History of Data Assimilation
ISDA-online

International Symposium
on Data Assimilation
22 February 2022

Data Assimilation:
From an Eventful Past to a Bright Future

Michael Ghil
Ecole Normale Supérieure, Paris, and
University of California, Los Angeles

The most recent results are joint with **Eviatar Bach** (ENS) and **Dan Crisan** (ICL)

Joint work with many people over the years
Please see, for names & references
https://dept.atmos.ucla.edu/tcd/ , &
https://www.researchgate.net/profile/Michael_Ghil

Outline

- Data in meteorology, oceanography and space physics
  - *in situ* & remotely sensed
- Basic ideas, data types, & issues
  - how to combine data with models
  - transfer of information
    - between (i) variables & (ii) regions
  - filters & smoothers
- Why does data assimilation work in planetary flows?
  - geostrophic adjustment
  - model & noise parameters – at & below grid scale
  - stability of the forecast-assimilation (FA) process
- Machine learning
- Novel areas of application
  - space physics
  - solid Earth
  - paleoclimate
  - detection & attribution (DADA)
  - macroeconomics
- Concluding remarks and bibliography

Main issues

- The solid earth stays put to be observed, the atmosphere, the oceans, & many other things, do not.
- Two types of information:

  - direct → observations, and

  - indirect → dynamics (from past observations) <= **ML!**

  both have errors.
- Combine the two in (an) optimal way(s)
- Advanced data assimilation methods provide such ways:

  - sequential estimation → the Kalman filter(s), and

  - control theory → the adjoint method(s)
- The two types of methods are essentially equivalent for simple linear systems (the duality principle)

Main issues (continued)

- Their performance differs for large nonlinear systems in:
  - accuracy, and
  - computational efficiency
- Study optimal combination(s) of, as well as improvements over, currently operational methods (4-D Var, EnKF).

Outline

- Data in meteorology, oceanography and space physics
  - *in situ* & remotely sensed
- Basic ideas, data types, & issues
  - how to combine data with models
  - transfer of information
    - between (i) variables & (ii) regions
  - filters & smoothers
- Why does data assimilation work in planetary flows?
  - geostrophic adjustment
  - model & noise parameters – at & below grid scale
  - stability of the forecast-assimilation (FA) process
- Machine learning
- Novel areas of application
  - space physics
  - solid Earth
  - paleoclimate
  - detection & attribution (DADA)
  - macroeconomics
- Concluding remarks and bibliography

Atmospheric data

Drifting buoys: P_s – 267

Cloud-drift: V – 2x2259

Aircraft: V – 2x1100

Ship & land surface: P_s, T_s, V_s – 4x3446

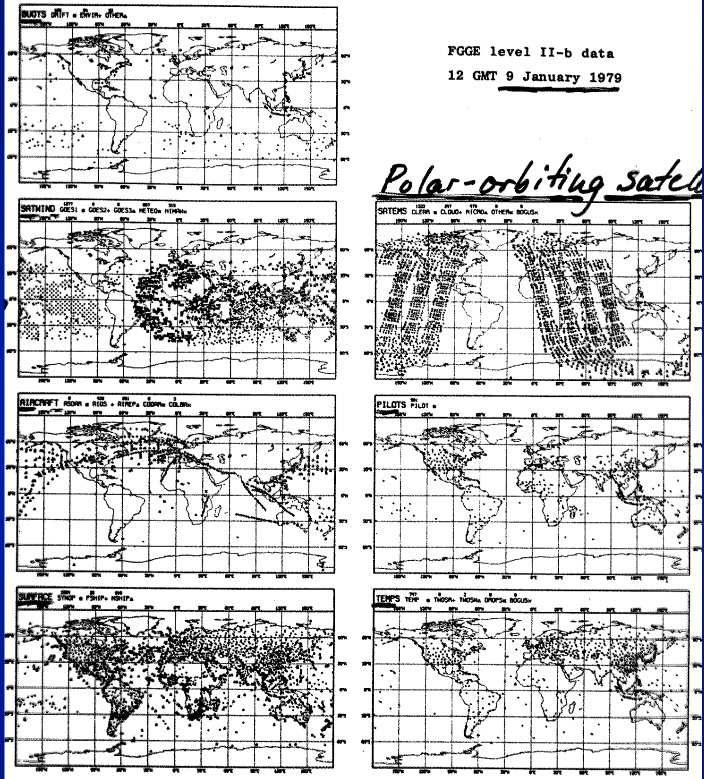


Fig. 1 An example of the different observing systems in use during the Global Weather Experiment.

Abbreviations used:

Aireps	Standard wind observations from aircraft
Asdars, Aids	High quality wind observations from aircraft
Buoys	Surface pressure observations from drifting buoys
Colba	Constant level balloons
Drops	Radiosondes dropped from aircraft
Pilots	Wind measurements from ascending balloons
Satems	Temperature measurements from polar orbiting satellites
Satwind	Cloud drift wind measurements from geostationary satellites
Ships	Surface observations from ships
Synops	Surface observations from land
Temps	Temperature, humidity and wind measurements from radiosondes

Bengtsson, Ghil & Källén (eds.): *Dynamic Meteorology, Data Assimilation Methods* (1981)

Polar orbiting satellites: T – 5x2048

Balloons : V – 2x581x10

Radiosondes : T, V - 3x749x10

Total no. of observations = $0(10^5)$ scalars per 12h–24h

⁂ $0(10^2)$ observations/[(significant *d-o-f*) x (significant Δt)]

Nowadays $0(10^7)$ obs. & more *d-o-f* of interest, too!

Ocean data – past

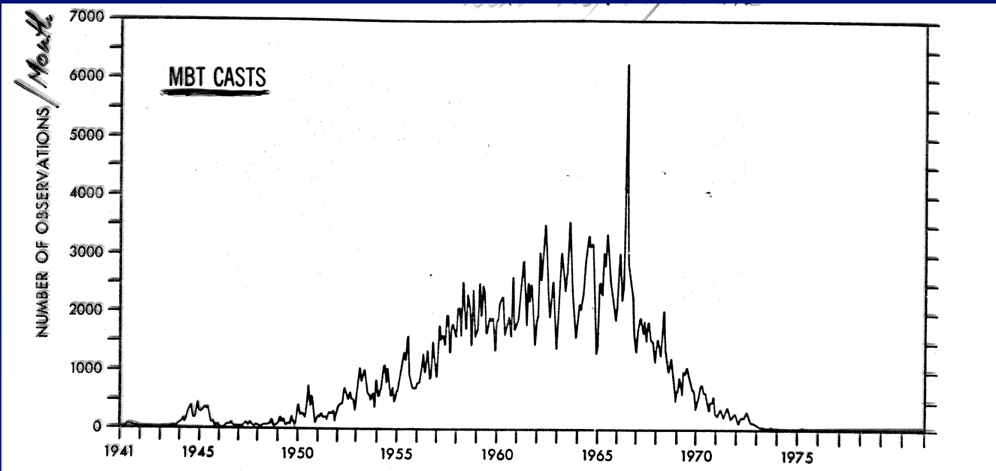


Figure 4.—Time series of MBT casts.

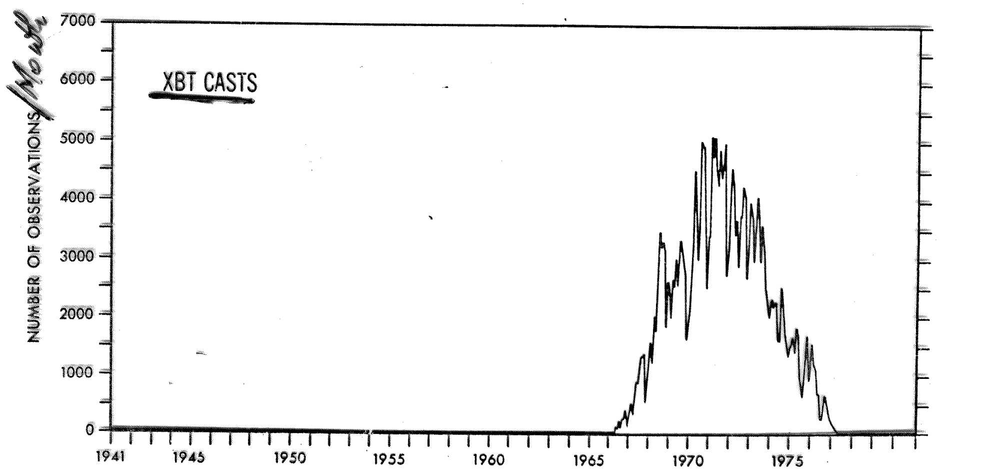
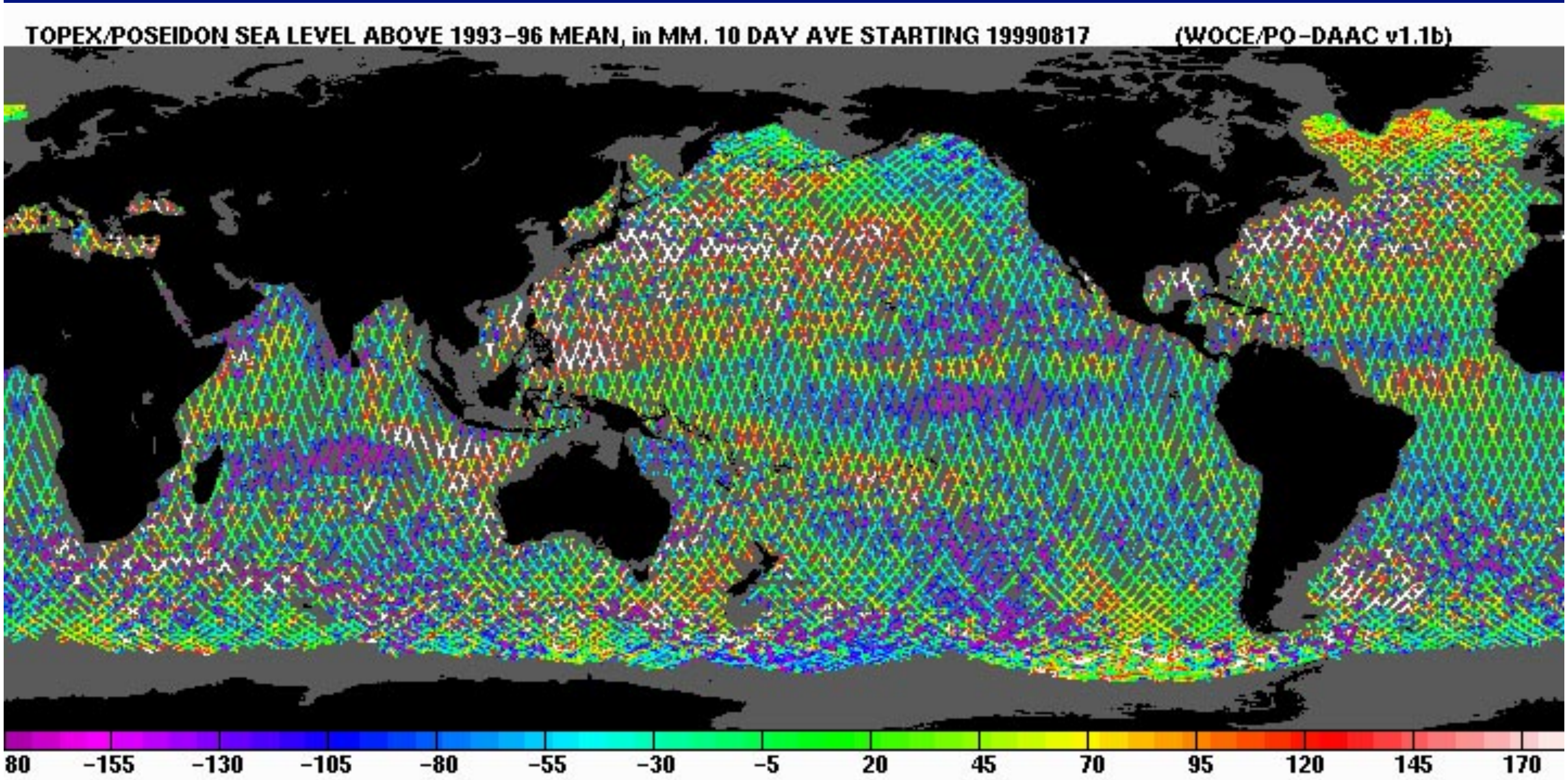


Figure 5.—Time series of XBT casts.

Total no. of
(oceanographic observations)/
(meteorological observations)
= $O(10^{-4})$ for the past; &
= $O(10^{-1})$ for the future :
Syd Levitus (1982).

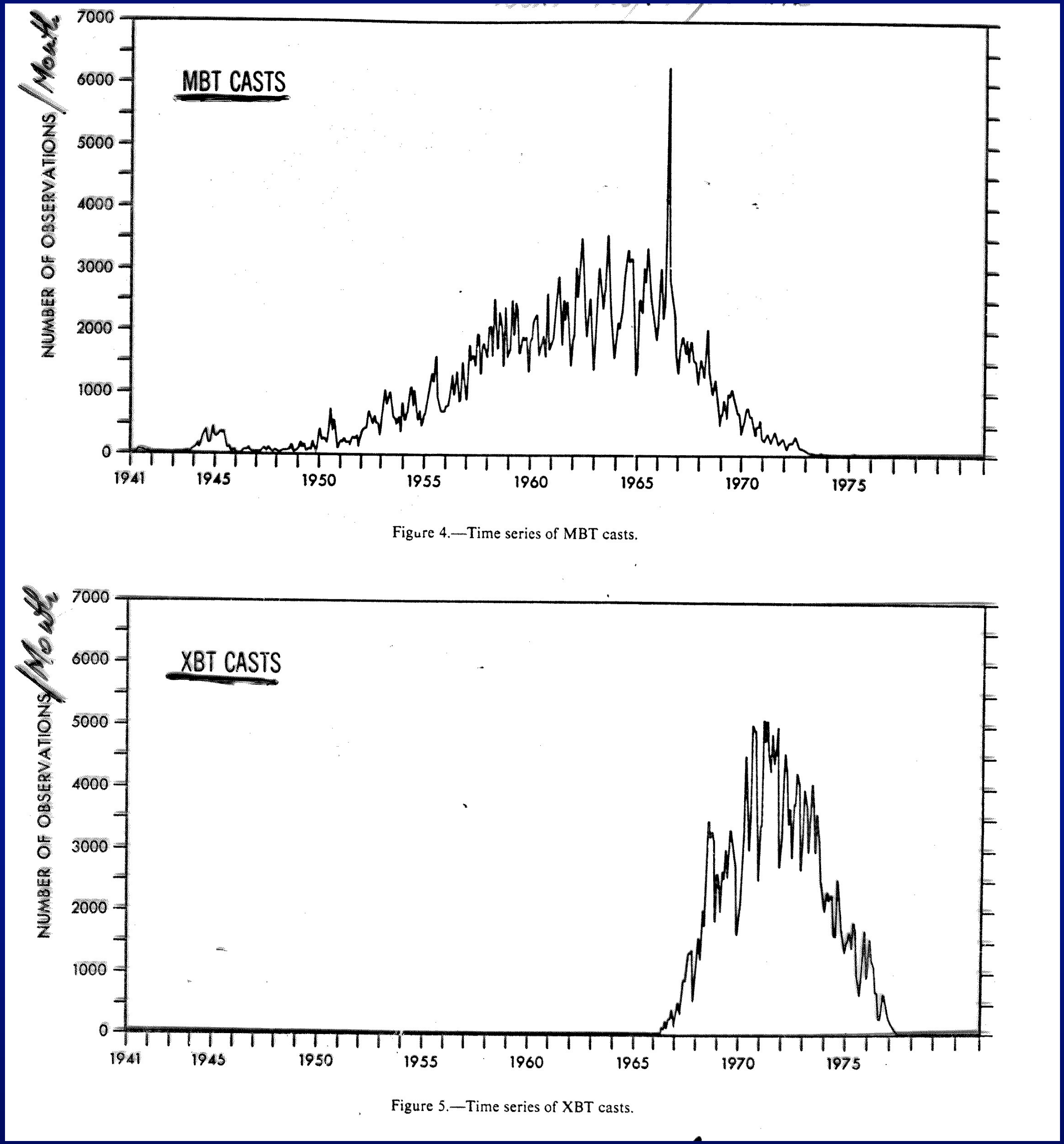
Ocean data – present & future

Altimetry ⇒ sea level; scatterometry ⇒ surface winds & sea state; acoustic tomography ⇒ temperature & density; etc.



Courtesy of Tong (“Tony”) Lee, JPL

Ocean data – past, present & future



Total no. of oceanographic observations/met. ob'sns
= $O(10^{-4})$ for the past; &
= $O(10^{-1})$ for the future :
Syd Levitus (1982).

- 3-D ocean obs. are still not really there: ocean tomography didn't work out but drifters are doing a better & better job.
- Two forms of DA for the coupled system:
  - – weak coupling: models coupled, DA not;
  - – strong coupling: both coupled; could work.

Outline

- Data in meteorology, oceanography and space physics
  - *in situ* & remotely sensed
- Basic ideas, data types, & issues
  - how to combine data with models
  - transfer of information
    - between (i) variables & (ii) regions
  - filters & smoothers
- Why does data assimilation work in planetary flows?
  - geostrophic adjustment
  - model & noise parameters – at & below grid scale
  - stability of the forecast-assimilation (FA) process
- Machine learning
- Novel areas of application
  - space physics
  - solid Earth
  - paleoclimate
  - detection & attribution (DADA)
  - macroeconomics
- Concluding remarks and bibliography

Evolution of DA

TABLE I. CHARACTERISTICS OF DATA ASSIMILATION SCHEMES IN OPERATIONAL USE AT THE END OF THE 1970s[a]

Organization or country	Operational analysis methods	Analysis area	Analysis/forecast
Australia	Successive correction method (SCM)	SH[d]	12 hr
	Variational blending techniques	Regional	6 hr
Canada	Multivariate 3-D statistical interpolation	NH[d] Regional	6 hr (3 hr for the surface)
France	SCM; wind-field and mass-field balance through first guess	NH	6 hr
	Multivariate 3-D statistical interpolation	Regional	
F.R. Germany	SCM. Upper-air analyses were built up, level by level, from the surface	NH	12 hr (6 hr for the surface)
	Variational height/wind adjustment		Climatology only as preliminary fields
Japan	SCM	NH	12 hr
	Height-field analyses were corrected by wind analyses	Regional	
Sweden	Univariate 3-D statistical interpolation	NH	12 hr
	Variational height/wind adjustment	Regional	3 hr
United Kingdom	Hemispheric orthogonal polynomial method		
	Univariate statistical interpolation (repeated insertion of data)	Global	6 hr
U.S.A.	Spectral 3-D analysis	Global	
	Multivariate 3-D statistical interpolation	Global	6 hr
U.S.S.R.	2-D[c] statistical interpolation	NH	12 hr
ECMWF[b]	Multivariate 3-D statistical interpolation	Global	6 hr

[a] After Gustafsson (1981).
[b] European Centre for Medium Range Weather Forecasts.
[c] 2-D is in a horizontal plane.
[d] Southern Hemisphere and Northern Hemisphere, respectively.

Transition from "early" to "mature" phase of DA in NWP:

- no Kalman filter (Ghil *et al.*, 1981 (*))
- no adjoint (Lewis & Derber, *Tellus*, 1985; Le Dimet & Talagrand**Tellus*, 1986)

(*) Bengtsson, Ghil & Källén (Eds., 1981), *Dynamic Meteorology: Data Assimilation Methods.*

M. Ghil & P. M.-Rizzoli (*Adv. Geophys.*, 1991).

Basic ideas of data assimilation and sequential estimation - I

Simple illustration

We want to estimate

T – the temperature of this room, based on the readings

T_1 and T_2 of two thermometers,

by a linear estimate $\hat{T} = \alpha_1 T_1 + \alpha_2 T_2$

The interpretation will be:

T_1 = T^f - **first guess** (of numerical forecast model)

T_2 = T^o - **observation** (R/S, satellite, etc.)

$\hat{T}$ = T^a - **objective analysis**

Basic ideas of data assimilation and sequential estimation - II

If the observations T_1 and T_2 are unbiased, and we want $\hat{T}$ to be unbiased, then $\alpha_1 + \alpha_2 = 1$,

so one can write

$\hat{T} = T_1 + \alpha_2 (T_2 - T_1)$: updating (sequential).

If T_1 and T_2 are uncorrelated, and have known standard deviations,

$A_1 = \sigma_1^{-2}$, $A_2 = \sigma_2^{-2}$,

then the minimum variance estimator(*) is

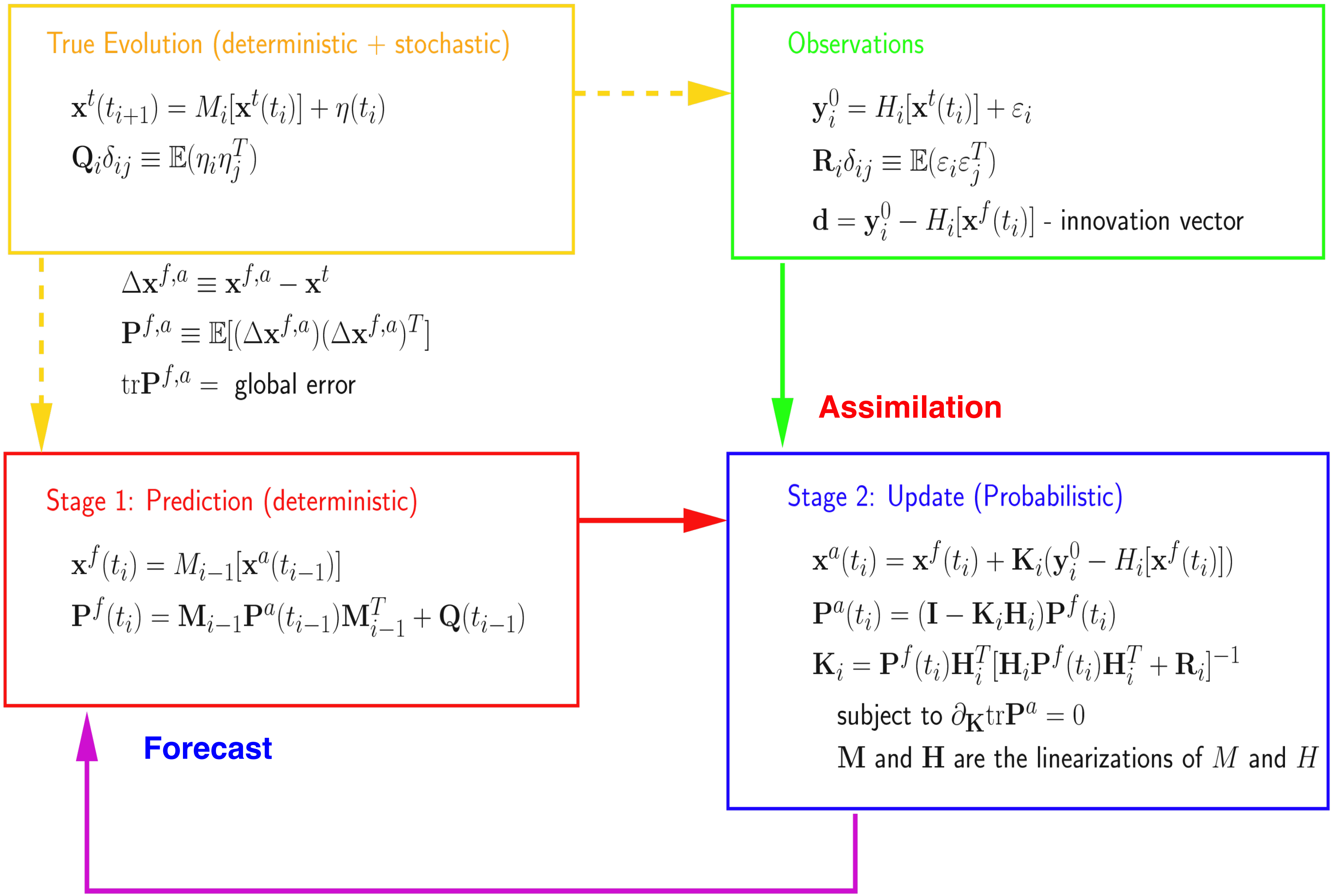
$$\hat{T} = T_1 + \frac{A_2}{A_1 + A_2}(T_2 - T_1),$$

and its accuracy is

$\hat{A} = (A_1 + A_2) \geq \max\{A_1, A_2\}$.

(*) BLUE = Best Linear Unbiased Estimator

(Extended) Kalman Filter (EKF)



Advection of information

Upper panel (NoSat):

Errors advected off the ocean

ϕ_{300}

Lower panel (Sat):

Errors drastically reduced, as info. now comes in, off the ocean

ϕ_{300}

{6h fcst} – {conventional (NoSat)}

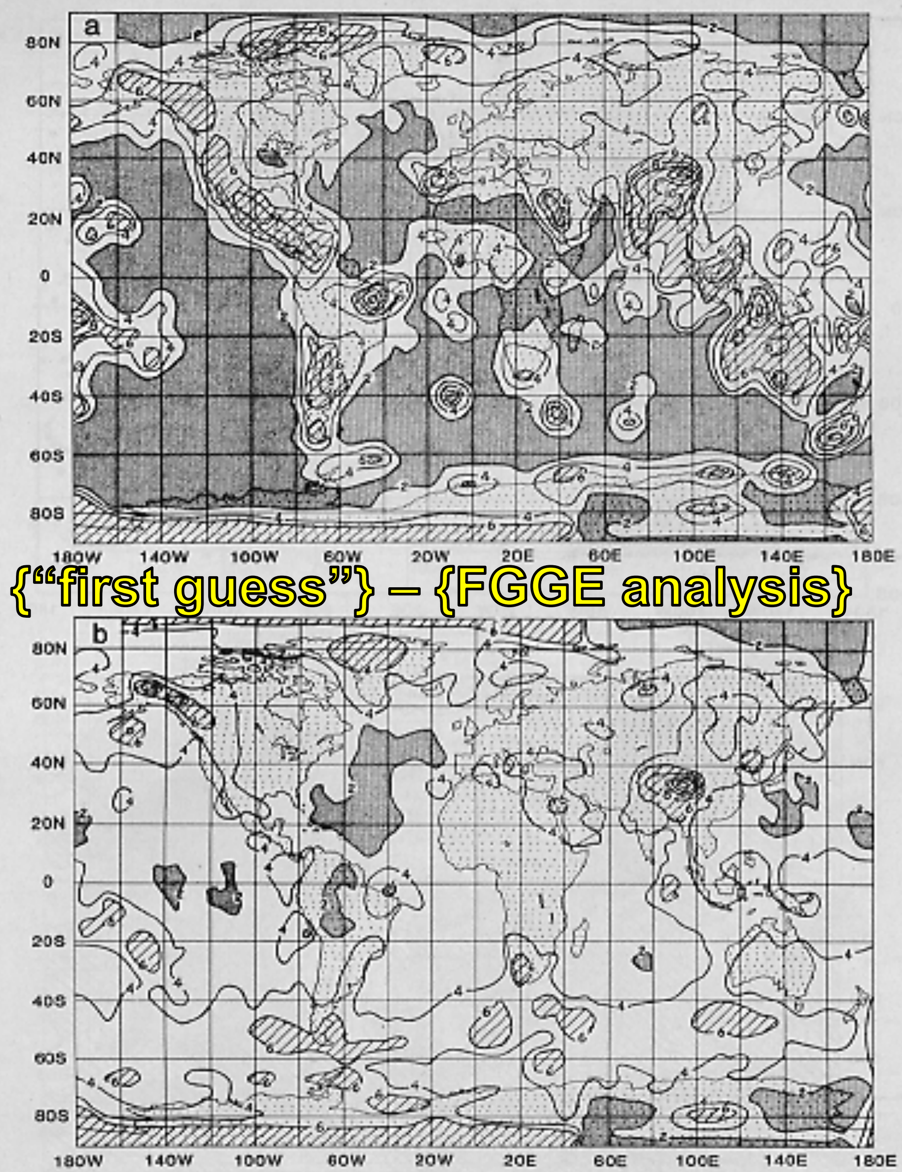


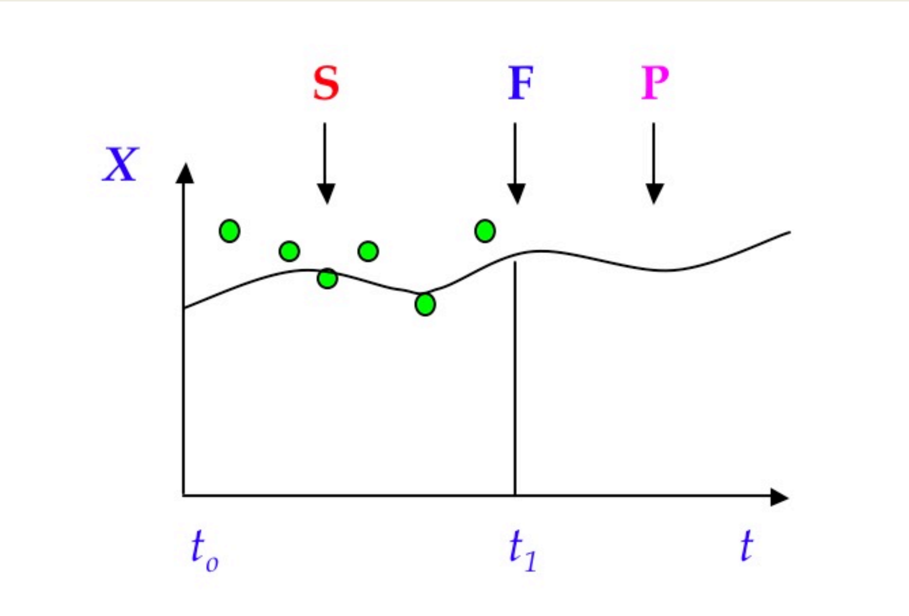
FIG. 5. The rms difference between the 6 h forecast of the 300 mb geopotential height field and the analysis for the period 5–21 January 1979. Contour interval is 20 m. a) Rms difference between the NOSAT analysis and forecast. b) Rms difference between the FGGE analysis and forecast.

Halem, Kalnay, Baker & Atlas

(*Bull. Amer. Meteorol. Soc.*, 1982)

The main products of estimation(*)

- Filtering (F) – "video loops"
- Smoothing (S) – full-length feature "movies"
- Prediction (Pr) – NWP, ENSO
- Parameter estimates (Pe) – all of the above + DADA (**)



Distribute all of this over the Web to scientists, and to the "person in the street"(or on the information superhighway).

In a general way: Have fun (†)

(*) F + S + P: N. Wiener (1949, MIT Press); Pe – a lot recently
(**) DA for Detection & Attribution; (†) or, these days, use machine learning (ML)

Outline

- Data in meteorology, oceanography and space physics
  - *in situ* & remotely sensed
- Basic ideas, data types, & issues
  - how to combine data with models
  - transfer of information
    - between (i) variables & (ii) regions
  - filters & smoothers
- Why does data assimilation work in planetary flows?
  - geostrophic adjustment
  - model & noise parameters – at & below grid scale
  - stability of the forecast-assimilation (FA) process
- Machine learning
- Novel areas of application
  - space physics
  - solid Earth
  - paleoclimate
  - detection & attribution (DADA)
  - macroeconomics
- Concluding remarks and bibliography

A little history – I

S V E N S K A   G E O F Y S I S K A   F Ö R E N I N G E N

VOLUME 2, NUMBER 4 **Tellus** NOVEMBER 1950

A QUARTERLY JOURNAL OF GEOPHYSICS

Numerical Integration of the Barotropic Vorticity Equation

By J. G. CHARNEY, R. FJÖRTOFT[1], J. von NEUMANN

The Institute for Advanced Study, Princeton, New Jersey[2]

(Manuscript received 1 November 1950)

1. Introduction

One of the aims of the meteorological computer project at Princeton is the application of the hydrodynamic equations to forecasting. Initial conditions for this problem would be furnished by the observations at a starting time, t_0. At this time it would be required to know the meteorological variables and some of their space derivatives at certain grid points. This could be accomplished by the normal modes of analysis, that is, by subjectively drawing isopleths, and approximating the derivatives by ratios of finite differences.

JOURNAL OF METEOROLOGY VOLUME 6

OBJECTIVE WEATHER-MAP ANALYSIS

By H. A. Panofsky

New York University[1]

(Manuscript received 7 February 1949)

ABSTRACT

Wind and pressure fields are fitted by third-degree polynomials in areas of the order of 10^6 square miles. Expressions involving derivatives of wind and pressure are computed and the question of computation of geostrophic deviations is re-examined. A method of connecting polynomials in separate areas is investigated.

The following conclusions are drawn:

1. Isopleths and streamlines drawn from the polynomials greatly resemble subjective isopleths and streamlines. In all cases studied, the smoothing seems to be adequate.

2. Horizontal divergence and vertical velocities can be determined as well from the polynomials objectively as by other subjective methods. The errors of observation influence the magnitude of these quantities considerably, but usually do not affect the sign.

3. On the scale of these measurements, reliable pressure gradients can be obtained objectively; however, the Laplacian of pressure is very much affected by the technique of analysis and by observational errors.

4. Reliable values of the geostrophic deviations can be obtained only under favorable conditions. Hence any method of integration of the fundamental equations which requires knowledge of the geostrophic deviations is to be avoided.

A little history – II

Use of Incomplete Historical Data to Infer the Present State of the Atmosphere

J. CHARNEY, M. HALEM[1] AND R. JASTROW[1]

Dept. of Meteorology, Massachusetts Institute of Technology, Cambridge, Mass.

22 August 1969

One of the principal objectives of the Global Atmospheric Research Program (GARP) is the acquisition of data which define the synoptic state of the atmosphere globally for use in long-range prediction. Since all proposed global sounding systems suffer limitations, the concept has arisen of a combination of several such systems permitting trade-offs among the meteorological parameters and between space and time.[2] The system

[1] Institute for Space Studies, Goddard Space Flight Center, NASA, New York, N. Y.

[2] See preface in "Plan for U. S. Participation in the Global Atmospheric Research Program," National Academy of Sciences, Washington, D. C., 1969.

"If it should prove to be possible to obtain the large-scale wind field from temperatures alone, the time-table for the implementation of GARP might be substantially advanced."

Time-Continuous Assimilation of Remote-Sounding Data and Its Effect on Weather Forecasting

M. GHIL[1], M. HALEM AND R. ATLAS

Laboratory for Atmospheric Sciences, NASA Goddard Space Flight Center, Greenbelt, MD 20771

(Manuscript received 7 April 1978, in final form 6 October 1978)

Stabilization of the forecast–assimilation (FA) system – I

Assimilation experiment with the 40-variable Lorenz (1996) model

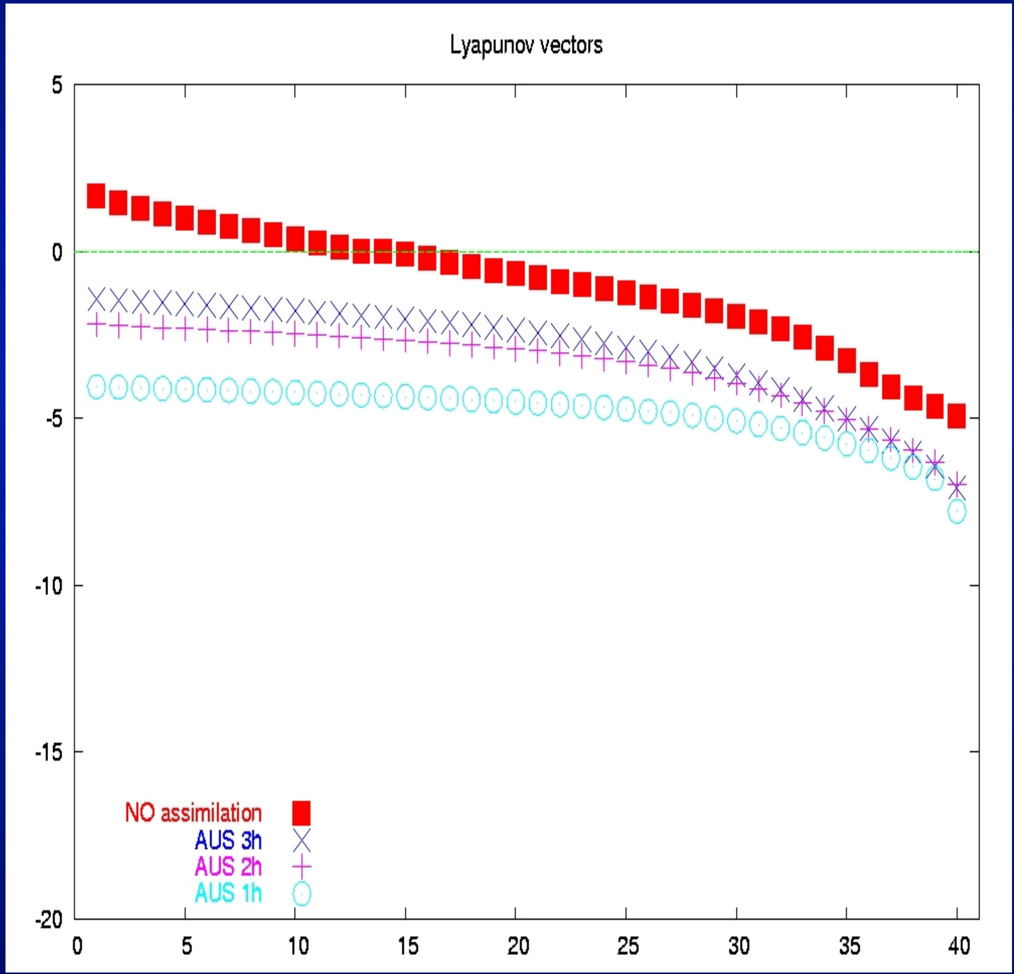
Spectrum of Lyapunov exponents:

Red: free system

Dark blue: AUS with 3-hr updates

Purple: AUS with 2-hr updates

Light blue: AUS with 1-hr updates



Carrassi, Ghil, Trevisan & Uboldi (*Chaos*, 2008)

DA as a random dynamical systems (RDS) problem

- Recall the forecast–assimilation (FA) steps of sequential estimation: in continuous time, & sloppy notation, one can write

$$\hat{x}^{\cdot} = (A - KH)\hat{x} + Kz$$

- Clearly Kz is a forcing by the observations z, with some weights K, optimal ("Kalman") or not – nudging, variational or what not.
- The mathematical framework of "open" dynamical systems is appropriate

  – skew-product flows (G. Sell)

  – pullback (Crauel & Flandoli, L. Arnold) or snapshot (C. Grebogi & E. Ott) attractors

References

H. Crauel and F. Flandoli. Attractors for random dynamical systems. *Probab. Theory Related Fields*, **100**(3):365–393, 1994.

F. J. Romeiras, C. Grebogi, and E. Ott, Multifractal properties of snapshot attractors of random maps, *Phys. Rev. A*, 41:784–799, 1990.

G. R. Sell. Non-autonomous differential equations and dynamical systems. *Trans. Amer. Math. Soc.*, **127**:241–283, 1967.

L.-S. Young, What are SRB measures, and which dynamical systems have them? *J. Stat. Phys.*, **108**, 733–754, 2002.

The sources of nonautonomous dynamics

Physically open vs. closed systems: fluxes of mass, momentum & energy between the system & its surroundings are present or not.

The mathematical framework of **nonautonomous dynamical systems (NDSs)** is appropriate for physically open ones, in which **the fluxes depend explicitly on time:**

- skew-product flows (G. Sell)
  $\dot{x} = f(x,q),\ \dot{q} = g(q),\ x \in \mathbb{R}^d,\ q \in \mathbb{R}^n$, with q the driving force for x.
- pullback (Flandoli, L. Arnold) or snapshot (C. Grebogi & E. Ott) attractors
  $\mathrm{d}X_t = f(X,q)\,\mathrm{d}t + \sigma(X)\,\mathrm{d}W_t$,
  where W_t is a Brownian motion in $\mathbb{R}^d$ and $\mathrm{d}t \sim (\mathrm{d}W)^2$.

More generally, studying **explicit time dependence** in forcing or coefficients requires NDSs.

The term **nonautonomous** is used both for the deterministic case and for a unified perspective on the deterministic & the random case.

The commonality between the two cases is (i) the independence & (ii) the semi-group property of the driving force, whether $q(t)$ or W_t.

Likewise, **pullback attractor (PBA)** is used both for the deterministic & the random case, while in the latter case uses more specifically the phrase **random attractor (RA)**.

Formally, the indexed family $\mathscr{A}$ of all pullback attracting sets $\mathcal{A}_t$ is termed the *pullback attractor* (PBA) of the NDS, if the following two conditions are fulfilled:

(i) each snapshot $\mathcal{A}_t$ is compact and the family $\mathscr{A} = \{\mathcal{A}(t)\}_{t \in R}$ is invariant with respect to the dynamics

$$X(t,s;X_0) \in \mathcal{A}_t \quad \forall s \leq t \text{ and } X_0 \in A_s; \text{ and} \tag{1}$$

(ii) the pullback attraction occurs for all times:

$$\lim_{s \to -\infty} |X(t,s;X_0) - \mathcal{A}_t| \to 0 \quad \forall t. \tag{2}$$

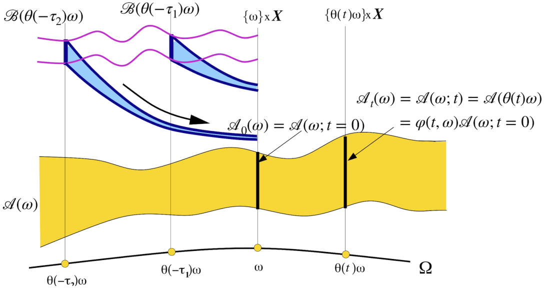


Figure: Schematic diagram of an **RA** $\mathscr{A}(\omega)$. From Ghil, Chekroun & Simonnet (2008).

Stabilization of the forecast–assimilation (FA) system – II

The rigorous proof of the FA process stabilizing the unstable, chaotic dynamics of planetary flows — based on the above ideas, along with other methods from stochastic calculus and nonlinear filtering — is given by Theorems I and II below:

Theorem I (nonlinear). If the coefficients f, h and σ and the measures π_0 and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ satisfy suitable conditions, then there exists $R = R(\pi_0, \mu)$ such that

$$\sup_{t \geq 0} \mathbb{E}[W_2(\pi_t^{\mu}, \pi_t)] \leq R.$$

Theorem II (linear). If the coefficients f, h and σ and the measures π_0 and $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ satisfy suitable conditions, then we have that

$$\lim_{t \to \infty} W_2(\pi_t^{\mu}, \pi_t) = 0.$$

Here, the **Wasserstein distance** is defined by

$$W_2(\mu, \nu) = \left(\inf \mathbb{E}\left[|X - Y|^2\right]\right)^{1/2}, \tag{3}$$

where $\mathbb{E}[Z]$ denotes the expected value of Z and the infimum is taken over all joint distributions of the random variables X and Y with marginals μ and ν, respectively.

D. Crisan and M. Ghil, 2022: Asymptotic behavior of the forecast–assimilation process with unstable dynamics, `arxiv:2202.02862`.

Outline

- Data in meteorology, oceanography and space physics
  - *in situ* & remotely sensed
- Basic ideas, data types, & issues
  - how to combine data with models
  - transfer of information
    - between (i) variables & (ii) regions
  - filters & smoothers
- Why does data assimilation work in planetary flows?
  - geostrophic adjustment
  - model & noise parameters – at & below grid scale
  - stability of the forecast-assimilation (FA) process
- Machine learning
- Novel areas of application
  - space physics
  - solid Earth
  - paleoclimate
  - detection & attribution (DADA)
  - macroeconomics
- Concluding remarks and bibliography

Use of Machine Learning (ML) in DA

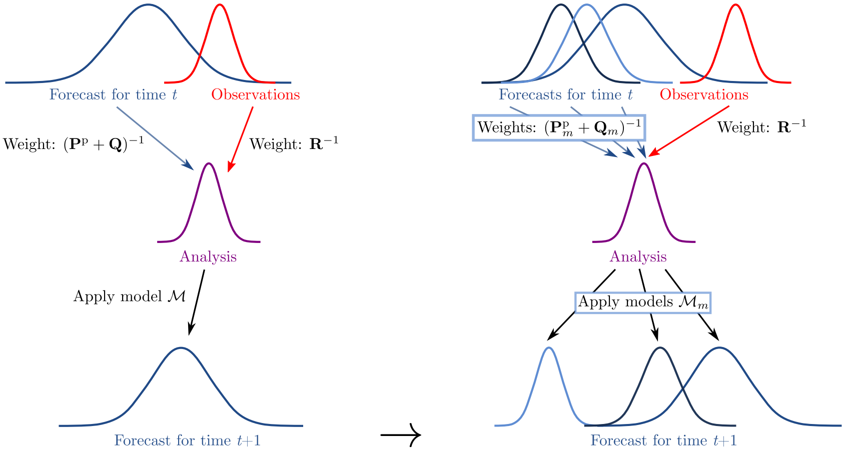
The applications of ML to DA have proliferated in recent years:

- For model error correction (Farchi et al., 2021).
- As the forecast model itself (Chattopadhyay et al., 2021).
  - Can be implemented as an online process, where DA is used to provide better training data for ML, and ML is used to forecast (Brajard et al., 2020).
- To emulate unresolved scales (Brajard et al., 2021).
- For learning the assimilation process directly (McCabe and Brown, 2021).

In forecasting, due to the high dimensionality of the atmosphere and its models, hybrid methods that combine ML forecasts with physical models are necessary (Pathak et al., 2018).

Multi-model data assimilation (MM-DA) – I

- Can we optimally combine ML, physical models, and observations for state estimation?
- Idea: use multi-model DA.



Multi-model data assimilation (MM-DA) – II

- Consider a generalization of the Kalman filter cost function: from

$$(\mathrm{x}-\mathrm{x}^{\mathrm{f}})^T(\mathrm{P}^{\mathrm{f}})^{-1}(\mathrm{x}-\mathrm{x}^{\mathrm{f}})+(\mathrm{Hx}-\mathrm{y})^T\mathrm{R}^{-1}(\mathrm{Hx}-\mathrm{y}), \tag{1}$$

  to

$$\sum_{m=1}^{M}(\mathrm{G}_m\mathrm{x}-\mathrm{x}^{\mathrm{f}}_m)^T(\mathrm{P}^{\mathrm{f}}_m)^{-1}(\mathrm{G}_m\mathrm{x}-\mathrm{x}^{\mathrm{f}}_m)+(\mathrm{Hx}-\mathrm{y})^T\mathrm{R}^{-1}(\mathrm{Hx}-\mathrm{y}), \tag{2}$$

  where each model m has its own forecast state $\mathrm{x}^{\mathrm{f}}_m$ with forecast error covariance matrix $\mathrm{P}^{\mathrm{f}}_m$ and operator G_m.
- MM-DA is also a generalization of the Bayesian formulation of the KF and is the best linear unbiased estimator.
- In Bach and Ghil (2022), we develop a multi-model EnKF and show that
  - it can outperform a multi-model ensemble,
  - as well as the best model in the ensemble.
- Future: test as a hybrid DA and forecasting method with physical and ML models.

Eviatar Bach and Michael Ghil. A multi-model ensemble Kalman filter for data assimilation and forecasting. *arXiv:2202.02272 [physics, stat]*, February 2022.

References

E. Bach and M. Ghil. A multi-model ensemble Kalman filter for data assimilation and forecasting. *arXiv:2202.02272 [physics, stat]*, Feb. 2022.

J. Brajard, A. Carrassi, M. Bocquet, and L. Bertino. Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: a case study with the Lorenz 96 model. *Journal of Computational Science*, page 101171, June 2020. ISSN 1877-7503. doi: 10.1016/j.jocs.2020.101171.

J. Brajard, A. Carrassi, M. Bocquet, and L. Bertino. Combining data assimilation and machine learning to infer unresolved scale parametrization. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2194):20200086, Apr. 2021. doi: 10.1098/rsta.2020.0086.

A. Chattopadhyay, M. Mustafa, P. Hassanzadeh, E. Bach, and K. Kashinath. Towards physically consistent data-driven weather forecasting: Integrating data assimilation with equivariance-preserving spatial transformers in a case study with ERA5. *Geoscientific Model Development Discussions*, pages 1–23, Apr. 2021. ISSN 1991-959X. doi: 10.5194/gmd-2021-71.

A. Farchi, M. Bocquet, P. Laloyaux, M. Bonavita, and Q. Malartic. A comparison of combined data assimilation and machine learning methods for offline and online model error correction. *Journal of Computational Science*, page 101468, Oct. 2021. ISSN 1877-7503. doi: 10.1016/j.jocs.2021.101468.

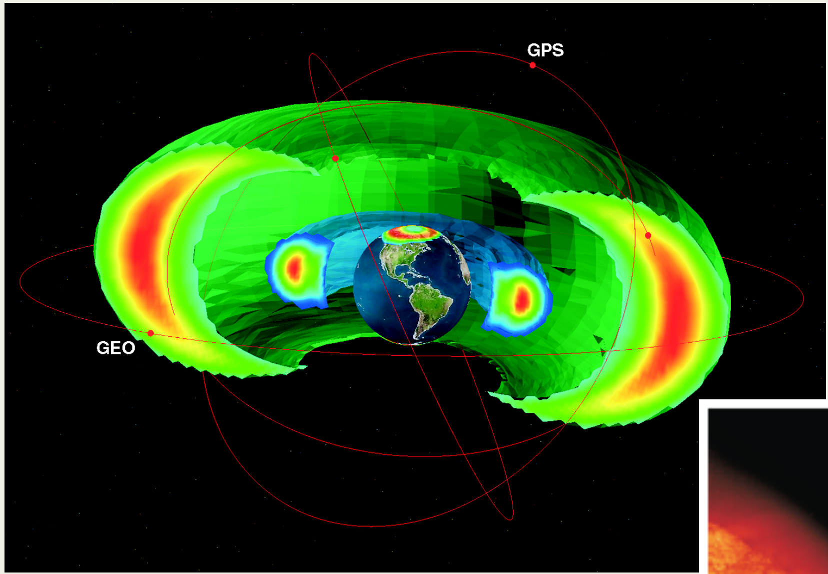
M. McCabe and J. Brown. Learning to Assimilate in Chaotic Dynamical Systems. *arXiv:2111.01058 [nlin]*, Nov. 2021.

J. Pathak, A. Wikner, R. Fussell, S. Chandra, B. R. Hunt, M. Girvan, and E. Ott. Hybrid forecasting of chaotic processes: Using machine learning in conjunction with a knowledge-based model. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(4):041101, Apr. 2018. ISSN 1054-1500. doi: 10.1063/1.5028373.

Outline

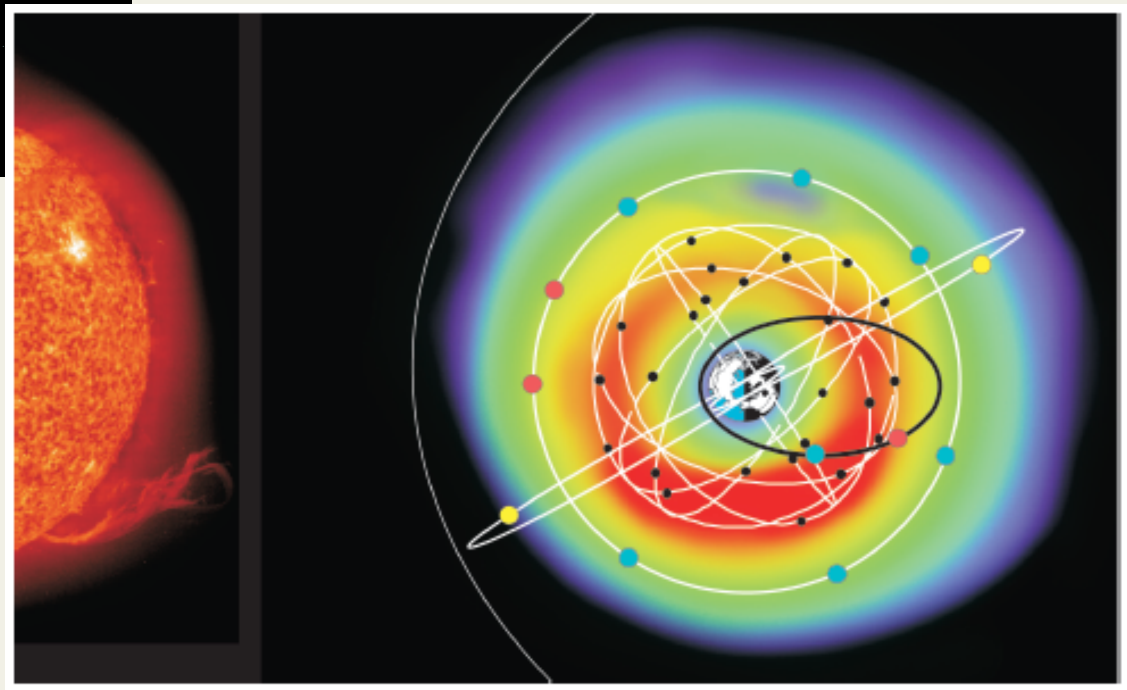
- Data in meteorology, oceanography and space physics
  - *in situ* & remotely sensed
- Basic ideas, data types, & issues
  - how to combine data with models
  - transfer of information
    - between (i) variables & (ii) regions
  - filters & smoothers
- Why does data assimilation work in planetary flows?
  - geostrophic adjustment
  - model & noise parameters – at & below grid scale
  - stability of the forecast-assimilation (FA) process
- Machine learning
- Novel areas of application
  - space physics
  - solid Earth
  - paleoclimate
  - detection & attribution (DADA)
  - macroeconomics
- Concluding remarks and bibliography

Space physics data



Two decades ago …

… and now



Space platforms in Earth’s magnetosphere

Detection & Attribution (D&A)

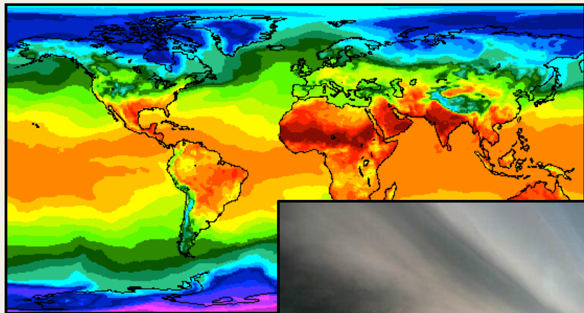
Observed forcing

Natural + anthropogenic forcing

Evidence of causal links?

Observed system behavior

Temperature, precipitation, …
Trends + events



Outline

- Data in meteorology, oceanography and space physics
  - – *in situ &* remotely sensed
- Basic ideas, data types, & issues
  - – how to combine data with models
  - – transfer of information
    - – between (i) variables & (ii) regions
  - – filters & smoothers
- Why does data assimilation work in planetary flows?
  - – geostrophic adjustment
  - – model & noise parameters – at & below grid scale
  - – stability of the forecast-assimilation (FA) process
- Machine learning
- Novel areas of application
  - – space physics
  - – solid Earth
  - – paleoclimate
  - – detection & attribution (DADA)
  - – macroeconomics
- Concluding remarks and bibliography

Computational advances

a) Hardware

- more computing power (CPU throughput)
- larger & faster memory (3-tier)

b) Software

- better numerical implementations of algorithms
- automatic adjoints
- block-banded, reduced-rank & other sparse-matrix algorithms
- better ensemble & particle filters, including localization
- efficient parallelization, ….

How much DA *vs.* forecast?

- Design integrated observing–forecast–assimilation systems!

Observing system design

- **Need no more (independent) observations than *d-o-f* to be tracked:**
  - "features" (Ide & Ghil, 1997a, b, *DAO*);
  - instabilities (Todling & Ghil, 1994 + Ghil & Todling, 1996, *MWR*);
  - trade-off between mass & velocity field (Jiang & Ghil, *JPO*, 1993);
  - **AUS** (Trevisan & colleagues) + **rigorous math** results (Crisan & Ghil).

- The cost of **advanced** DA is **much** less than that of instruments & platforms:
  - at best use DA **instead** of instruments & platforms.
  - at worst use DA to determine which instruments & platforms (**advanced** OSSE)

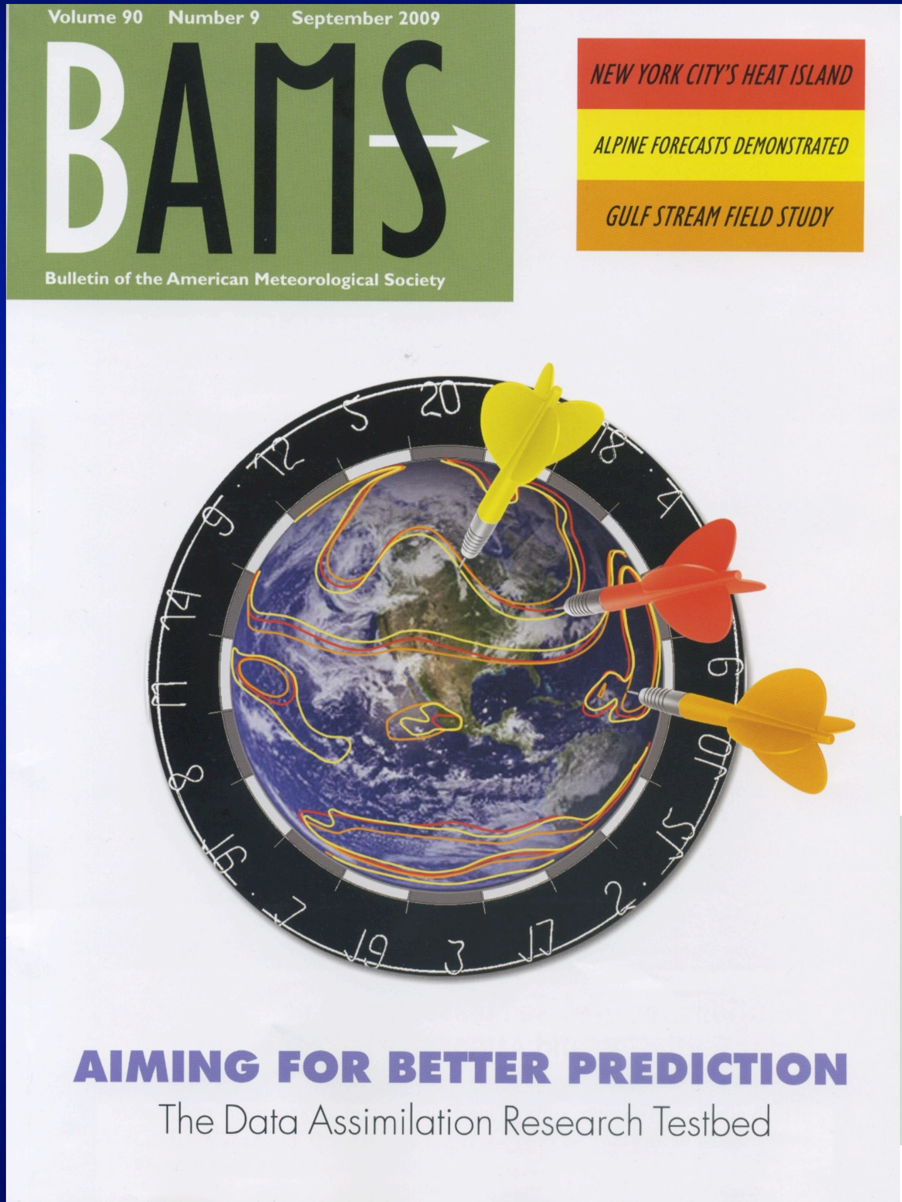
- **Use any** observations, if forward modeling is possible (observing operator H)
  - satellite images, 4-D observations => **ML**;
  - pattern recognition in observations and in phase-space statistics => **ML**.

Conclusions

- Theoretical concepts can play a useful role in devising better practical algorithms, and vice-versa.
- Judicious choices of observations and method can stabilize the forecast-assimilation cycle.
- Trade-off between cost of observations and of data assimilation.

- *Assimilation of ocean data in the coupled O–A system is useful.*
- *They help estimate both ocean and coupling parameters.*
- *Changes in estimated parameters compensate for model imperfections.*

DA Research Testbed (DART)



ARTICLES

THE DATA ASSIMILATION RESEARCH TESTBED

A Community Facility

BY JEFFREY ANDERSON, TIM HOAR, KEVIN RAEDER, HUI LIU, NANCY COLLINS, RYAN TORN, AND AVELINO AVELLANO

DART, developed and maintained at the National Center for Atmospheric Research, provides well-documented software tools for data assimilation education, research, and development.

Data assimilation combines observations with model forecasts to estimate the state of a physical system. Developed in the 1960s (Daley 1991; Kalnay 2003) to provide initial conditions for numerical weather prediction (NWP; Lynch 2006), data assimilation can do much more than initialize forecasts. Repeating the NWP process after the fact using all available observations and state-of-the-art data assimilation produces reanalyses, the best available estimate of the atmospheric state (Kistler et al. 2001; Uppala et al. 2005; Compo et al. 2006). Data assimilation can estimate the value of existing or hypothetical observations (Khare and Anderson 2006a; Zhang et al. 2004). Applications include predicting efficient flight paths for planes that release dropsondes (Bishop et al. 2001) and assessing the potential impact of a new satellite instrument before it is built or launched (Mourre et al. 2006). Data assimilation tools can also be used to evaluate forecast models, identifying quantities that are poorly predicted and comparing models to assess relative strengths and weaknesses. Data assimilation can guide model development by estimating values for model parameters that are most consistent with observations (Houtekamer et al. 1996; Aksoy et al. 2006). Assimilation is now used also for the ocean (Keppene and Rienecker 2002; Zhang et al. 2005), land surface (Reichle et al. 2002), cryosphere (Stark et al. 2008), biosphere (Williams et al. 2004), and chemical constituents (Constantinescu et al. 2007). Assimilation tools under different names are used in other areas of geophysics, engineering, economics, and social sciences.

The Data Assimilation Research Testbed (DART) is an open-source community facility that provides software tools for data assimilation research,

AFFILIATIONS: ANDERSON, HOAR, RAEDER, LIU, COLLINS—NCAR* Data Assimilation Research Section, Boulder, Colorado; TORN—Department of Earth and Atmospheric Sciences, University at Albany, State University of New York, Albany, New York; ARELLANO—NCAR Atmospheric Chemistry Division, Boulder, Colorado
*The National Center for Atmospheric Research is sponsored by the National Science Foundation
CORRESPONDING AUTHOR: Jeffrey Anderson, NCAR, P.O. Box 3000, Boulder, CO 80307-3000
E-mail: jla@ucar.edu

The abstract for this article can be found in this issue, following the table of contents.

DOI:10.1175/2009BAMS2618.1
In final form 8 April 2009

The DA Maturity Index of a Field

- **Pre-DA:** few data, poor models
  - The theoretician: Science is truth, don't bother me with the facts!
  - The observer/experimentalist: Don't ruin my beautiful data with your lousy model!!
- **Early DA:**
  - Better data, so-so models.
  - Stick it (the obs'ns) in – direct insertion, nudging.
- **Advanced DA:**
  - Plenty of data, fine models.
  - E(n)KF, 4-D Var (2nd duality); UKF, particle filters, etc.
- **Post-industrial DA:**

  (Satellite) images → (weather) forecasts, climate "movies" …

Concluding remarks

We've come a long way in 60 years — some advances are laborious and incremental (e.g., sequential vs. control-theoretical methods), but others are fresh and exciting.

The latter include new areas of application

– biology, geomagnetism, paleoclimate, space physics, …, DADA

as well as novel methodological challenges

– multi-scale and multi-model problems (MM-DA)

– various forms of machine learning,

– inverse problems for evolution equations, including climate simulation & sensitivity studies, & uncertainty quantification (UQ)

Technological advances both pose new problems (massive data sets, higher resolution, …) and help solve them.

Overall, it's a brave new world, in which data and models actively speak to each other, and we do so to both: enjoy!

"Miss Peterson, may I go home? I can't assimilate any more data today."

J.B. Handelsman (5/31/1969)

General References

Bengtsson, L., 1975: *Four-Dimensional Data Assimilation of Meteorological Observations*, GARP Publ. Ser., No. 15, WMO-ICSU, Geneva, Switzerland, 76 pp.

Bengtsson, L., M. Ghil and E. Källén (Eds.), 1981. *Dynamic Meteorology: Data Assimilation Methods*, Springer-Verlag, 330 pp.

Daley, R., 1991. *Atmospheric Data Analysis*. Cambridge Univ. Press, Cambridge, U.K., 460 pp.

Ghil, M., and P. Malanotte-Rizzoli, 1991. Data assimilation in meteorology and oceanography. *Adv. Geophys*., **33**, 141–266.

Bennett, A. F., 1992. *Inverse Methods in Physical Oceanography*. Cambridge Univ. Press, 346 pp.

Malanotte-Rizzoli, P. (Ed.), 1996. *Modern Approaches to Data Assimilation in Ocean Modeling*. Elsevier, Amsterdam, 455 pp.

Wunsch, C., 1996. *The Ocean Circulation Inverse Problem*. Cambridge Univ. Press, 442 pp.

Ghil, M., K. Ide, A. F. Bennett, P. Courtier, M. Kimoto, and N. Sato (Eds.), 1997. *Data Assimilation in Meteorology and Oceanography: Theory and Practice*, Meteorological Society of Japan and Universal Academy Press, Tokyo, 496 pp.

Perec, G., 1969: *La Disparition*, Gallimard, Paris.

Van Leeuwen, P.-J., 2009: Particle filtering in geophysical systems, in *Special Math. Advances DA Collection*, *Mon. Wea. Rev*., **137**, 4089–4114.

Asch, M., M. Bocquet and M. Nodet, 2016. *Data Assimilation: Methods, Algorithms, and Applications*, SIAM, Philadelphia.

More geoscientific references

Charney, J. G., M. Halem and R. Jastrow, 1969: Use of incomplete historical data to infer the present state of the atmosphere. *J. Atmos. Sci.*, **26**, 1160–1163.

Evensen, G., 2003: The Ensemble Kalman Filter: theoretical formulation and practical implementation, *Ocean Dynamics*, **53,** 343–367, doi:10.1007/s10236-003-0036-9.

Ghil, M., S. Cohn, J. Tavantzis, K. Bube, and E. Isaacson, 1981: Applications of estimation theory to numerical weather prediction, *Dynamic Meteorology: Data Assimilation Methods*, L. Bengtsson, M. Ghil and E. Källén (Eds.), Springer Verlag, pp. 139–224.

Ghil, M., K. Ide, A. F. Bennett, P. Courtier, M. Kimoto, and N. Sato (Eds.), 1997. *Data Assimilation in Meteorology and Oceanography: Theory and Practice*, Meteorological Society of Japan and Universal Academy Press, Tokyo, 496 pp.

Ide, K., P. Courtier, M. Ghil, and A. Lorenc, 1997: Unified notation for data assimilation: Operational, sequential and variational. *J. Meteor. Soc. Japan,* **75**, 181–189.

Kalnay, E., 2003. *Atmospheric Modeling, Data Assimilation and Predictability*. Cambridge Univ. Press, Cambridge/London, UK, 341 pp.

Kondrashov, D., C.-j. Sun, and M. Ghil, 2008: Data assimilation for a coupled ocean-atmosphere model. Part II: Parameter estimation, *Mon. Wea. Rev.*, **136**, 5062–5076, doi: 10.1175/2008MWR2544.1.

Kondrashov, D., Y. Shprits, and M. Ghil, 2011: Log-normal Kalman filter for assimilating phase-space density data in the radiation belts, *Space Weather*, **9**, S11006, doi:10.1029/2011SW000726.

Roques, L., M. D. Chekroun, M. Cristofol, S. Soubeyrand and M. Ghil, 2014: Parameter estimation for energy balance models with memory, *Proc R. Soc. A*, **470**, 20140349, http://dx.doi.org/10.1098/rspa.2014.0349 .

Some more mathematical references

Bach, E., and M. Ghil, 2022: A multi-model ensemble Kalman filter for data assimilation and forecasting, arXiv:2202.02272.

Bain, A., and D. Crisan. *Fundamentals of Stochastic Filtering*, Springer, 2008.

Carrassi, A., M. Ghil, A. Trevisan and F. Uboldi, 2008: Data assimilation as a nonlinear dynamical systems problem: Stability and convergence of the prediction-assimilation system, *Chaos*, 18(2), 023112, doi: 10.1063/1.2909862.

Crisan, D., and M. Ghil, 2022: Asymptotic behavior of the forecast–assimilation process with unstable dynamics, arXiv:2202.02862 .

Crisan, D., and B. Rozovsky (editors), 2011: *The Oxford Handbook of Nonlinear Filtering*, Oxford University Press.

Farhat, A., E. Lunasin, and E. S. Titi, 2019: A data assimilation algorithm: The paradigm of the 3D Leray-α model of turbulence, in M.B. Ayed, M.A. Jendoubi, Y. Rébaï, H. Riahi, and H. Zaag (editors), *Partial Differential Equations Arising From Physics and Geometry*, Cambridge University Press, pp. 253–273.

Gottwald, G. and S. Reich, 2021: Combining machine learning and data assimilation to forecast dynamical systems from noisy partial observations. *Chaos*, 31(10):101103.

Law, K.J.H., A.M. Stuart, and K.C. Zygalakis, 2015: *Data Assimilation: A Mathematical Introduction*, Springer.

Reich, S., and C. Cotter, 2015: *Probabilistic Forecasting and Bayesian Data Assimilation*, Cambridge University Press.

Reserve slides

A little history – III

Charney et al., *JAS*, 1969: "*If it [were] possible to obtain the large-scale wind field from temperatures alone, the time-table for the implementation of GARP might be substantially advanced.*"

Ghil et al., *MWR*, 1979: "Satellite-sounding data, though promising, provide only one of the basic variables, viz., temperature. Moreover, while the conventional data sets are available simultaneously over the entire globe, temperature sounding data are obtained only in asynoptic, time-continuous fashion. In one of the earliest efforts at using remote-sounding temperatures for NWP, Charney et al. (1969) put forward the conjecture that a complete knowledge of the continuous temperature history of the atmosphere will determine other initial state variables, in particular the winds.

Tellus (1980), *32*, 198–206

The compatible balancing approach to initialization, and four-dimensional data assimilation

By MICHAEL GHIL, *Courant Institute of Mathematical Sciences, New York University, 251 Mercer Street, New York, N.Y. 10012, U.S.A. and Laboratory for Atmospheric Sciences, NASA Goddard Space Flight Center, Greenbelt, MD. 20771, U.S.A.*

A little history – IV

Charney et al., *JAS*, 1969: "*If it [where] possible to obtain the large-scale wind field from temperatures alone, the time-table for the implementation of GARP might be substantially advanced.*"

In a rotating Cartesian (x,y)-coordinate system the linearized shallow-water equations (SWEs) are

$$
\begin{aligned}
u_t + \phi_x - fv &= 0, \\
v_t + \phi_y + fu &= 0, \\
\phi_t + \Phi(u_x + v_y) &= 0,
\end{aligned}
$$

with (u,v) the velocity, φ the geopotential, and f the Coriolis parameter. In this model, knowing the history of the mass field means knowing $\phi(x, y, t)$ and hence (ϕ_t, ϕ_{tt}) at any given instant.

Differentiating the three equations above w.r.t. x, y and t, respectively, and substituting the mixed derivatives u_{xt} and v_{yt} from the first two into the third one, yields:

$$
\Phi f(v_x - u_y) - \Phi(\phi_{xx} + \phi_{yy}) + \phi_{tt} = 0.
$$

This equation, together with the continuity equation above, leads to the Cauchy-Riemann system for the velocity components below:

$$
\begin{aligned}
u_x + v_y &= -\phi_t/\Phi, \\
u_y - v_x &= -f^{-1}(\Delta\phi - \phi_{tt}/\Phi),
\end{aligned}
$$

where Δ is the Laplacian. Similar diagnostic relations for the velocities were obtained for nonlinear SWEs and baroclinic models. E. Titi & colleagues obtained recently rigorous & more general results.

Parameter Estimation

a) Dynamical model

$dx/dt = M(x, \mu) + \eta(t)$

$y^o = H(x) + \varepsilon(t)$

Simple (EKF) idea – augmented state vector

$d\mu/dt = 0$, $X = (x^T, \mu^T)^T$

b) Statistical model

$L(\rho)\eta = w(t)$, L – AR(MA) model, $\rho = (\rho_1, \rho_2, \ldots. \rho_M)$

Examples: 1) Dee *et al*. (*IEEE*, 1985) – estimate a few parameters in the covariance matrix $Q = E(\eta, \eta^T)$; also the bias $<\eta> = E\eta$;

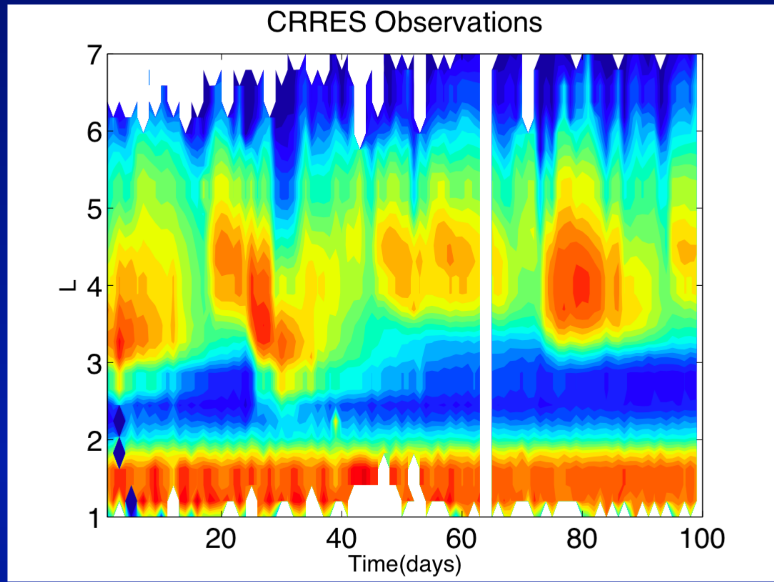
2) POPs - Hasselmann (1982, Tellus); Penland (1989, *MWR*; 1996, *Physica D*); Penland & Ghil (1993, *MWR*)

3) $dx/dt = M(x, \mu) + \eta$: Estimate both M & Q from data (Dee, 1995, *QJ*), Nonlinear approach: Empirical mode reduction (EMR: Kravtsov *et al*., *J. Clim*., 2005; Kondrashov *et al*., *J. Clim*., 2005, *J. Atmos. Sci*., 2006; Kravtsov *et al*., in Palmer & Williams (Eds.), Cambridge U. P., 2010; Strounine *et al*., *Physica D*, 2010)

Parameter estimation for space physics – II

HERRB-1D code (Y. Shprits) – estimating phase-space density *f* and electron lifetime τ_L:

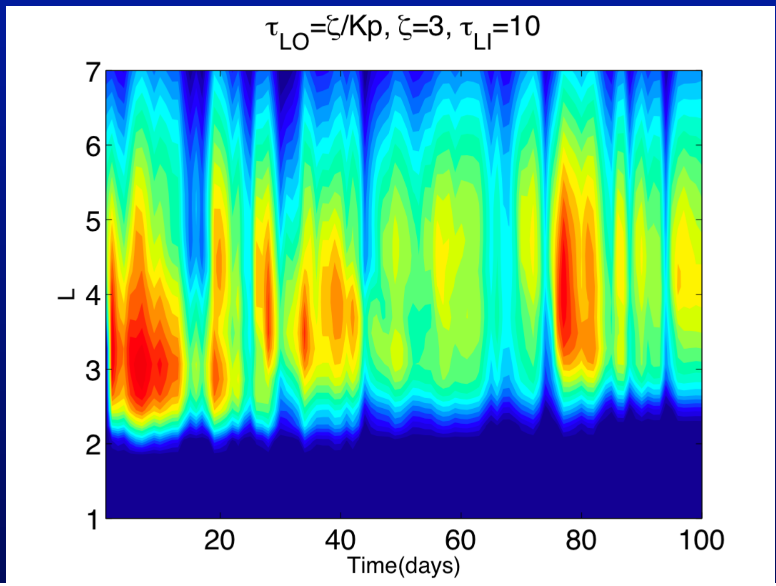
$$\frac{\partial f}{\partial t} = L^2 \frac{\partial}{\partial L}\left(L^{-2} D_{LL} \frac{\partial f}{\partial L}\right) - \frac{f}{\tau_L}$$

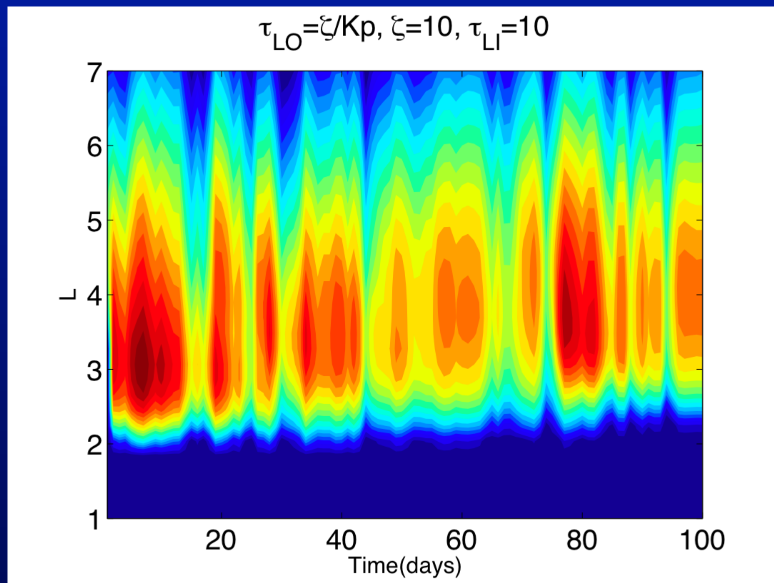


Different lifetime parameterizations for plasmasphere – out/in:

$\tau_{Lo} = \zeta / K_p(t)$; $\tau_{Li} = \text{const.}$

What are the **optimal** lifetimes to match the observations best?





Parameter estimation for space physics – III

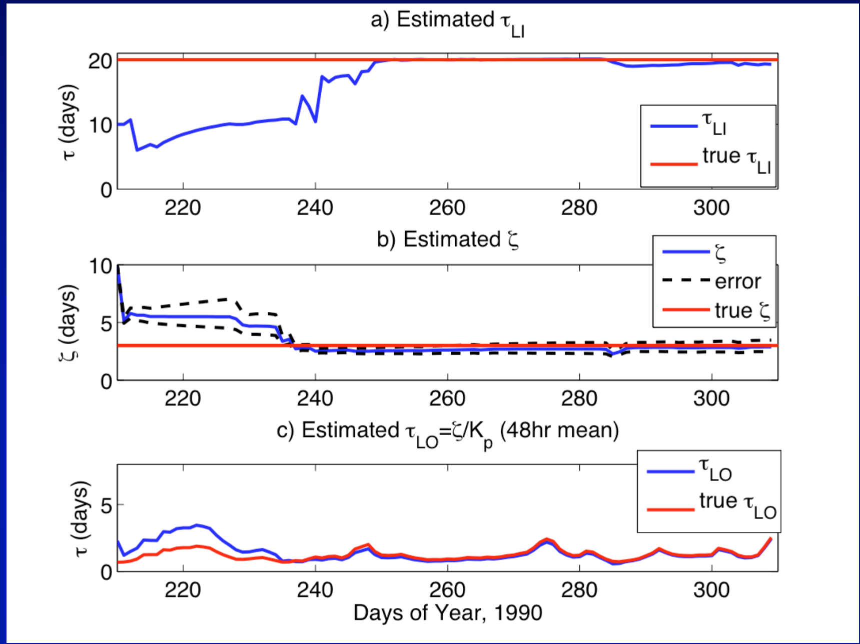
Daily observations from the "truth" —

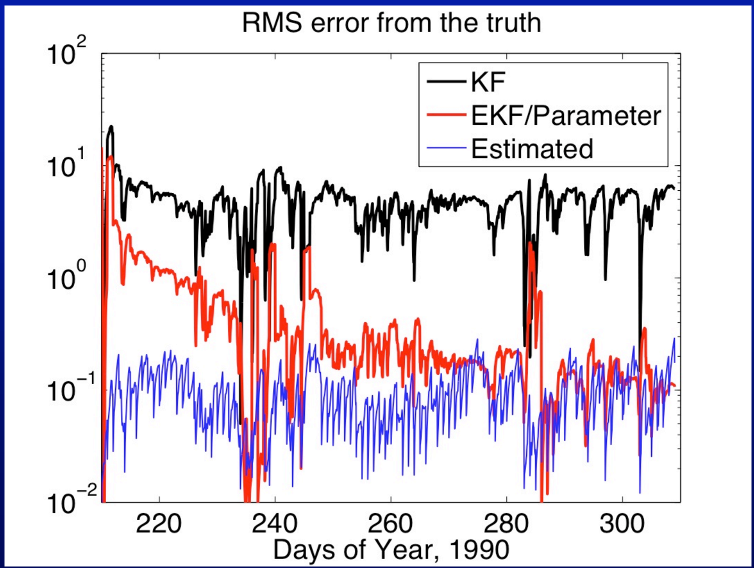
$\tau_{Lo} = \zeta/K_p$, $\zeta = 3$, and $\tau_{LI} = 20$ —

are used to correct the model's "wrong" parameters, $\zeta = 10$ and $\tau_{LI} = 10$.

The estimated error $tr(P^f) \approx$ actual.

When the parameters' assumed uncertainty is large enough, their EKF estimates converge rapidly to the "truth".





Black – actual errors for state estimation only

Red – actual errors for state and parameter estimation

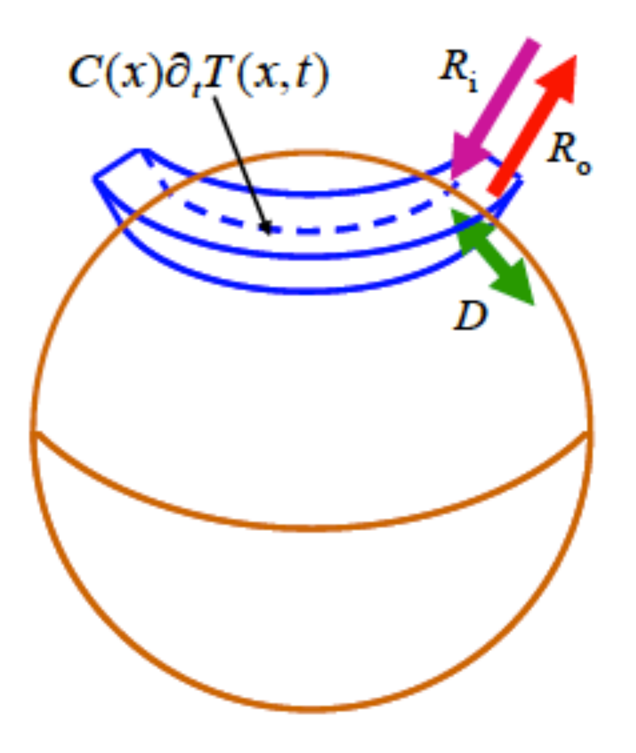
Blue – EKF-estimated error (tr P_k^f)

Parameter estimation for energy balance models with memory (EBMMs) – I

One considers a 1-D paleoclimate model governed by an EBM for zonally averaged surface air temperatures $T(t, x)$:

$$c(x)\,\frac{\partial T}{\partial t} = \frac{\partial}{\partial x}\left(k(x)\,\frac{\partial T}{\partial x}\right) + \mu\,Q(x)[1 - a(x,T)] - g(x,T);$$

here $R_\mathrm{i} = \mu\,Q(x)[1 - a(x,T)]$ is the absorbed solar radiation, with $a = a(x, T)$ the planetary albedo, and $R_\mathrm{o} = g(x, T)$ is the terrestrial radiation, modified by the greenhouse effect, while $0 \le x \le 1$ is a meridional variable. The albedo depends on past temperatures, because of the long time needed to build up and melt ice sheets.



Zonal belt with heat capacity $C(x)$ and temperature $T(t, x)$, subject to incoming radiation R_i, outgoing radiation R_o, and meridional diffusion D.

Ghil (*JAS*, 1976), Bhattacharya, Ghil & Vulis (*JAS*, 1983), Roques, Chekroun *et al.* (*PRS-A*, 2014)

Evolution of DA – II

TABLE IV. DUALITY RELATIONSHIPS BETWEEN STOCHASTIC ESTIMATION AND DETERMINISTIC CONTROL[a]

A. Continuous (linear) Kalman Filter

System Model	$\dot{\mathbf{w}}^t(t) = F(t)\mathbf{w}^t(t) + G(t)\mathbf{b}^t(t), \quad \mathbf{b}^t(t) \sim N[0, Q(t)]$
Measurement Model	$\mathbf{w}^o(t) = H(t)\mathbf{w}^t(t) + \mathbf{b}^o(t), \quad \mathbf{b}^o(t) \sim N[0, R(t)]$
State estimation	$\dot{\mathbf{w}}^a(t) = F(t)\mathbf{w}^a(t) + K(t)[\mathbf{w}^o(t) - H(t)\mathbf{w}^a(t)], \quad \mathbf{w}^a(0) = \mathbf{w}^a_0$
Error covariance propagation (Riccati Equation)	$\dot{P}(t) = F(t)P(t) + P(t)F^T(t) + G(t)Q(t)G^T(t) - K(t)R(t)K^T(t), \quad P(0) = P_0$
Kalman Gain	$K(t) = P(t)H^T(t)R^{-1}(t)$
Initial conditions	$E[\mathbf{w}^t(0)] = \mathbf{w}^a_0, \quad E\{[\mathbf{w}^t(0) - \mathbf{w}^a_0][\mathbf{w}^t(0) - \mathbf{w}^a_0]^T\} = P_0$
Assumptions	$R^{-1}(t)$ exists $E\{\mathbf{b}^t(t)[\mathbf{b}^o(t')]^T\} = 0$
Performance Index	$p^{f,a}(t) = E\{[\mathbf{w}^{f,a} - \mathbf{w}^t][\mathbf{w}^{f,a} - \mathbf{w}^t]^T\}$

B. Continuous (linear) Optimal Control

System Model	$\dot{\mathbf{w}}^t(t) = \tilde{F}(t)\mathbf{w}(t) + \tilde{H}(t)\mathbf{u}(t)$
Measurement Model	$\mathbf{w}^o(t) = \mathbf{w}(t)$ (all system variables are measured)
Performing control	$\mathbf{u}(t) = -\tilde{K}(t)\mathbf{w}(t)$
Performance propagation (Riccati Equation)	$\dot{\tilde{P}}(t) = -\tilde{F}^T(t)\tilde{P}(t) - \tilde{P}(t)\tilde{F}(t) - \tilde{Q}(t) + \tilde{P}(t)\tilde{H}(t)\tilde{K}(t)$
Control Gain	$\tilde{K}(t) = \tilde{R}^{-1}(t)\tilde{H}(t)\tilde{P}(t)$
Terminal conditions	$\mathbf{w}(t_f) = 0$ $\mathbf{P}(t_f) = \tilde{Q}_f$
Cost function	$J[\mathbf{w}, \mathbf{u}] = \mathbf{w}_f^T\tilde{Q}_f\mathbf{w}_f + \int_0^{t_f} [\mathbf{w}^T(t)\tilde{Q}(t)\mathbf{w}(t) + \mathbf{u}^T(t)\tilde{R}(t)\mathbf{u}(t)]\, dt$

C. Estimation-Control Duality

Estimation	Control
t_0 initial time	t_f final time
$\mathbf{w}(t)$ unobservable state variable of random process	$\mathbf{w}(t)$ observable state variable to be controlled
$\mathbf{w}^o(t)$ random observations	$\mathbf{u}(t)$ deterministic control
$F(t)$ dynamic matrix	$\tilde{F}^T(t)$ dynamic matrix
$Q(t)$ covariance matrix for the model errors	$\tilde{Q}(t)$ quadratic matrix defining acceptable errors on model variables
$H(t)$ effect of observations on state variables	$\tilde{H}(t)$ effect of control on state variables
$P(t)$ covariance of estimation error under optimization	$\tilde{P}(t)$ quadratic performance under optimization
$K(t)$ weighting on observation for optimal estimation	$\tilde{K}(t)$ weighting on state for optimal control

[a] (A), Kalman filter as the optimal solution for the former problem; (B), optimal solution for the latter problem; (C), equivalences between the two (after Kalman, 1960, and Gelb, 1974, Section 9.5; courtesy of R. Todling).

Cautionary note:

"Pantheistic" view of DA:

- variational ~ KF;
- 3- & 4-D Var ~ 3- & 4-D PSAS or EnKF.

Fashionable to claim it's all the same but it's not:

- **God** is in **everything**,
- **but the devil is in the details.**

M. Ghil & P. M.-Rizzoli (*Adv. Geophys.*, 1991).

Overall Conclusion

- No observing system without data assimilation and no assimilation without dynamics[a]
- Quote of the day: "You cannot step into the same river[b] twice[c]"
  (Heracleitus, *Trans. Basil. Phil. Soc. Miletus*, *cca.* 500 B.C.)

[a] of state and errors
[B] Meandros
[c] "You cannot do so even once" (subsequent development of "flux" theory by Plato, *cca.* 400 B.C.)

Τα πάντα ρεί = Everything flows