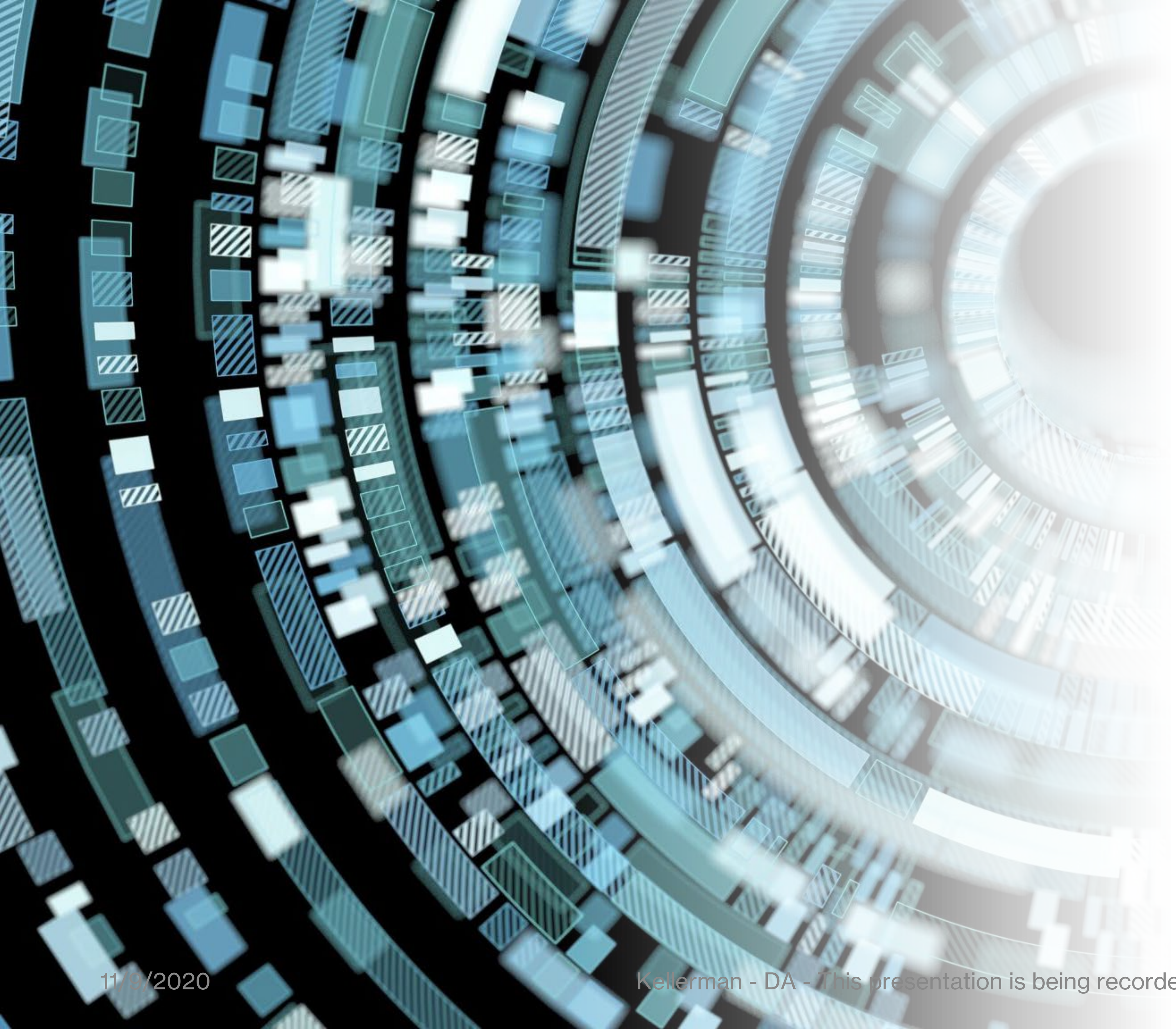


Data Assimilation

Adam C. Kellerman

Department of Earth, Planetary, and Space Sciences, UCLA

Outline

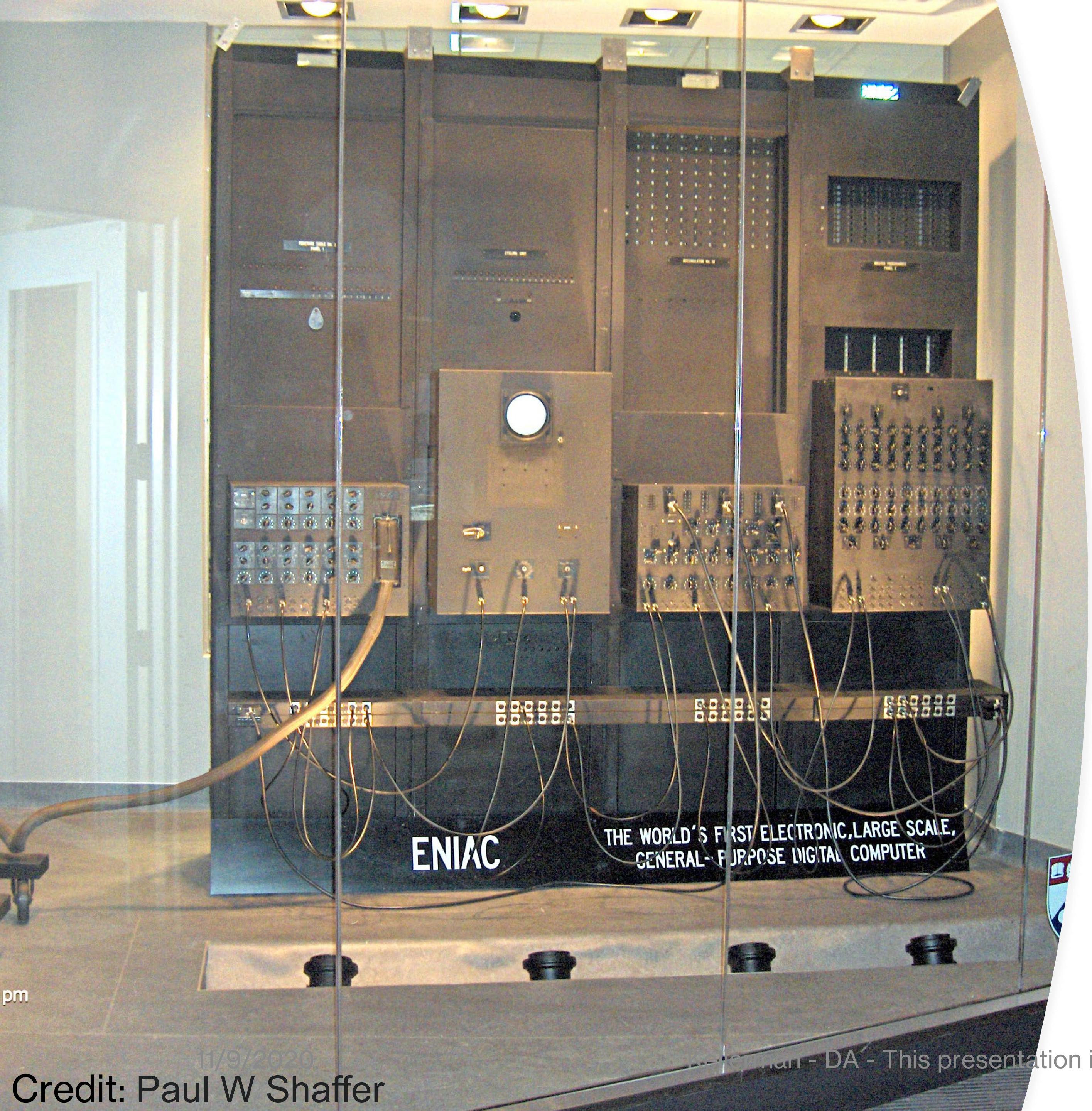
- Brief history of NWP and data assimilation
- Empirical methods of data assimilation with examples
- Data assimilation in the radiation belts – key advances
- Now and into the future

Credit: Alf Lannerbaeck

Artist's impression by Alf Lannerbaeck, published by the Sw

History of Data Assimilation

- Begins with numerical weather prediction (NWP) - the most common techniques were developed through NWP
- Vilhelm Bjerknes (Norwegian) - weather forecasting is a deterministic initial value problem (1904)
  - "A sufficiently accurate knowledge of the state of the atmosphere at the initial time"
  - "A sufficiently accurate knowledge of the laws according to which one state of the atmosphere develops from another"
- Bjerknes - surface observation stations and the Bergen school of synoptic and dynamic meteorology
- Three of Bjerknes' students were fundamental to NWP; Rossby, Eliassen, and Fjortoft
- Lewis Fry Richardson (1922) – first to propose a practical solution – conducted a 6-hour forecast by hand
- Richardson had a vision that 64,000 *computers* would be needed, within a massive factory to race the weather for the whole globe



Credit: Paul W Shaffer

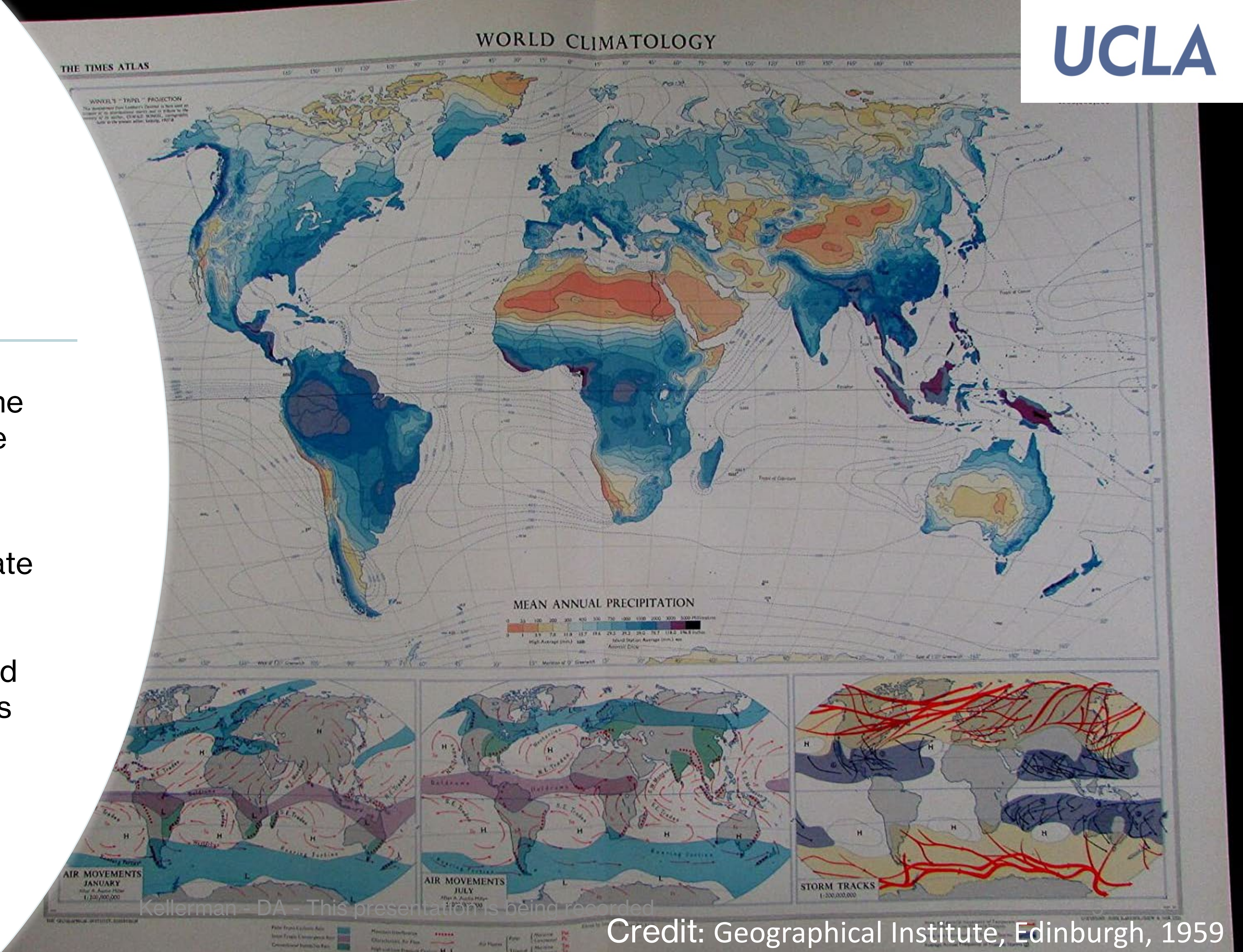
History of Data Assimilation

- Predicting the evolution of the atmosphere requires finding numerical solutions to partial differential equations (PDEs)
- In 1928 Courant, Freidreichs, and Lewy (Germans) studied approximate solutions to PDE's through finite differences
- In 1945 the Electronic Numerical Integrator and Computer (ENIAC) was developed
- ENIAC was donated to Pennsylvania University, upgraded and then moved to Arberdeen, Maryland
- The first NWP was made in 1950 by Charney and Neumann, using the ENIAC
- Specification of the initial condition or state of the atmosphere quickly developed into its own theoretical and practical field

UCLA

History of Data Assimilation

- With added complexity, the degrees of freedom in the models far exceeded the number of observations
- Introduce background state and *a priori* information (Bergthorsson and Doos, 1955), e.g. climatology, and later, short-term forecasts



Credit: Geographical Institute, Edinburgh, 1959

What is Data Assimilation?

- The initial condition or state is referred to as the analysis
- The act of combining the data and model to obtain the initial state is known as objective analysis, while the act of repeating a forecast-analysis cycle over time is referred to as data assimilation
- The aim (in NWP) is to obtain the best forecast, and hence find the analysis which delivers on that aim - not necessary the "best" analysis
- The use of data assimilation to (re)construct a historic dataset is known as reanalysis

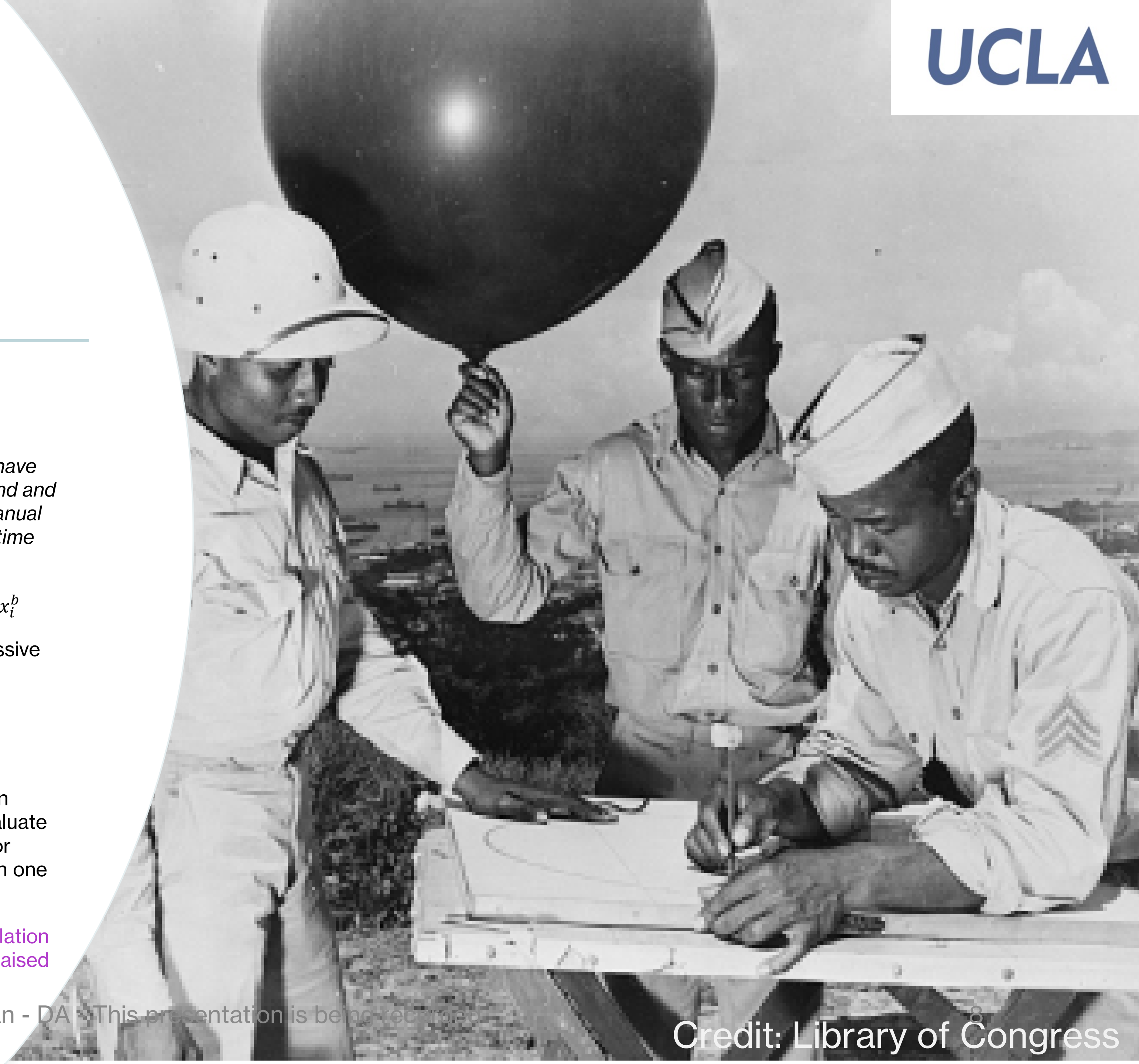
Data Assimilation – Empirical Methods

- Direct Insertion – assume observations are reality
- Nudging - also known as Newtonian relaxation (*Hoke and Anthes* [1976], *Kistler* [1974]), the method nudges the solution towards the observations using a relaxation time τ, e.g. $\frac{\partial x}{\partial t} = f + \frac{x_{obs} - x}{\tau}$
- A smaller τ results in faster convergence of the solution towards the observations
- *Hoke and Anthes* [1976], suggested that τ be chosen such that this term is “similar in magnitude to less dominant terms” in the PDE

Data Assimilation – Empirical Methods

Successive Corrections Method (SCM)

- Developed (first) by Bergthorsson and Doos (1955)
- *"The first attempts at numerical weather forecasting on a routine basis have been characterized by a combination of tedious manual work on one hand and electronic computations with extremely high speed on the other…the manual part of these operations consumes time that is out of proportion to the time required for the machine computation."*
- The first estimate of the state at point i, x_i^0 is given by the background x_i^b
- After the first estimate, the following iterations are obtained by "successive corrections"
- $x_i^{n+1} = x_i^n + \frac{\sum_{k=1}^{K_i^n} w_{ik}^n (x_k^o - x_k^n)}{\sum_{k=1}^{K_i^n} w_{ik}^n + \varepsilon^2}$
- x_i^n is the nth iteration estimation at grid point i, x_k^o is the kth observation surrounding the grid point i, x_k^n is the value of the nth field estimate evaluate at the observation point k, ε^2 is ratio of observation to background error variance. Lastly, the weights may be determined by any method, though one that uses squared difference between grid points is common.
- The power of this technique to allow the computer to make the interpolation to grid points, based on weights and regression, addressing the issue raised above.



Credit: Library of Congress

Least Squares Estimation

We are interested in determining the relationship between two or more variables, by minimizing the error or residuals between variables

Least squares – central axiom in most DA schemes based on statistical estimation theory – Attributed to Gauss (~1794) though first published by Legendre (1805) then Gauss (1809)

Least Squares – "Toddler Example"

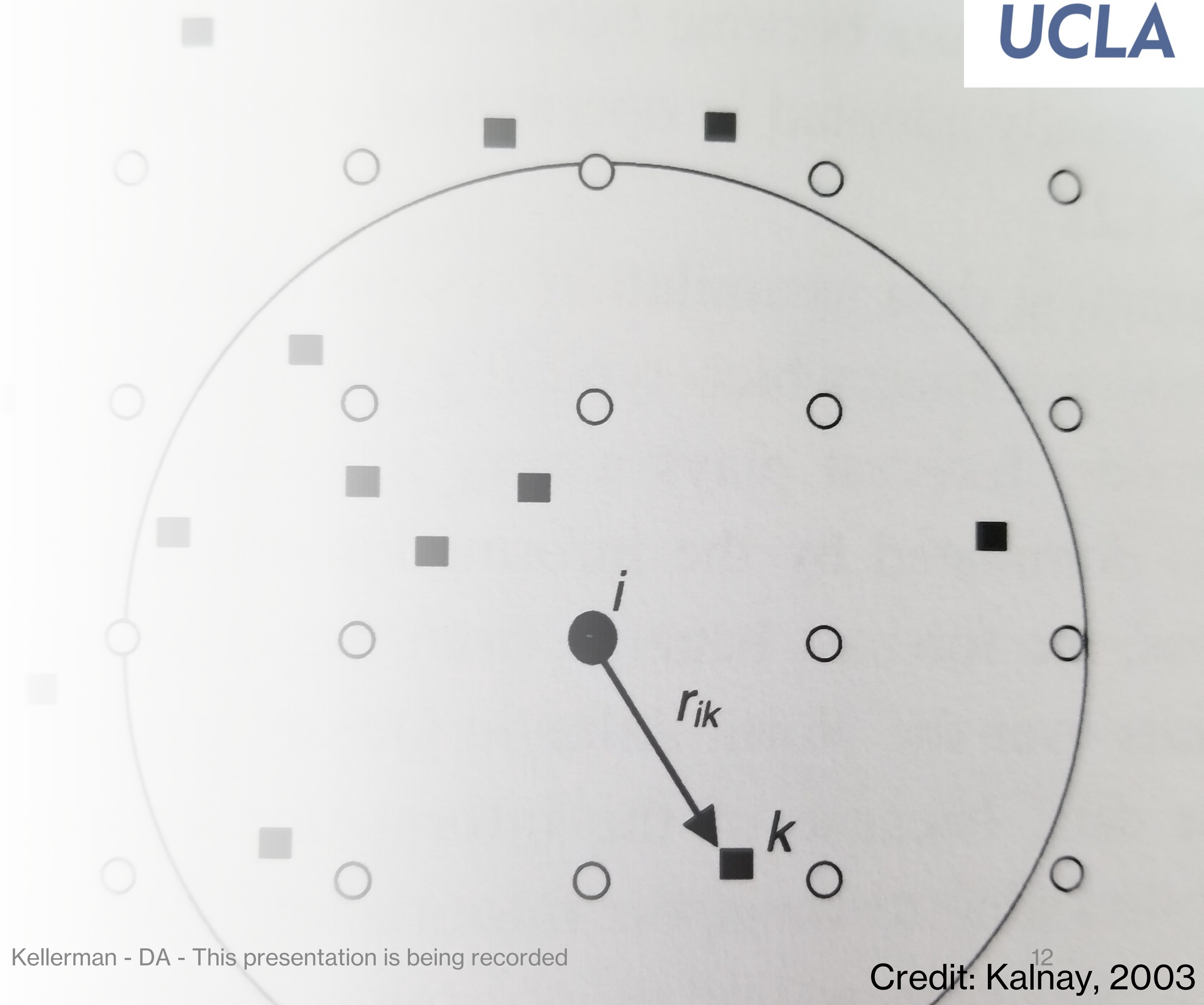
- A toddler wants to know the weight of a jellybean jar x, in a variation of the original guessing game, and her parents decide to help by asking two other older children, what their guesses were:
- $x_1 = x_T + \varepsilon_1, x_2 = x_T + \varepsilon_2$
- The parents assume that the two guesses are unbiased, as each of the older children are experts
- $E(\varepsilon_1^2) = \sigma_1^2, E(\varepsilon_2^2) = \sigma_2^2$
- The parents try to estimate the truth (analysis) by using a linear combination of the two estimates
- $x_a = a_1 x_1 + a_2 x_2, \quad a_1 + a_2 = 1$
- They recognize that x_a will be the best estimate of x_T if they choose coefficients to minimize the mean squared error of x_a
- $\sigma_a^2 = E[(x_a - x_T)^2] = E[\left(a_1(x_1 - x_T) + a_2(x_2 - x_T)\right)^2]$

Data Assimilation – "Toddler Example"

- They then derive a solution for a_1 and a_2,
- $a_1 = \frac{\frac{1}{\sigma_1^2}}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}} = \frac{\sigma_2^2}{\sigma_1^2+\sigma_2^2}$
- $a_2 = \frac{\frac{1}{\sigma_2^2}}{\frac{1}{\sigma_1^2}+\frac{1}{\sigma_2^2}} = \frac{\sigma_1^2}{\sigma_1^2+\sigma_2^2}$
- They understand that the value of the coefficients are proportional to the precision of the guesses (inverse of the guess error variances), and so
- $\frac{1}{\sigma_a^2} = \frac{1}{\sigma_1^2} + \frac{1}{\sigma_2^2}$
- Which states that, if the variances are known, the sum of the precision of each guess is equal to the precision of the analysis and x_a will be the best linear unbiased estimate (BLUE) of the true weight of the jellybean jar x_T.
- After some discussion with the two children regarding their prior estimates, the parents arrive at values for σ_1 and σ_2, which they use to compute a and b to obtain a guess for the toddler

Data Assimilation – Some more history

- First example of an automated, global, objective analysis algorithm (*Panofsky* [1949]) was based on a two-dimensional polynomial interpolation
- A local polynomial interpolation scheme based on least squares was first used (*Gilchrist and Cressman* [1954]) to minimize the difference between observations and model within a radius of influence – localization



Credit: Kalnay, 2003

Cost Function

- Our enterprising parents get motivated, and want to determine if they can arrive at a solution through a maximum likelihood approach. In this case, the analysis is the most likely value of x, given the two observations
- $\Pr\{x_1|x_T\} = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x_1-x_T)^2}{2\sigma_1^2}}$
- $\Pr\{x_2|x_T\} = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(x_2-x_T)^2}{2\sigma_2^2}}$
- The parents recognize that the most likely value of x_T, given x_1 and x_2, is the one that minimizes their product, the join probability
- $\Pr\{x_1|x_T\} \Pr\{x_2|x_T\} = \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{(x_1-x_T)^2}{2\sigma_1^2}-\frac{(x_2-x_T)^2}{2\sigma_2^2}}$

Cost Function

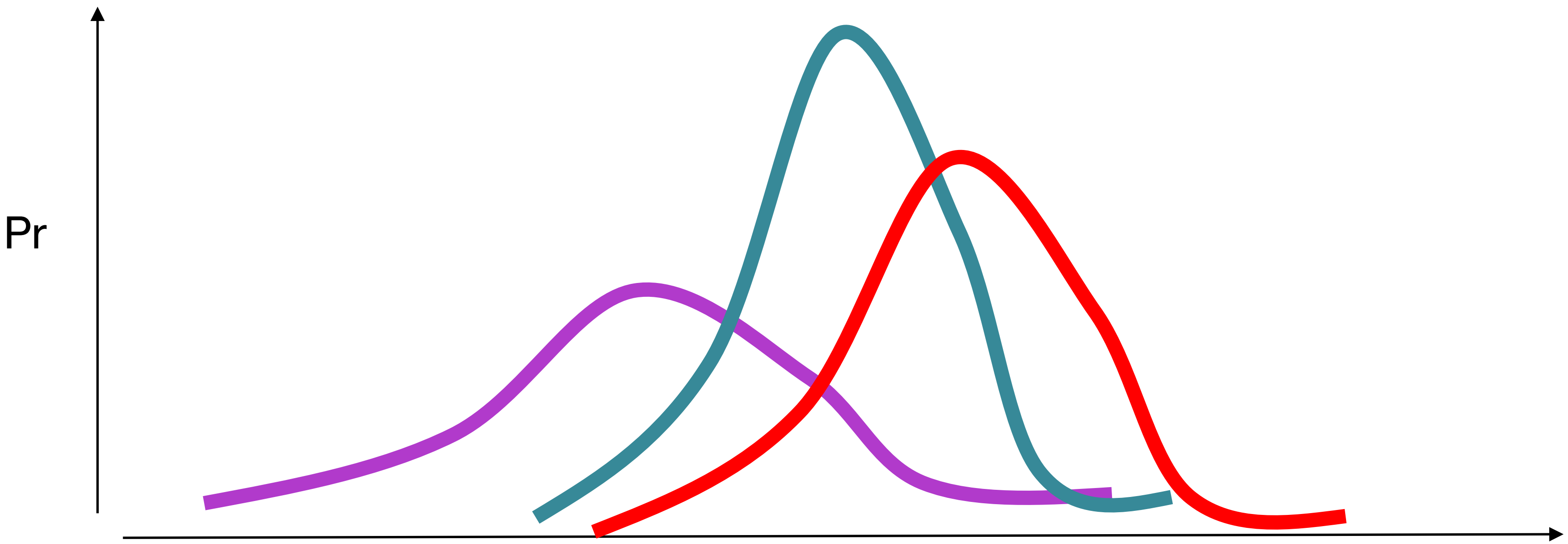
- The maximum likelihood can be determined by taking the logarithm of the result
- $\ln(\Pr\{x_1|x_T\}\Pr\{x_2|x_T\}) = \max[const. - \frac{(x_1-x_T)^2}{2\sigma_1^2} - \frac{(x_2-x_T)^2}{2\sigma_2^2}$
- Which provides the cost function f, which they need to minimize, representing the sum of the square of the difference of the estimate to the two guesses, weighted by the precision of each guess.
- $f(x) = \frac{1}{2}\left[\frac{(x_T-x_1)^2}{\sigma_1^2} + \frac{(x_T-x_2)^2}{\sigma_2^2}\right]$

Bayesian Derivation

- One parent then asks the other, what if we can only ask one child at a time? We'd have to account for the prior guess, before estimating x_T.
- The parents understand that Baye's rule is derived from the Laws of probability
- $\Pr\{A,B\} = \Pr\{A|B\}\Pr\{B\} = \Pr\{B|A\}\Pr\{A\}$
- Assuming that the first observation x_1 is already made
- $\Pr\{x_T|x_2\} = \frac{\Pr\{x_2|x_T\}\ \Pr\{x_T\}}{\Pr\{x_2\}} = \frac{\frac{1}{\sqrt{2\pi}\sigma_2}e^{-\frac{(x_2-x_T)^2}{2\sigma_2^2}}\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x_1-x_T)^2}{2\sigma_1^2}}}{\Pr\{x_2\}}$
- where x_T is based on all prior information, including x_1
- $\Pr\{x_T|x_2\}$ – update (posterior) estimate of the guess based on all prior information, and the new guess x_2
- The least-squares and cost function approaches are equivalent, which holds in higher dimensional problems

Bayesian Derivation

$$\Pr\{A|BC\} = \frac{\Pr\{B|AC\}\ \Pr\{A|C\}}{\Pr\{B|C\}}$$



Data Assimilation – Sequential Assimilation

How do we utilize this knowledge to conduct data assimilation?

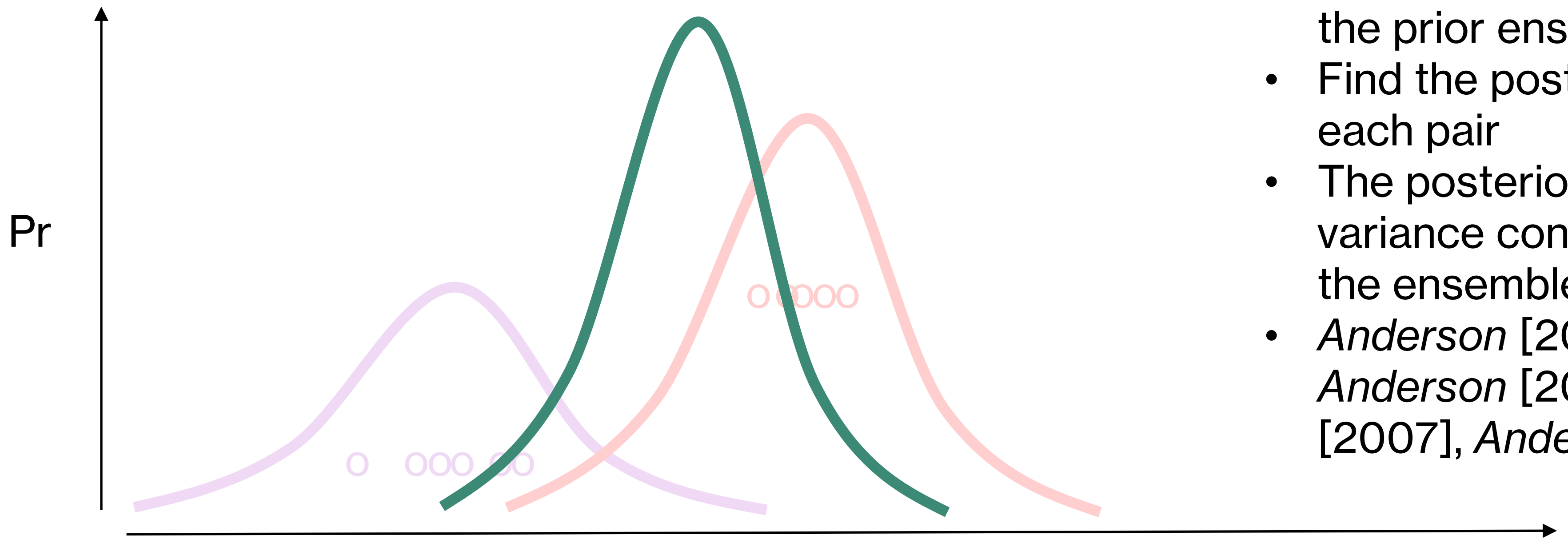
- $x_a = x_b + W(x_o - x_b)$
- x_b is background, first guess, model forecast, or prior analysis
- x_o is a new observation
- $x_o - x_b$ is the increment or innovation
- The weight $W = \sigma_b^2(\sigma_b^2 + \sigma_o^2)^{-1}$, and $\sigma_a^2 = (\sigma_b^{-2} + \sigma_o^{-2})^{-1}$

Data Assimilation – Sequential Assimilation

- This scheme is used in optimal interpolation (OI), where we would assume some inflation α of the initial error covariance
- $\sigma^2_{b,t+1} = \alpha\sigma^2_{a,t}$, with $\alpha \geq 1$
- If we consider the case where x_b is given by a model forecast, through solution of a PDE, we can compute the forecast error covariance $\sigma^2_{b,t+1}$ from the model itself.
- This leads to the Kalman filter (*Kalman,* 1960)
- The power of the technique is to quickly assimilate information and approach the truth, without having to conduct statistics on multiple observations at a given time
- $x_a = x_b + W(x_o - x_b)$
- See *Ghil and Malanotte-Rizzoli* [1991] and *Gil et al.* [1997] for an excellent review of data assimilation methods

Data Assimilation - Ensembles

$$\Pr\{A|BC\} = \frac{\Pr\{B|AC\}\ \Pr\{A|C\}}{\Pr\{B|C\}}$$

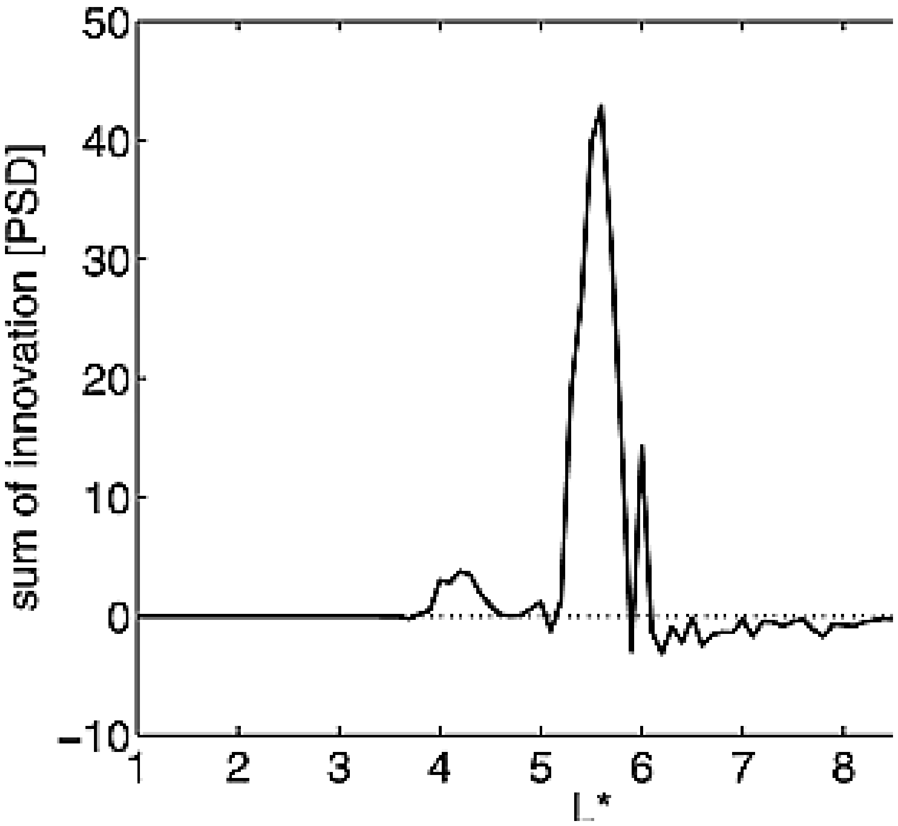


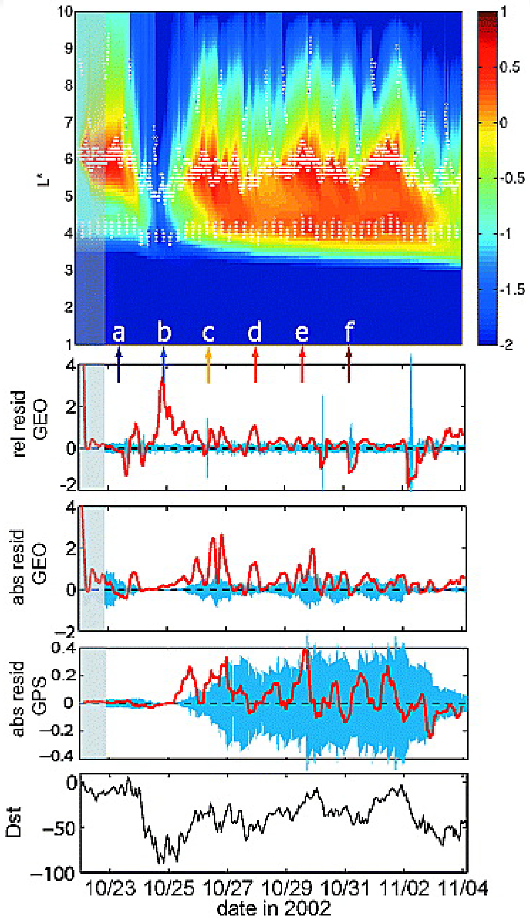
- Associate random draws from the observations with each sample in the prior ensemble – EnKF
- Find the posterior distribution of each pair
- The posterior sample mean and variance converges to exact as the ensemble size is increased
- *Anderson* [2001], *Zhang and Anderson* [2003], *Anderson* [2007], *Anderson* [2009]

Data Assimilation – Radiation Belts

- Objective analysis methods have been developed (e.g. *Friedel et al.*, 2000)
- First data assimilation using a physics-based model was *Naehr and Toffoletto* [2005], using EKF and one-dimensional radial diffusion model
- Demonstrated that the EKF outperformed direct insertion methods for a specific case.
- Direct insertion was used to reconstruct the radiation belts over multiple years (*Maget et al.*, 2007)

Data Assimilation – Radiation Belts

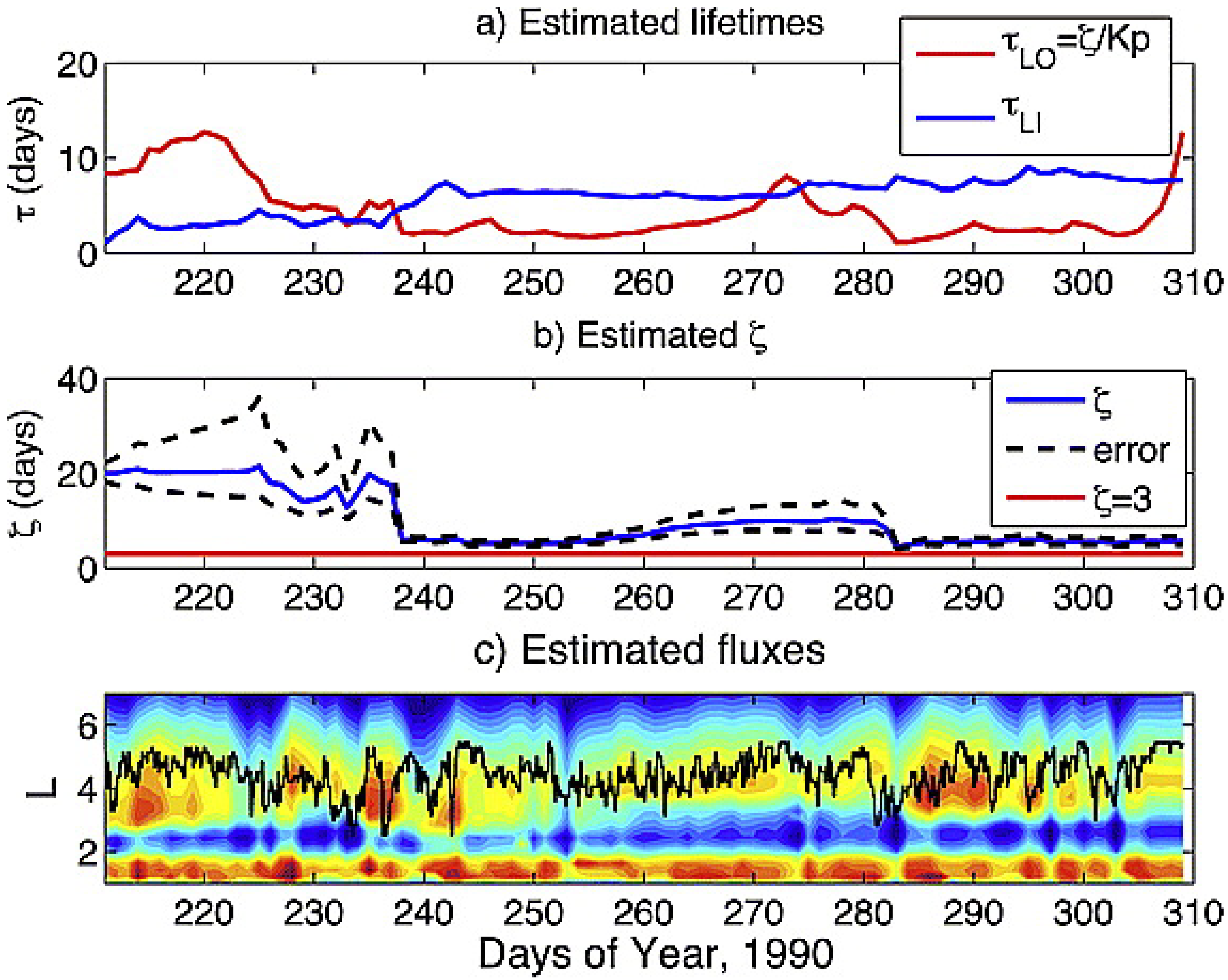




- *Koller et al,* [2007] – Ensemble Kalman Filter, parameter estimation through an augmented state vector (PSD on the outer boundary)
- Three LANL-GEO spacecraft, Polar, and one GPS spacecraft
- One-dimensional radial diffusion model
- Determined that a local heating source was necessary
- Later confirmed by *Shprits et al.* [2007] and *Ni et al.* [2009a], *Daae et al.* [2011], *Schiller et al.* [2012]

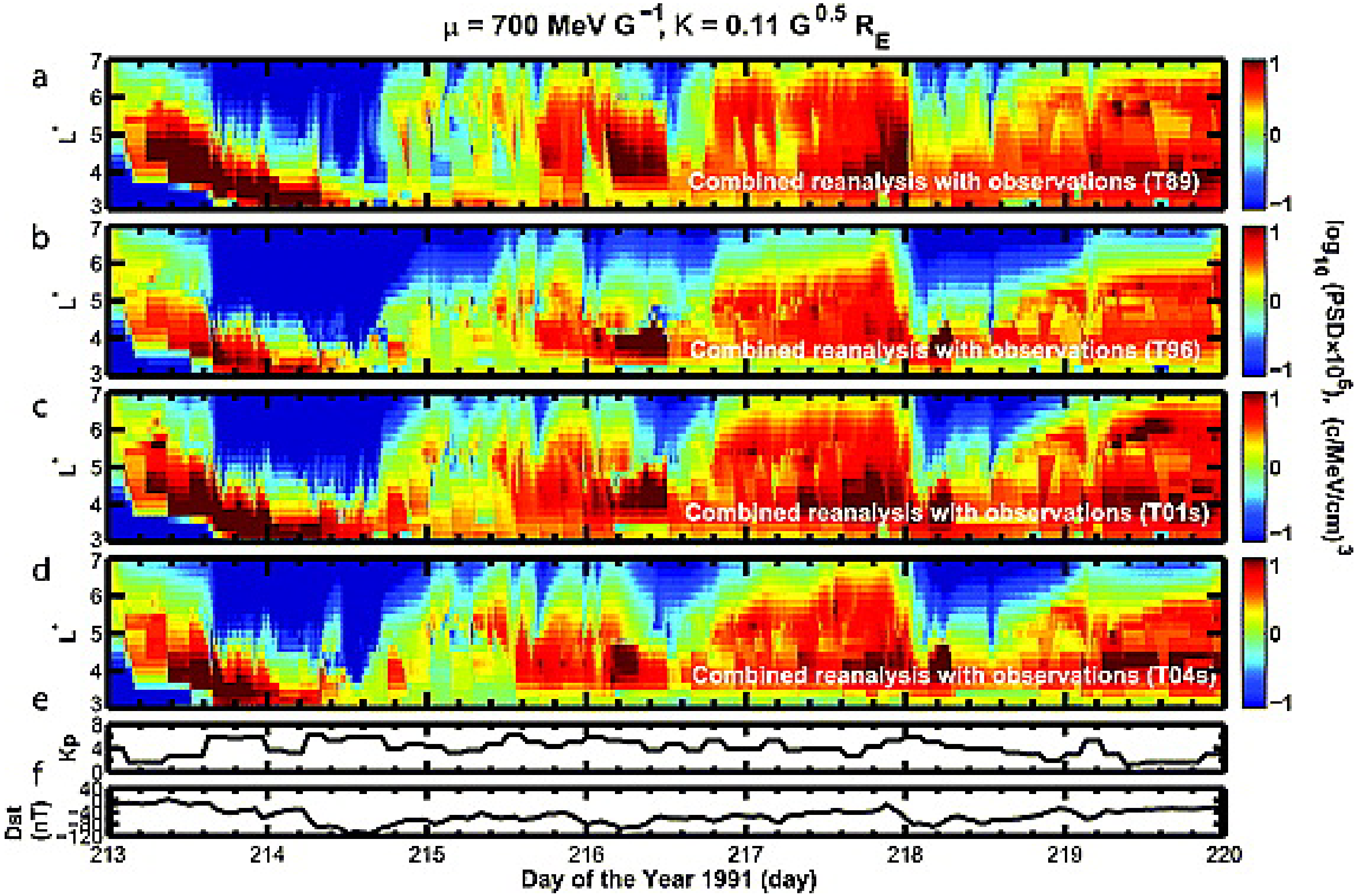
Data Assimilation – Radiation Belts

- *Kondrashov et al.* [2007] - EKF, parameter estimation, CRRES
- Estimated lifetimes and electron number flux

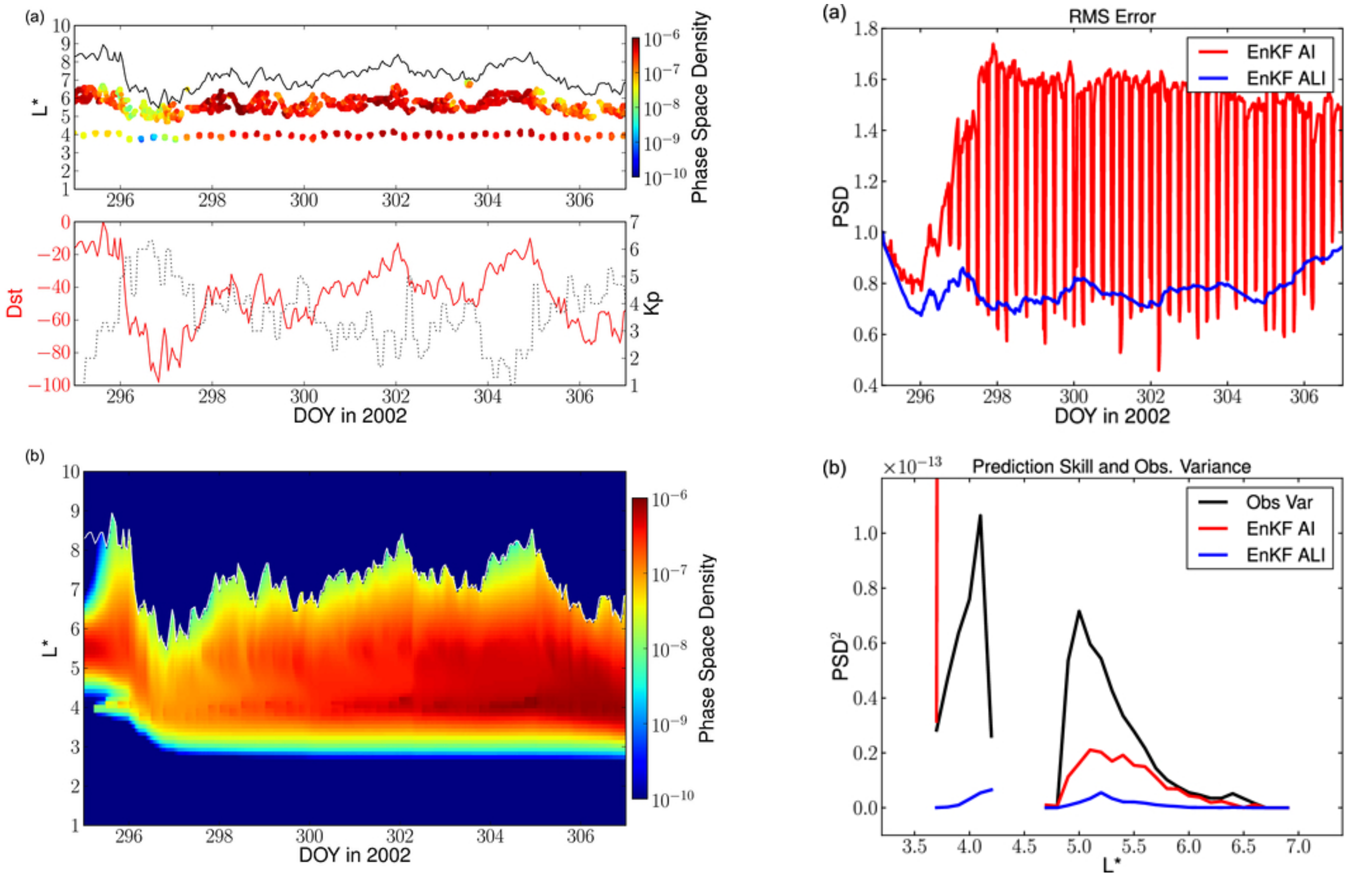


Data Assimilation – Radiation Belts

- *Ni et al.* [2009b] – CRRES, Akebono, GEO1989, GEO1990 – Kalman Filter, DLL
- Tested four global magnetic field models and found relative insensitivity to the model used



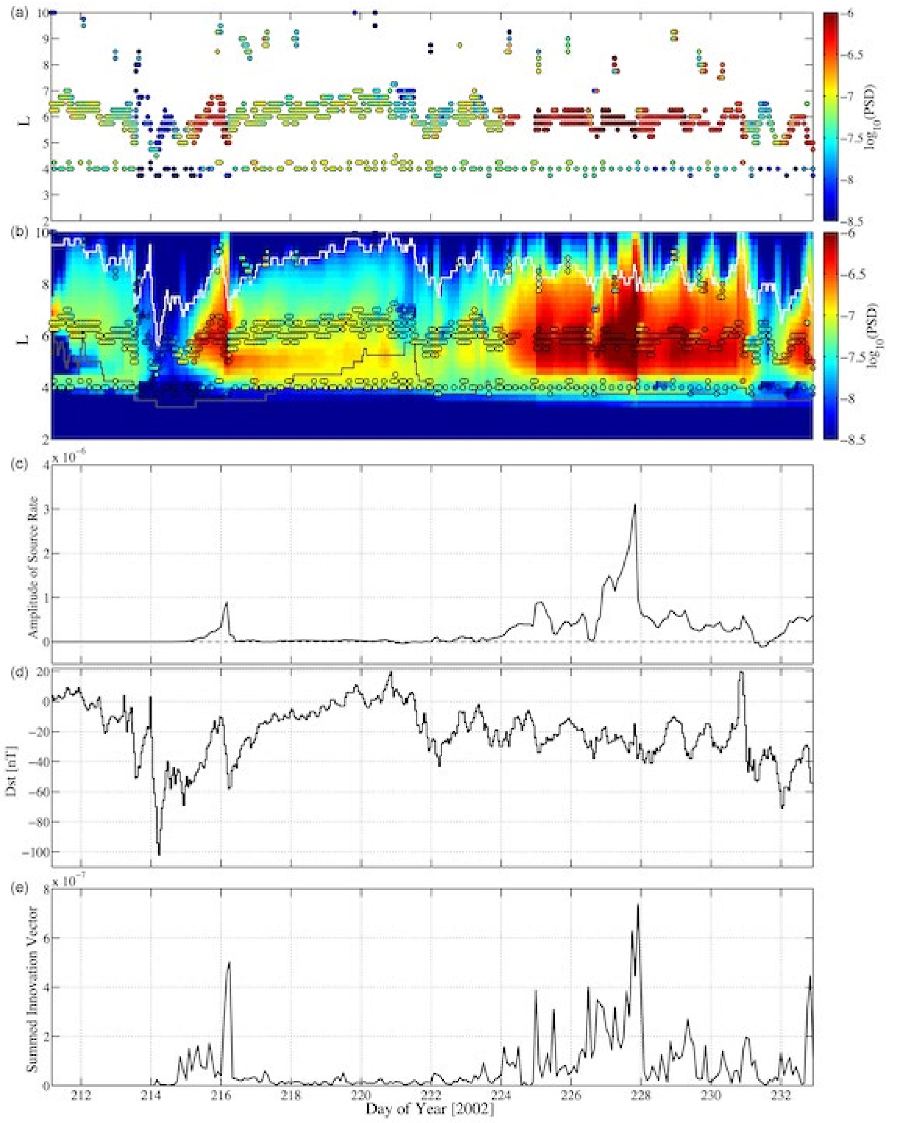
Data Assimilation – Radiation Belts



- *Godinez and Koller* [2012] – 1D radial diffusion, ensemble Kalman filter with localized adaptive inflation EnKF-LAI
- LANL-GEO, Polar, GPS
- One of the best ways to account for model uncertainty

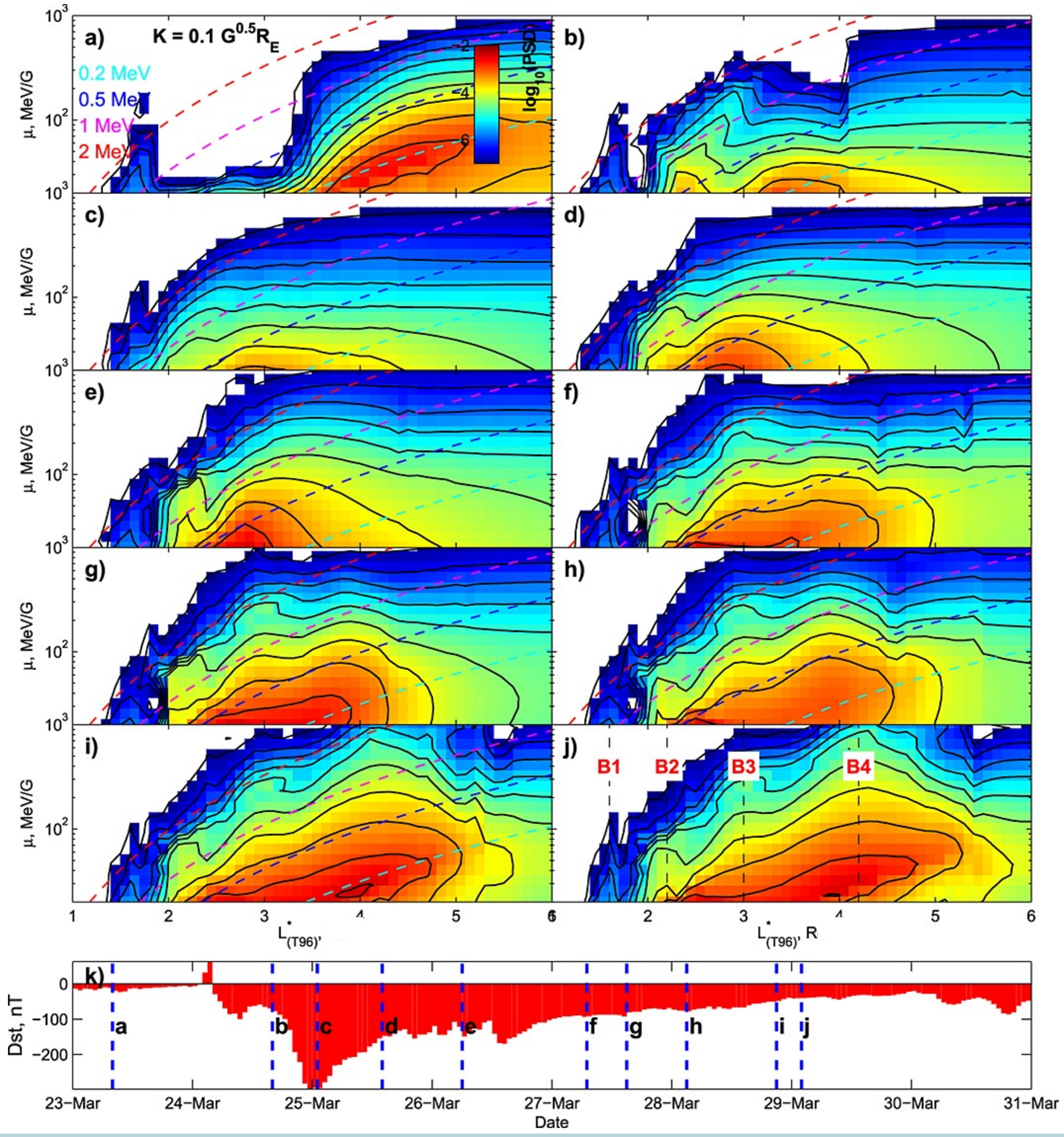
Data Assimilation – Radiation Belts

- *Schiller et al.* [2012] – 1D radial diffusion, ensemble Kalman Filter with augmented state vector
- LANL-GEO, Polar, GPS
- Assimilation (updating) the source term provides a more accurate simulation result, confirming the importance of local heating for enhancement of relativistic electrons



Data Assimilation – Radiation Belts

- *Kellerman et al.* [2014] – CRRES, Akebono, LANL-GEO, GPS split-operator Kalman Filter, 3-D model, VERB 2.0
- Data errors set via PSD matching distributions
- Prompt injection, substorm injection followed by a dropout, leading to a remnant belt, then another injection and local heating
- Additional work by *Podladchikova* [2014a, 2014b], *Kellerman* 2018, *Cervantes* [2020a, 2020b]



Where we need to be

Depends on the question, it is **use-case specific**

A similar approach can be applied for each case: define metrics for success and work towards them – **AUL's**, RL's, or TRL's

Quantified errors for our model inputs/data and adaptive inflation and localization to account for unknown errors

Probabilistic approach to modeling and assimilation

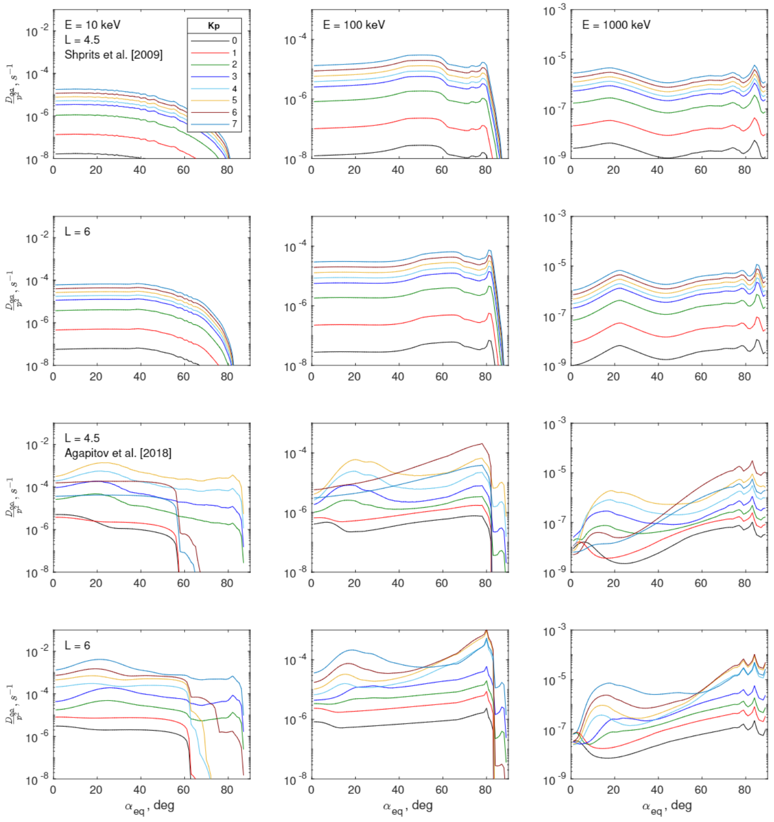
Understand which data are required to specify the environment, such that our models **satisfy the metrics** in each specific use case (e.g. MLT, altitude, energy, time-scale, spatial scale)

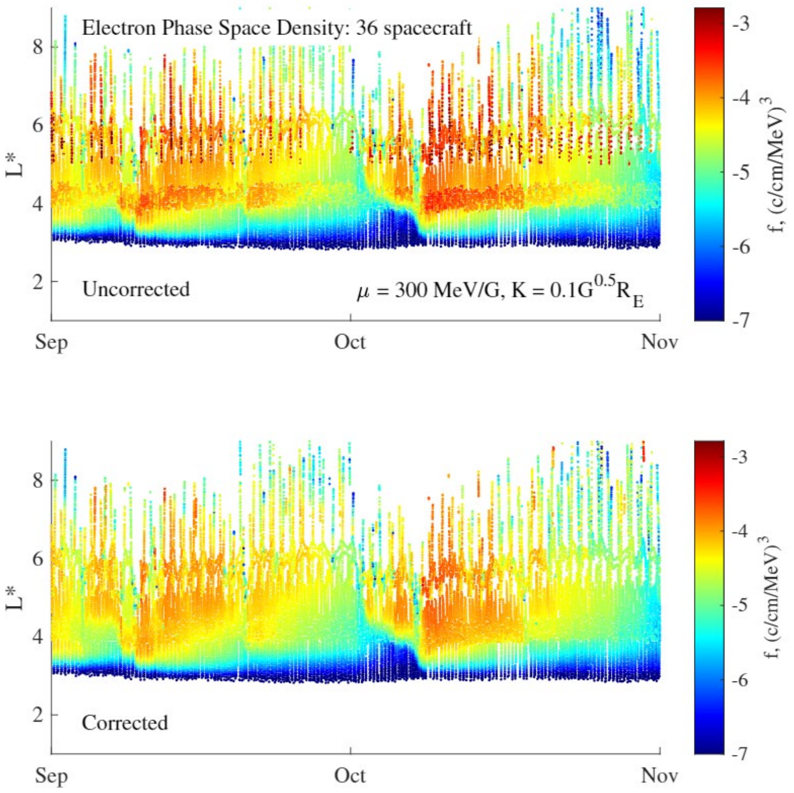
How to get there

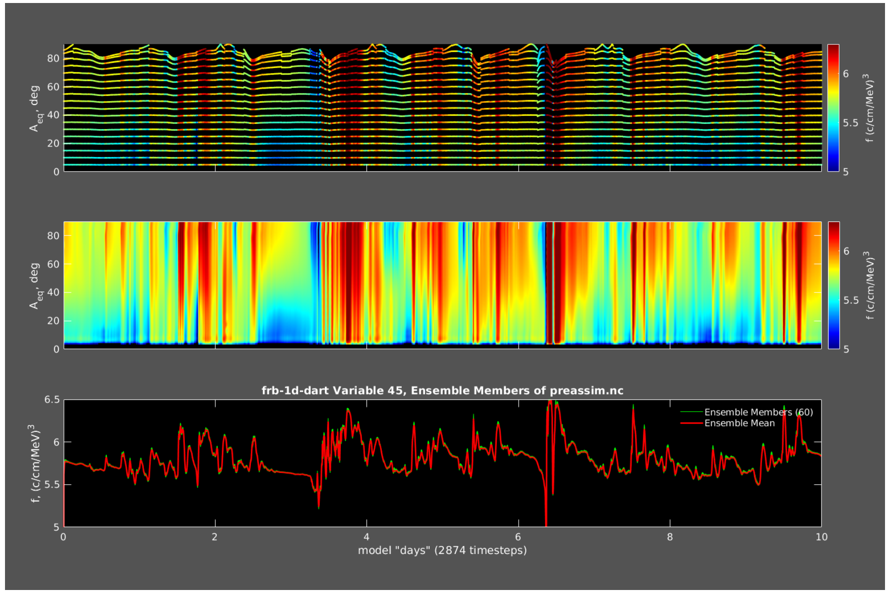
- Define clear **metrics** (AUL's)
- Utilize **OSE's** to determine model sensitivity to existing observational datasets – reference reanalysis
- Employ **OSSE's** to explore the impact and optimal use of observing systems before they are developed and deployed
- **Quantify** the errors in our **data and model inputs**, or estimate them as part of the assimilation system
- **Parameterize non-diffusive and sub-scale effects** that are not captured within our model frameworks
- Employ separate-state-variable-**adaptive inflation** and **localization** schemes to account for unknown model error

Some of My Recent Work

- Implemented **newer wave models** (Agapitov et al., 2018), needed for NR in OSSE
- Statistical error and bias in PSD **computed**
- **Utilizing DART** to harness "tried, tested, and scalable" assimilative schemes







Data Assimilation – Sequential Assimilation

- Start by writing the equation with M, the model operator providing the forecast state
- $x_{b,t+1} = M(x_{a,t}) + \varepsilon_M$
- Assuming unbiased Gaussian errors
- $Q^2 = E(\varepsilon_M^2)$
- $\varepsilon_{b,t+1} = x_{b,t+1} - x_{T,t+1} = M(x_{a,t}) - M(x_{T,t}) + \varepsilon_M = \mathbf{M}\varepsilon_{a,t} + \varepsilon_M$
- $\sigma_{b,t+1}^2 = E(\varepsilon_{b,t+1}^2) = \mathbf{M}^2\sigma_{a,t}^2 + Q^2$
- Here, $\mathbf{M} = \frac{\partial M}{\partial x}$ is the tangent linear model operator
- This formulation can be expanded to multiple dimensions and variables, leading to the formal definition of the Kalman Filter

OSE – Observing System Experiment

- Reference radiation belt reanalysis with all data
- Systematically exclude observational datasets and validate the new reanalysis against the reference reanalysis
- The metrics for success should be clearly defined

OSSE – Observing System Simulation Experiment

For the OSSEs to produce accurate quantitative results, all of the components of the OSSE system must be realistic. (Hoffman and Atlas, 2016) This means that:

1. The NR, which is used to represent the radiation belts, should be generated by a state-of-the-art numerical model
2. There should be realistic differences between the NR model and the model used for assimilation and forecasting;
3. The assimilation methodology must conform to current or future practices
4. Observations should be simulated with realistic coverage and accurately calibrated errors
5. The entire OSSE system must be validated to ensure that the accuracy of analyses and forecasts and that the impact of existing observing systems in the OSSE are comparable to the accuracies and impacts of the same observing systems in the real world