Dynamical Systems, Sequential Estimation, and Estimating Parameters

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

Joint work with
D. Kondrashov and J. D. Neelin, UCLA; C.-J. Sun, NASA Goddard; A. Carrassi, U. of Ferrara; A. Trevisan, ISAC-CNR, Bologna; F. Uboldi, Milano; and many others: please see http://www.atmos.ucla.edu/tcd/
Outline

• Data in meteorology and oceanography
  - *in situ* & remotely sensed
• Basic ideas, data types, & issues
  - how to combine data with models
  - transfer of information
    - between variables & regions
  - stability of the fcst.–assimilation cycle
  - filters & smoothers
• Parameter estimation
  - model parameters
  - noise parameters – at & below grid scale
• Subgrid-scale parameterizations
  - deterministic (“classic”)
  - stochastic – “dynamics” & “physics”
• Novel areas of application
  - space physics
  - shock waves in solids
  - macroeconomics
• Concluding remarks
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);
    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), as well as improvements over currently operational methods (OI, 4-D Var, PSAS, EnKF).
Space physics data

Space platforms in Earth’s magnetosphere
Extended Kalman Filter (EKF)

**Sequential Data Assimilation: Extended Kalman Filtering**

**True Evolution**
\[
\begin{align*}
\frac{dx}{dt} &= m(x^p, t) + d\eta(t) \\
\frac{dQ}{dt} &= \mathbb{E}\left[\left(d\eta(t)\right)\left(d\eta(t)^T\right)\right]
\end{align*}
\]
(2.1)
(2.2)

\[
\Delta x^p = x^p - x^i
\]
(2.3)

\[
P^p = \mathbb{E}\left[\left(\Delta x^p\right)\left(\Delta x^p\right)^T\right]
\]
(2.4)

\[
\text{tr}P^p = \text{global error}
\]

**Stage 1: Prediction**
\[
\frac{d}{dt}x^p = m(x^p, t)
\]
(2.5)

\[
\frac{d}{dt}P^p = MP^p - P^p(M)^T + Q(t)
\]
(2.6)

\[
M = \frac{\partial}{\partial x} m(x, t)|_{x=x^p}
\]
(2.7)

**Stage 2: Update (probabilistic analysis)**
\[
x^i = x^p + Kd^p
\]
(2.12)

\[
P^i = (I - KH)P^p
\]
(2.13)

\[
K = P^pH^T\left(HP^pH^T + R\right)^{-1}
\]
(2.14)

subject to \(\partial_K \text{tr}P^p = 0\)

**Fig. 1.** A flow-chart representation of the EKF method (see Table 1 for definitions of the symbols).
Basic concepts: barotropic model

Shallow-water equations in 1-D, linearized about \((U,0,\Phi)\), \(fU = -\Phi_y\)
\(U = 20\) ms\(^{-1}\), \(f = 10^{-4}\) s\(^{-1}\), \(\Phi = gH\), \(H \approx 3\) km.

\[
\begin{align*}
    u_t + U u_x + \phi_x - f v &= 0 \\
    v_t + U v_x + f u &= 0 \\
    \phi_t + U \phi_x + \Phi u_x - f U v &= 0
\end{align*}
\]

PDE system discretized by finite differences, periodic B. C.
\(H_k\): observations at synoptic times, over land only.

Ghil et al. (1981), Cohn & Dee (Ph.D. theses, 1982 & 1983), etc.
Conventional network

Relative weight of observational vs. model errors

\[ P_\infty = \frac{QR}{Q + (1 - \Psi^2)R} \]

(a) \( Q = 0 \Rightarrow P_\infty = 0 \)

(b) \( Q \neq 0 \Rightarrow (i), (ii) \) and (iii):

(i) “good” observations
\[ R \ll Q \Rightarrow P_\infty \approx R; \]

(ii) “poor” observations
\[ R \gg Q \Rightarrow P_\infty \approx Q/(1 - \Psi^2); \]

(iii) always (provided \( \Psi^2 < 1 \))
\[ P_\infty \leq \min \{R, Q/(1 - \Psi^2)\}. \]
Advection of information

Upper panel (NoSat):

Errors advected off the ocean

Lower panel (Sat):

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

Halem, Kalnay, Baker & Atlas

(BAMS, 1982)
Outline

• Data in meteorology and oceanography
  - *in situ* & remotely sensed

➤ Basic ideas, data types, & issues
  - how to combine data with models
  - stability of the fcst.–assimilation cycle
  - filters & smoothers

• Parameter estimation
  - model parameters
  - noise parameters – at & below grid scale

• Subgrid-scale parameterizations
  - deterministic (“classic”)
  - stochastic – “dynamics” & “physics”

• Novel areas of application
  - space physics
  - shock waves in solids
  - macroeconomics

• Concluding remarks
Error components in forecast–analysis cycle

\[
\begin{align*}
P_f^f & \approx P_a^a + \Delta t (2AP_a^a + Q) \\
\text{first-guess} & \quad \text{analysis} \\
\text{error} & \quad \text{error} \\
& \quad \text{id. twins} \\
& \quad \text{error} \\
& \quad \text{growth} \\
& \quad \text{modeling} \\
& \quad \text{error}
\end{align*}
\]

\[
(\Psi' = e^{A\Delta t} \approx 1 + A\Delta t)
\]

The relative contributions to error growth of

- analysis error
- intrinsic error growth
- modeling error (stochastic?)
Assimilation of observations: Stability considerations

Free-System Dynamics (sequential-discrete formulation): Standard breeding

Forecast state: model integration from a previous analysis

\[ x_{n+1}^f = M(x_n^a) \]

Corresponding perturbative (tangent linear) equation

\[ \delta x_{n+1}^f = M \delta x_n^a \]

Observationally Forced System Dynamics (sequential-discrete formulation): BDAS

If observations are available and we assimilate them:

Evolutive equation of the system, subject to forcing by the assimilated data

\[ x_{n+1}^a = \left[ I - KH \right] M(x_n^a) + Ky_{n+1}^o \]

Corresponding perturbative (tangent linear) equation, if the same observations are assimilated in the perturbed trajectories as in the control solution

\[ \delta x_{n+1}^a = \left[ I - KH \right] M \delta x_n^a \]

- The matrix \((I - KH)\) is expected, in general, to have a stabilizing effect;
- the free-system instabilities, which dominate the forecast step error growth, can be reduced during the analysis step.

Joint work with A. Carrassi, A. Trevisan & F. Uboldi
Stabilization of the forecast–assimilation system – I

Assimilation experiment with a low-order chaotic model
- Periodic 40-variable Lorenz (1996) model;
- Assimilation algorithms: replacement (Trevisan & Uboldi, 2004), replacement + one adaptive obs’n located by multiple replication (Lorenz, 1996), replacement + one adaptive obs’n located by BDAS and assimilated by AUS (Trevisan & Uboldi, 2004).

BDAS: Breeding on the Data Assimilation System
AUS: Assimilation in the Unstable Subspace
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, 2006, submitted
Stabilization of the forecast–assimilation system – III

Assimilation experiment with an intermediate atmospheric circulation model
- 64-longitudinal x 32-latitudinal x 5 levels periodic channel QG-model (Rotunno & Bao, 1996)
- Perfect-model assumption
- Assimilation algorithms: 3-DVar (Morss, 2001); AUS (Uboldi et al., 2005; Carrassi et al., 2006)

Observational forcing ⇒ Unstable subspace reduction

- **Free System**
  - Leading exponent: \( \lambda_{\text{max}} \approx 0.31 \text{ days}^{-1} \);
  - Doubling time \( \approx 2.2 \text{ days} \);
  - Number of positive exponents: \( N^+ = 24 \);
  - Kaplan-Yorke dimension \( \approx 65.02 \).

- **3-DVar–BDAS**
  - Leading exponent: \( \lambda_{\text{max}} \approx 6x10^{-3} \text{ days}^{-1} \);

- **AUS–BDAS**
  - Leading exponent: \( \lambda_{\text{max}} \approx -0.52x10^{-3} \text{ days}^{-1} \)
Outline

• Data in meteorology and oceanography
  - in situ & remotely sensed
• Basic ideas, data types, & issues
  - how to combine data with models
  - filters & smoothers
  - stability of the fcst.-assimilation cycle

➤ Parameter estimation
  - model parameters
  - noise parameters – at & below grid scale

• Subgrid-scale parameterizations
  - deterministic (“classic”)
  - stochastic – “dynamics” & “physics”

• Novel areas of application
  - space physics
  - shock waves in solids
  - macroeconomics

• Concluding remarks
Parameter Estimation

a) Dynamical model

\[
\frac{dx}{dt} = M(x, \mu) + \eta(t)
\]
\[
y^o = H(x) + \varepsilon(t)
\]
Simple (EKF) idea – augmented state vector
\[
\frac{d\mu}{dt} = 0, \quad X = (x^T, \mu^T)^T
\]

b) Statistical model

\[
L(\rho)\eta = w(t), \quad L \sim AR(MA) \text{ model, } \rho = (\rho_1, \rho_2, \ldots, \rho_M)
\]
Examples: 1) Dee et al. (IEEE, 1985) – estimate a few parameters in the covariance matrix \( Q = E(\eta, \eta^T) \); also the bias \( \langle \eta \rangle = E\eta \);


3) \( \frac{dx}{dt} = M(x, \mu) + \eta \): Estimate both \( M & Q \) from data (Dee, 1995, QJ), Nonlinear approach: Empirical mode reduction (Kravtsov et al., 2005, Kondrashov et al., 2005)
Estimating noise – I

\[ Q_1 = Q_{\text{slow}}, \quad Q_2 = Q_{\text{fast}}, \quad Q_3 = 0; \]
\[ R_1 = 0, \quad R_2 = 0, \quad R_3 = R; \]
\[ Q = \sum \alpha_i Q_i, \quad R = \sum \alpha_i R_i; \]
\[ \alpha(0) = (6.0, 4.0, 4.5)^T; \]
\[ Q(0) = 25*I. \]


Poor convergence for \( Q_{\text{fast}} \)?
Estimating noise – II

Same choice of $\alpha(0)$, $Q_i$, and $R_i$ but

$$\Theta(0) = 25 \times \begin{bmatrix} 1 & 0.8 & 0 \\ 0.8 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$


Good convergence for $Q_{fast}$!
Sequential parameter estimation

- "State augmentation" method – uncertain parameters are treated as additional state variables.
- Example: one unknown parameter $\mu$

\[
\begin{bmatrix}
\bar{x}_k \\
\mu_k
\end{bmatrix}
= \begin{bmatrix}
x_k \\
\mu_k
\end{bmatrix}
= \begin{bmatrix}
F(x_{k-1}, \mu_{k-1}) \\
\mu_{k-1}
\end{bmatrix}
+ \begin{bmatrix}
\epsilon_k \\
\mu_k
\end{bmatrix}
\]

\[
y_k^o
= \begin{bmatrix}
H & 0 \\
0 & 0
\end{bmatrix}
\begin{bmatrix}
x_k \\
\mu_k
\end{bmatrix}
+ \epsilon^0
= \bar{H}\bar{x}_k + \epsilon^0
\]

\[
\bar{x}_k^o
= \bar{x}_k^f + \bar{K}(y_k^o - \bar{H}\bar{x}_k^f);
\bar{K}
= \bar{P}^f\bar{H}^T(\bar{H}\bar{P}^f\bar{H}^T + R)^{-1}
\]

- The parameters are not directly observable, but the cross-covariances drive parameter changes from innovations of the state:

\[
\bar{P}^f
= \begin{bmatrix}
P_{xx}^f & P_{x\mu}^f \\
P_{\mu x}^f & P_{\mu\mu}^f
\end{bmatrix};
\bar{K}
= \begin{bmatrix}
P_{xx}^fH^T \\
P_{\mu x}^fH^T
\end{bmatrix}
(HP_{xx}^fH^T + R)^{-1}
\]

- Parameter estimation is always a nonlinear problem, even if the model is linear in terms of the model state: use Extended Kalman Filter (EKF).
Parameter estimation for coupled O-A system

- Intermediate coupled model (ICM: Jin & Neelin, JAS, 1993)
- Estimate the state vector $W = (T', h, u, v)$, along with the coupling parameter $\mu$ and surface-layer coefficient $\delta_s$ by assimilating data from a single meridional section.
- The ICM model has errors in its initial state, in the wind stress forcing & in the parameters.
Coupled O-A Model (ICM) vs. Observations

SSTA for westward-propagating regime: $\delta_s = 0.8, \mu = 0.56$

SSTA for delayed-oscillator regime: $\delta_s = 0, \mu = 0.76$

SSTA in NCAR–NCEP Reanalysis
Convergence of Parameter Values – I

a) Ocean–atmosphere coupling coefficient

b) Surface–layer coefficient

Identical-twin experiments
Convergence of Parameter Values – II

a) Central Pacific SSTA

b) Iterative $\mu$ estimate

c) Iterative $\delta_s$ estimate

Real SSTA data
EKF results with and w/o parameter estimation

SSTA from EKF with fixed $\mu = 0.76$, $\delta_s = 0$

SSTA from EKF with $\mu$ and $\delta_s$ estimation

SSTA difference of EKF $(\mu, \delta_s)$ estimation and NCEP–NCAR

Longitude
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 filters
- 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).

➤ 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;
  - pattern recognition in observations and in phase-space statistics.
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\(^c\)”

\(^a\) 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 ρεει = Everything flows
Evolution of DA – I

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.
Evolution of DA – II

**Cautionary note:**

“Pantheistic” view of DA:
- variational ~ KF;
- 3- & 4-D Var ~ 3- & 4-D PSAS.

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

---

**Table IV. Duality Relationships Between Stochastic Estimation and Deterministic Control**

<table>
<thead>
<tr>
<th>A. Continuous (linear) Kalman Filter</th>
</tr>
</thead>
<tbody>
<tr>
<td>System Model</td>
</tr>
<tr>
<td>$\mathbf{w}(t) = F(t)\mathbf{w}(t) + G(t)\mathbf{b}(t)$</td>
</tr>
<tr>
<td>Measurement Model</td>
</tr>
<tr>
<td>$\mathbf{b}(t) \sim N(0, Q(t))$</td>
</tr>
<tr>
<td>State estimation</td>
</tr>
<tr>
<td>$\mathbf{w}(t) = F(t)\mathbf{w}(t) + H(t)\mathbf{y}(t)$</td>
</tr>
<tr>
<td>Error covariance propagation</td>
</tr>
<tr>
<td>$P(t) = P(t) + P(t)F(t)H(t) + G(t)\mathbf{Q}(t)G(t)^T$</td>
</tr>
<tr>
<td>Kalman Gain</td>
</tr>
<tr>
<td>$K(t) = P(t)H(t)R^{-1}(t)$</td>
</tr>
<tr>
<td>Initial conditions</td>
</tr>
<tr>
<td>$E[\mathbf{w}(0)] = \mathbf{w}_0$, $E[(\mathbf{w}(0) - \mathbf{w}_0)[(\mathbf{w}(0) - \mathbf{w}_0)^T] = \mathbf{P}_0$</td>
</tr>
<tr>
<td>Assumptions</td>
</tr>
<tr>
<td>$R^{-1}(t)$ exists</td>
</tr>
<tr>
<td>Performance Index</td>
</tr>
<tr>
<td>$J[\mathbf{w}, \mathbf{u}] = \mathbf{w}_0^T\mathbf{Q}_0\mathbf{w}_0 + \int_0^T [(\mathbf{w}(t)\mathbf{Q}(t)\mathbf{w}(t) + \mathbf{u}(t)\mathbf{R}(t)\mathbf{u}(t)) dt$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>B. Continuous (linear) Optimal Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>System Model</td>
</tr>
<tr>
<td>$\mathbf{w}(t) = F(t)\mathbf{w}(t) + \mathbf{b}(t)$</td>
</tr>
<tr>
<td>Measurement Model</td>
</tr>
<tr>
<td>$\mathbf{b}(t) = \mathbf{b}(t)$ (all system variables are measured)</td>
</tr>
<tr>
<td>Performing control</td>
</tr>
<tr>
<td>$u(t) = -K(t)\mathbf{w}(t)$</td>
</tr>
<tr>
<td>Performance propagation</td>
</tr>
<tr>
<td>$P(t) = P(t) - K(t)P(t)K(t) + R(t)\mathbf{Q}(t)\mathbf{R}(t)$</td>
</tr>
<tr>
<td>Riccati Equation</td>
</tr>
<tr>
<td>$\mathbf{K}(t) = \mathbf{R}^{-1}(t)\mathbf{H}(t)\mathbf{P}(t)$</td>
</tr>
<tr>
<td>Terminal conditions</td>
</tr>
<tr>
<td>$\mathbf{w}(t) = \mathbf{w}_0$, $\mathbf{P}(t) = \mathbf{Q}_0$</td>
</tr>
<tr>
<td>Cost function</td>
</tr>
<tr>
<td>$J[\mathbf{w}, \mathbf{u}] = \mathbf{w}_0^T\mathbf{Q}_0\mathbf{w}_0 + \int_0^T [(\mathbf{w}(t)\mathbf{Q}(t)\mathbf{w}(t) + \mathbf{u}(t)\mathbf{R}(t)\mathbf{u}(t)) dt$</td>
</tr>
</tbody>
</table>

| C. Estimation-Control Duality         |

<table>
<thead>
<tr>
<th>Estimated duality</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_0$ initial time</td>
</tr>
<tr>
<td>$w(t)$ unobservable state variable of random process</td>
</tr>
<tr>
<td>$w(t)$ random observations</td>
</tr>
<tr>
<td>$F(t)$ dynamic matrix</td>
</tr>
<tr>
<td>$Q(t)$ covariance matrix for the model errors</td>
</tr>
<tr>
<td>$H(t)$ effect of observations on state variables</td>
</tr>
<tr>
<td>$P(t)$ covariance of estimation error under optimization</td>
</tr>
<tr>
<td>$K(t)$ weighting on observation for optimal estimation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Control duality</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_f$ final time</td>
</tr>
<tr>
<td>$w(t)$ observable state variable to be controlled</td>
</tr>
<tr>
<td>$u(t)$ deterministic control</td>
</tr>
<tr>
<td>$\mathbf{F}(t)$ dynamic matrix</td>
</tr>
<tr>
<td>$\mathbf{Q}(t)$ quadratic matrix defining acceptable errors on model variables</td>
</tr>
<tr>
<td>$\mathbf{H}(t)$ effect of control on state variables</td>
</tr>
<tr>
<td>$\mathbf{P}(t)$ quadratic performance under optimization</td>
</tr>
<tr>
<td>$\mathbf{K}(t)$ weighting on state for optimal control</td>
</tr>
</tbody>
</table>

* (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).
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.
  • EKF, 4-D Var (2nd duality).

• Post-industrial DA:
  (Satellite) images --> (weather) forecasts, climate “movies” …


Parameter Estimation

a) Dynamical model

\[
d\frac{x}{dt} = M(x, \mu) + \eta(t) \\
yo = H(x) + \varepsilon(t)
\]
Simple (EKF) idea – augmented state vector
\[
d\mu/dt = 0, \quad X = (x^T, \mu^T)^T
\]

b) Statistical model

\[
L(\rho)\eta = w(t), \quad L - \text{AR(MA) model, } \rho = (\rho_1, \rho_2, ..., \rho_M)
\]
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 (Kravtsov et al., 2005, Kondrashov et al., 2005)