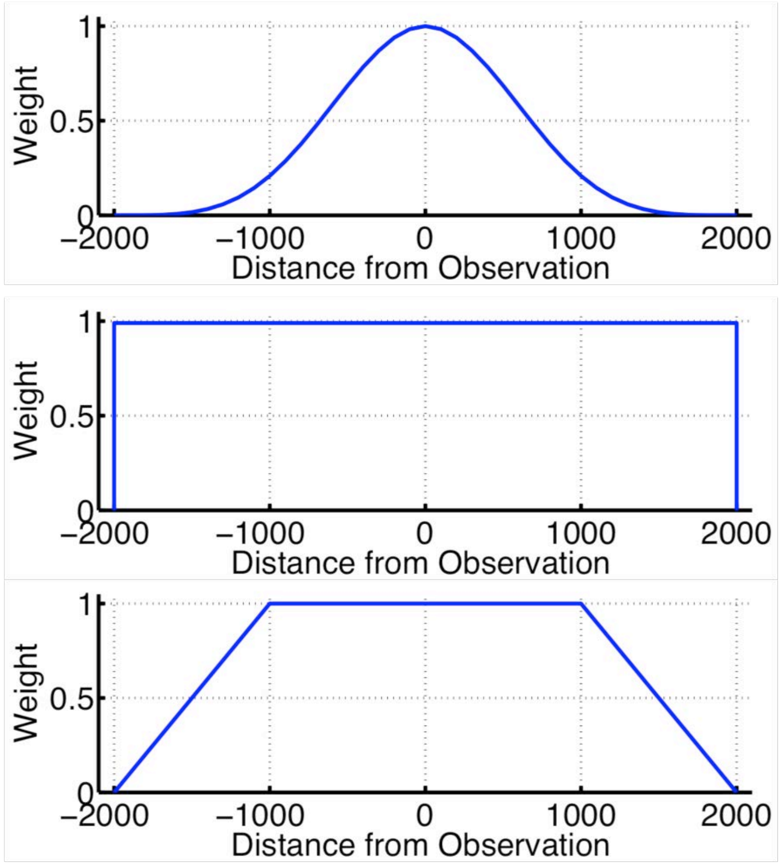
DART provides several localization options

1. Different shapes for the localization function are available.
   Controlled by *select_localization* in `&cov_cutoff_nml`.

1=> Gaspari-Cohn

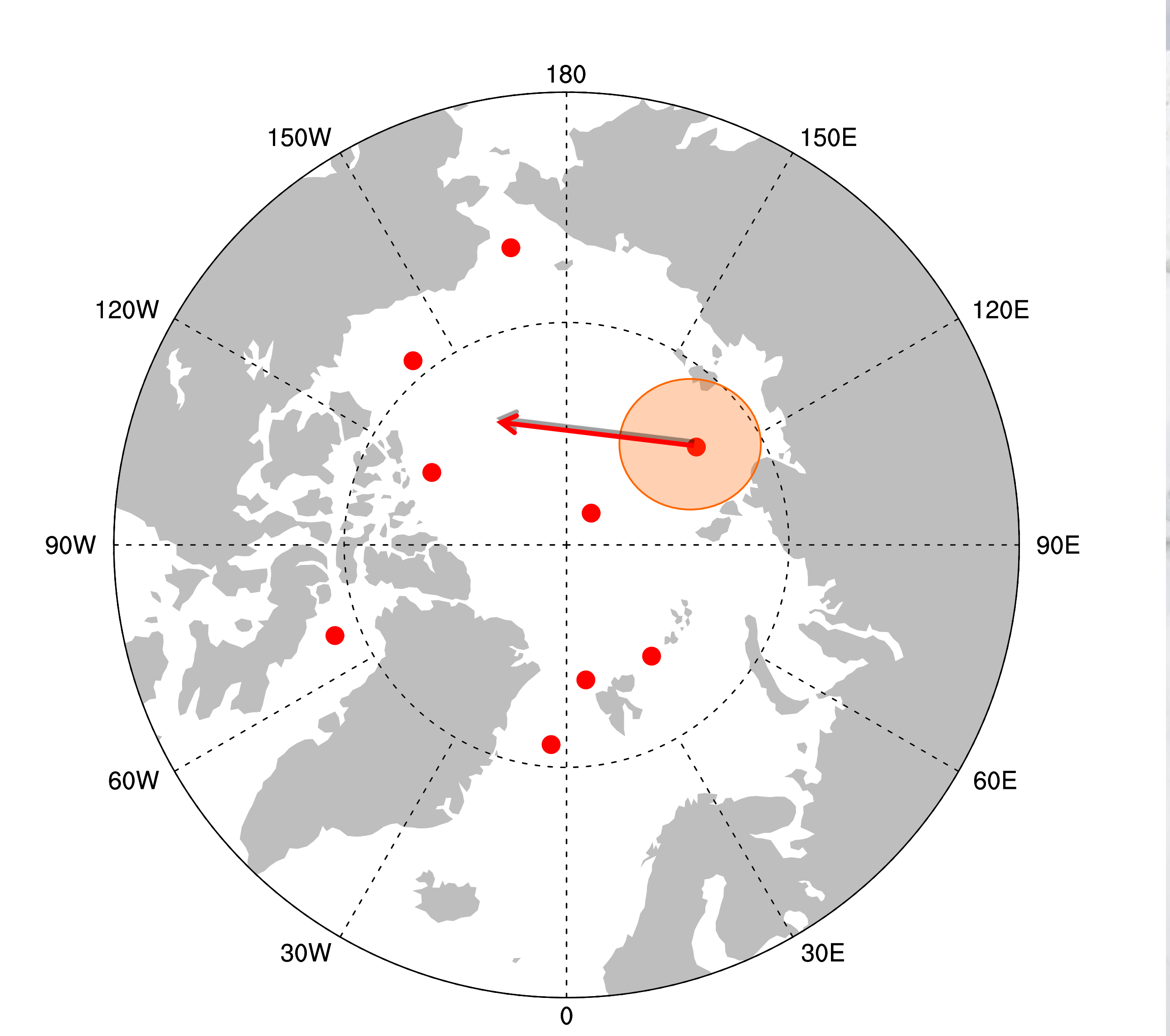
2=> Boxcar

3=> Ramped Boxcar

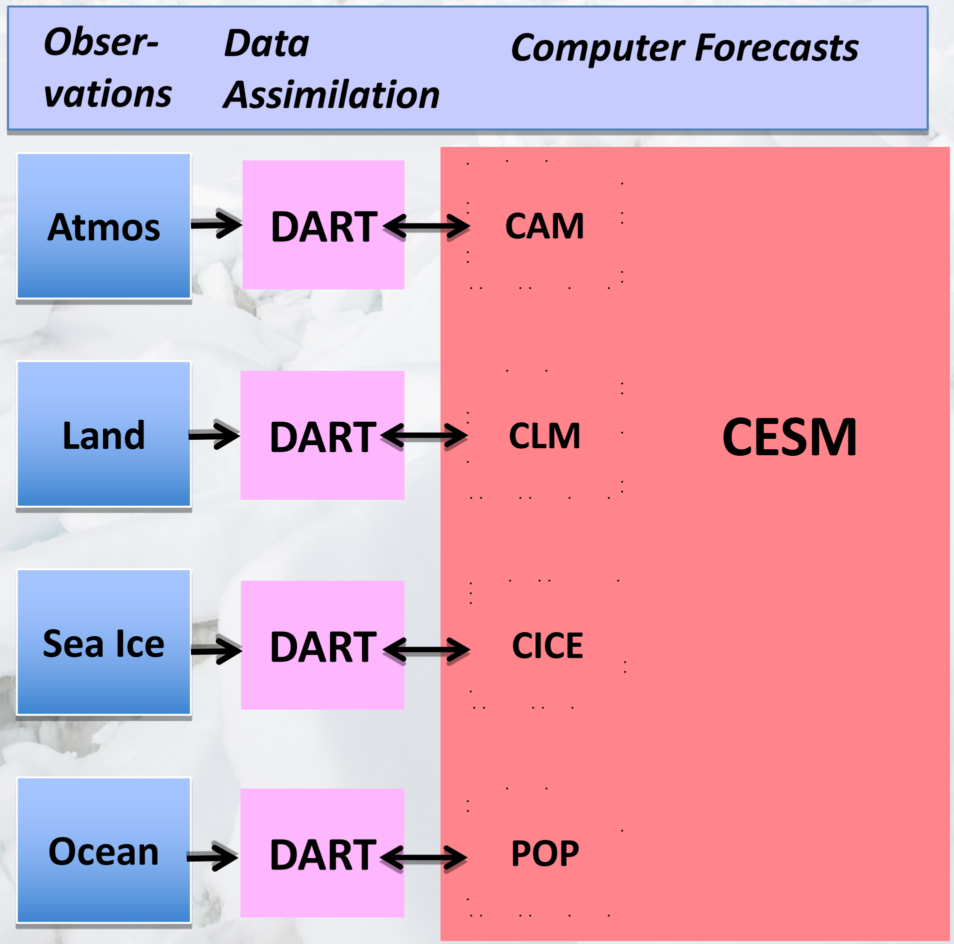


2. Halfwidth of localization function set by *cutoff* in `&assim_tools_nml`

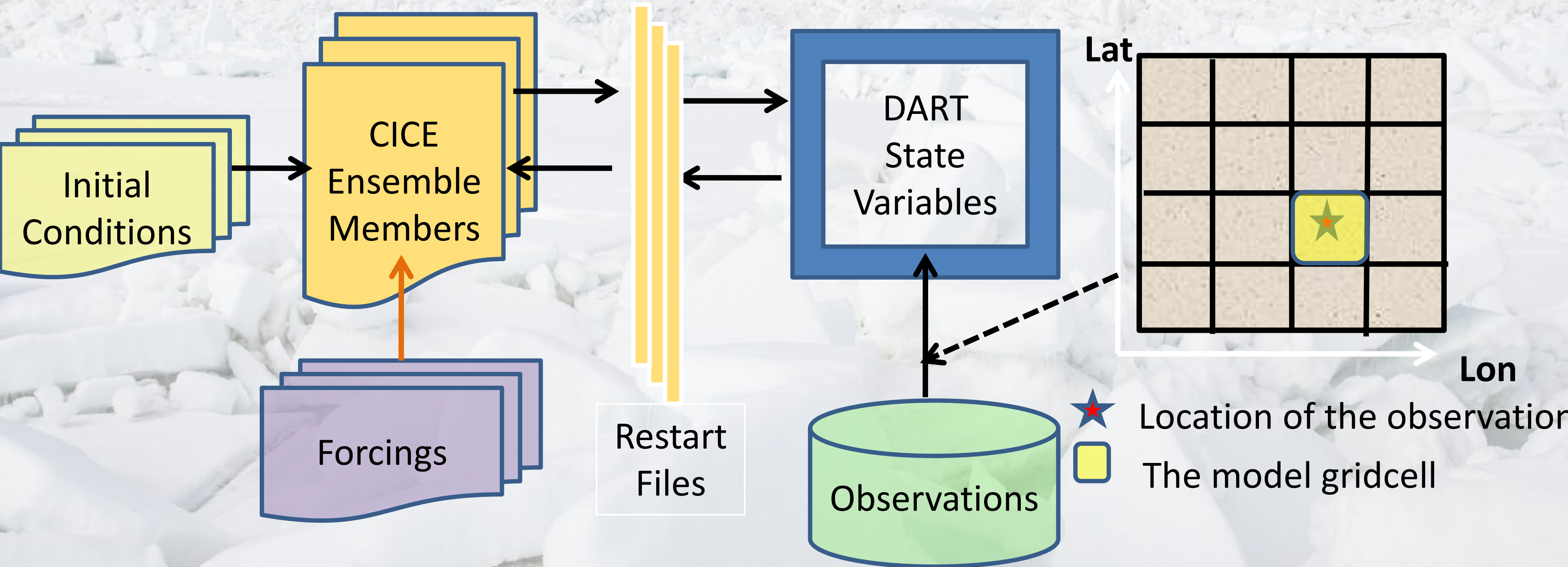
Localization



The Data Assimilation System Of CESM/DART



The Coupled DART and CICE



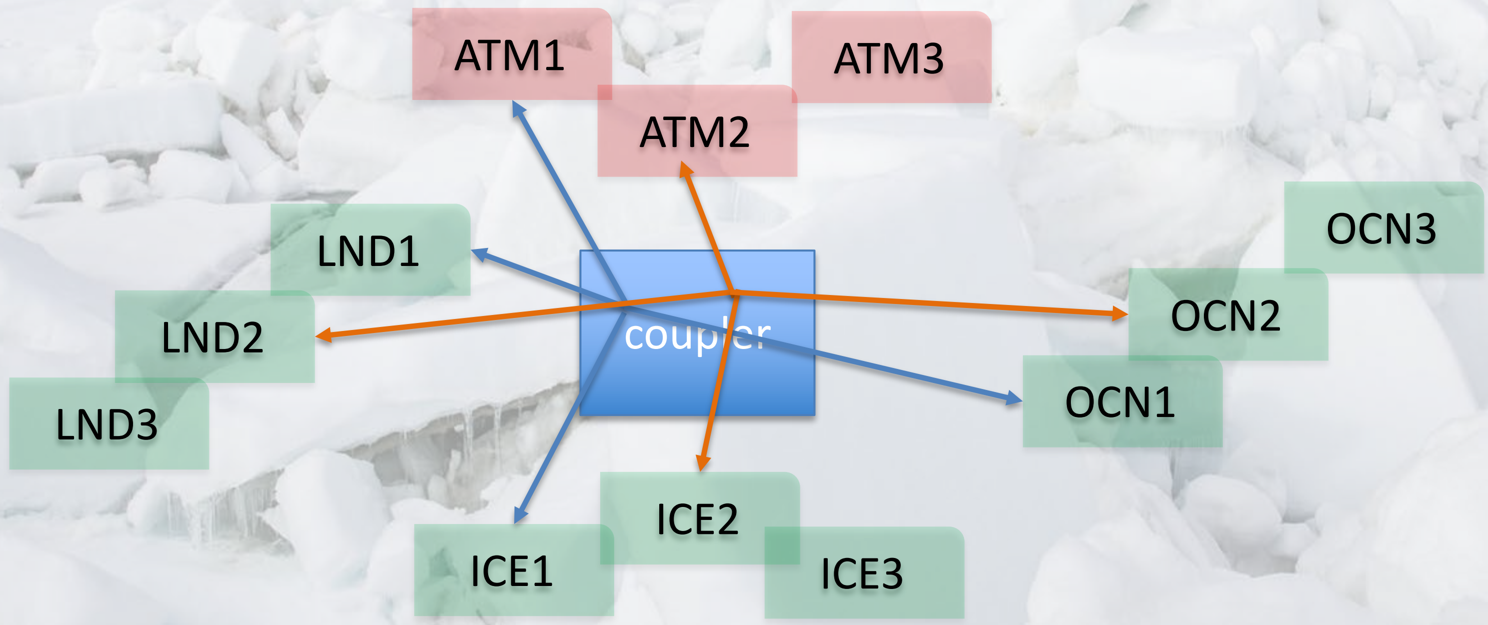
Executables: filter, dart_to_cice

Things to do

- Log on Cheyenne
- Copy the folder of day4 into your account, read the guidelines
- Download the DART code
- Check out the cesm2_0_ensemble_setup jobscript
  - Ensemble atmospheric forcing
  - Ensemble sets of parameters
  - Ensemble initial conditions

- Check out the observation sequence files

- Compset: DTEST
  - 2000_DATM%NYF_SLND_CICE_DOCN%SOM_DROF%NYF_SGLC_SWAV_TEST
  - Data atmosphere, slab ocean, active sea ice
- The ensemble-capability of CESM
  - One executable, multiple instances



Let's run 10 ensemble members

Specify the namelist files

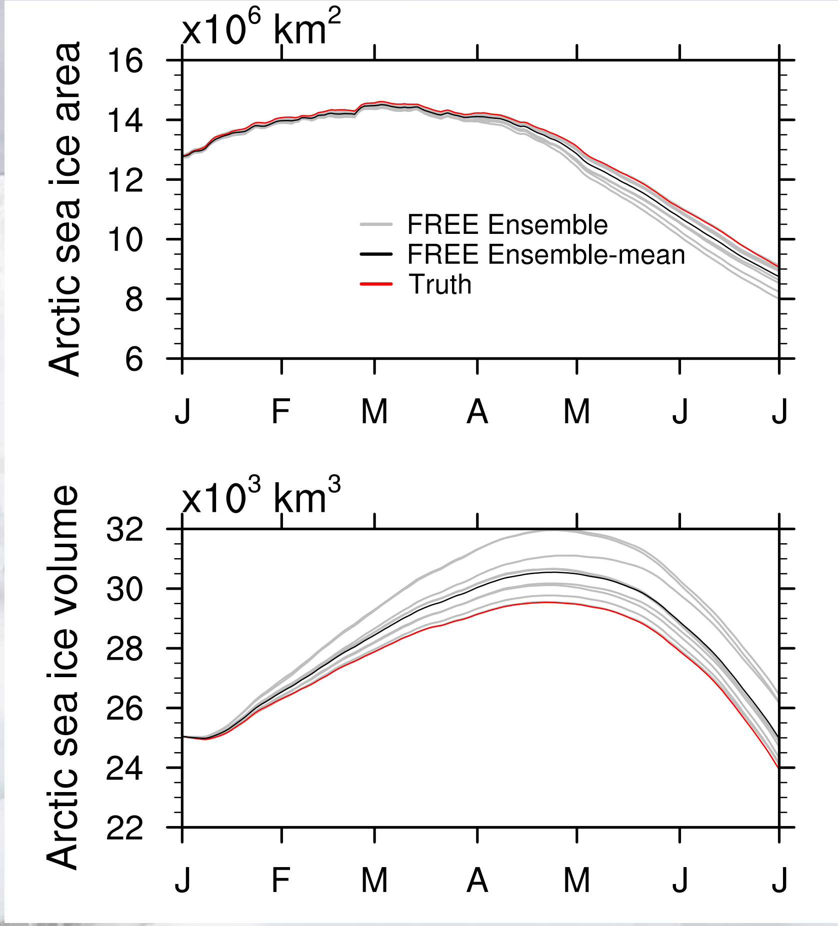
- user_nl_datm_0001

```
streams   = 'datm.streams.txt.CPLHIST3HrWx.Solar_0001             2001 2001 2003',
            'datm.streams.txt.CPLHIST3HrWx.Precip_0001            2001 2001 2003',
            'datm.streams.txt.CPLHIST3HrWx.nonSolarNonPrecip_0001 2001 2001 2003'
            'datm.streams.txt.presaero.clim_2000_0001 1 1 1'
vectors   = 'u:v'
mapmask   = 'nomask',
            'nomask',
            'nomask',
            'nomask'
tintalgo = 'coszen',
            'nearest',
            'linear',
            'linear'
~
```

- user_nl_cice_0001

```
ice_ic    = '//scratch/01548/yfzhang/inputdata_cam/ice/cice/2001-01-01/cice5
_sp2000_ens30.cice_0001.r.2001-01-01-00000.nc'
histfreq_n =  1,1,1,1,1
histfreq   = 'd','m','x','x','x'
f_sst = 'dmxxx'
f_sss = 'dmxxx'
f_frzmlt = 'dmxxx'
f_frz_onset = 'dmxxx'
f_aicen = 'dmxxx'
Cf    =  45.2172
r_snw   =  1.61738
```

The free run with 10 ensemble members



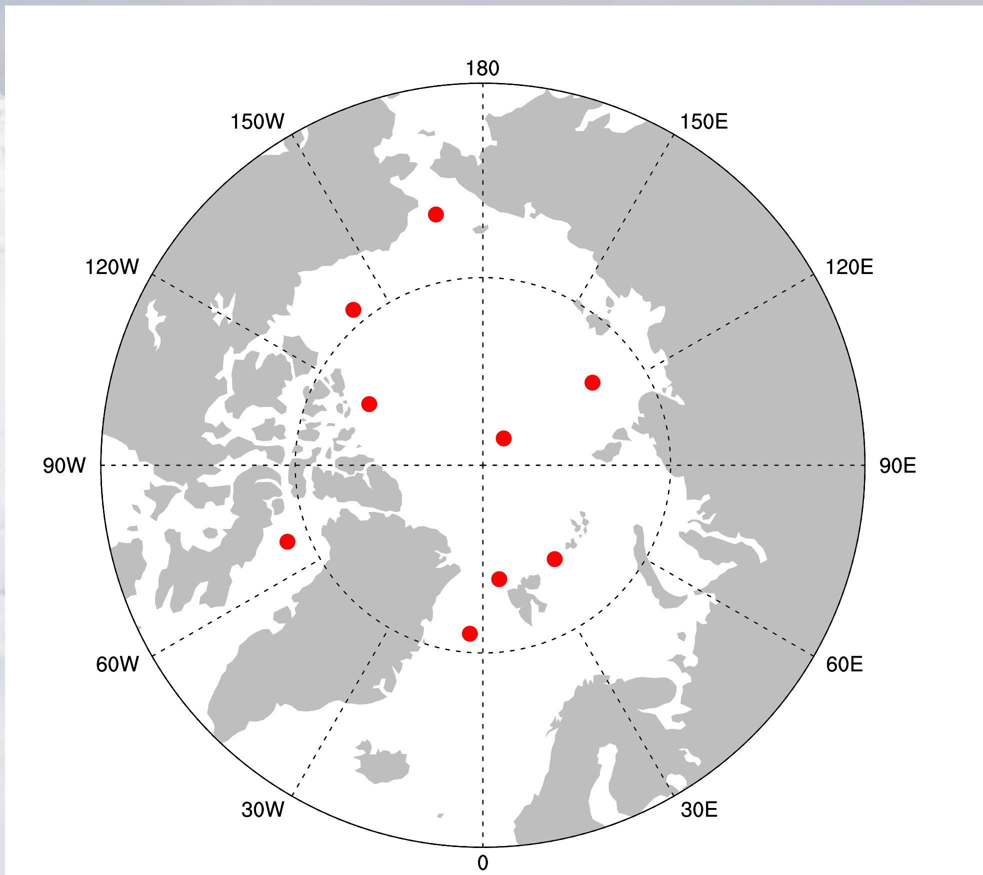
- We chose one ensemble member as the "truth"
- and add some noise to make "synthetic observations": 0.4m random error is added to sea ice thickness (SIT)

observation sequence files

- DART has its own format for observations
  - obs_seq.YYYY-MM-DD-SSSSS

```
obs_sequence
obs_kind_definitions
           1
           5 SAT_SEAICE_AGREG_THICKNESS
  num_copies:            1  num_qc:            1
  num_obs:            1  max_num_obs:            1
observation
Data QC
  first:            1  last:            1
 OBS            1
   1.37225902080536
  0.000000000000000E+000
          -1           -1           -1
obdef
loc3d
    0.652776735853137       1.40541362362681       0.000000000000000     -1
kind
           5
     0     147653
  0.160000000000000
```

Choose your favorite observation spot



Beaufort
Chukchi
FramStrait
GreenlandSea
Kara
NofArchip
NPole
BaffinBay
Barents

~yfzhang/PWS2018/day4/obs_seqs/$observation_spot/