Climate, Assimilation of Data and Models

When Data Fail Us

JUAN M. RESTREPO
Group Leader
Uncertainty Quantification Group

University of Arizona

June 2010

Uncertainty Quantification Group

- Faculty
  - JMR (Math, Physics, Atmospheric Sciences)
  - Shankar Venkataramani, Kevin Lin, Hermann Flaschka (Math)
  - Shlomo Neuman, Larry Winter (Hydrology)
  - Rabi Bhattacharya, Walt Piegorsch (Math, Statistics)
  - Kobus Barnard, Alon Efrat (Computer Science)
- Post-Docs
  - P. Krause (estimation)
  - J. Ramírez (probability)
  - P. Dostert (scientific computing)
  - T.-T. Shieh (variational methods)
- Graduate Students: 15 (Brad Weir, Darin Comeau, Suz Tolwinski)
- Undergraduate Students: 2 (Jason Ditmann)

UQG

Focus Problems

Climate/Weather | Hydrogeology | Computer Vision

Complex Problems. Problems in which progress will result from a combination of statistical methods, and deterministic and stochastic mathematics.

Focus Problems

Climate/Weather

Hydrogeology

Computer Vision

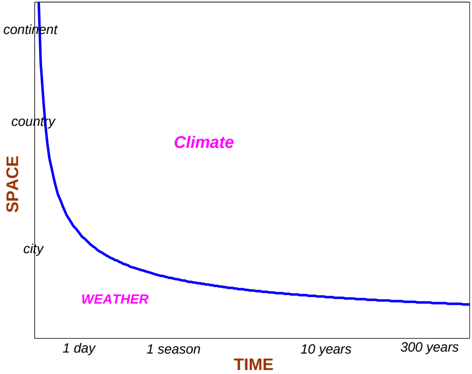
Complex Problems. Problems in which progress will result from a combination of statistical methods, and deterministic and stochastic mathematics.

Focus Problems

Climate/Weather | Hydrogeology | Computer Vision

Complex Problems. Problems in which progress will result from a combination of statistical methods, and deterministic and stochastic mathematics.

Earth's Climate: Forced/Dissipative/Thermodynamic System



The Meteorology vs. Climate Problem

- Climate: poor and/or sparse data. Very large spatio-temporal ranges (oceans). Challenging interations: ocean's interaction with the atmosphere and ice.
  - Aim is to discern variability and understanding dynamics
  - Bayesian data assimilation is viable strategy
  - Need to make more use of models, need to develop parameterizations
  - Need to focus more on non-Gaussian issues as well as the use of data assimilation combined with time of transit dynamics
- Meteorology: lots of data, very nonlinear, very stiff
  - Societal imperative: forecasts
  - Ensemble forecasting
  - Reduced state space representation
  - Large deviation theory could be explored here

The Meteorology vs. Climate Problem

- Climate: poor and/or sparse data. Very large spatio-temporal ranges (oceans). Challenging interations: ocean's interaction with the atmosphere and ice.
  - Aim is to discern variability and understanding dynamics
  - Bayesian data assimilation is viable strategy
  - Need to make more use of models, need to develop parameterizations
  - Need to focus more on non-Gaussian issues as well as the use of data assimilation combined with time of transit dynamics
- Meteorology: lots of data, very nonlinear, very stiff
  - Societal imperative: forecasts
  - Ensemble forecasting
  - Reduced state space representation
  - Large deviation theory could be explored here

The Meteorology vs. Climate Problem

- Climate: poor and/or sparse data. Very large spatio-temporal ranges (oceans). Challenging interations: ocean's interaction with the atmosphere and ice.
  - Aim is to discern variability and understanding dynamics
  - Bayesian data assimilation is viable strategy
  - Need to make more use of models, need to develop parameterizations
  - Need to focus more on non-Gaussian issues as well as the use of data assimilation combined with time of transit dynamics
- Meteorology: lots of data, very nonlinear, very stiff
  - Societal imperative: forecasts
  - Ensemble forecasting
  - Reduced state space representation
  - Large deviation theory could be explored here

The Meteorology vs. Climate Problem

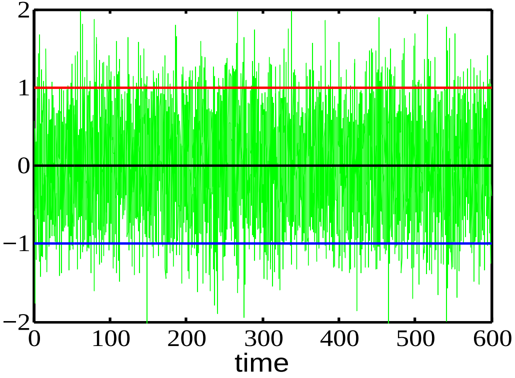
- Climate: poor and/or sparse data. Very large spatio-temporal ranges (oceans). Challenging interations: ocean's interaction with the atmosphere and ice.
  - Aim is to discern variability and understanding dynamics
  - Bayesian data assimilation is viable strategy
  - Need to make more use of models, need to develop parameterizations
  - Need to focus more on non-Gaussian issues as well as the use of data assimilation combined with time of transit dynamics
- Meteorology: lots of data, very nonlinear, very stiff
  - Societal imperative: forecasts
  - Ensemble forecasting
  - Reduced state space representation
  - Large deviation theory could be explored here

The Meteorology vs. Climate Problem

- Climate: poor and/or sparse data. Very large spatio-temporal ranges (oceans). Challenging interations: ocean's interaction with the atmosphere and ice.
  - Aim is to discern variability and understanding dynamics
  - Bayesian data assimilation is viable strategy
  - Need to make more use of models, need to develop parameterizations
  - Need to focus more on non-Gaussian issues as well as the use of data assimilation combined with time of transit dynamics
- Meteorology: lots of data, very nonlinear, very stiff
  - Societal imperative: forecasts
  - Ensemble forecasting
  - Reduced state space representation
  - Large deviation theory could be explored here

What Models tell Us about Data, the Non-Gaussian Issue

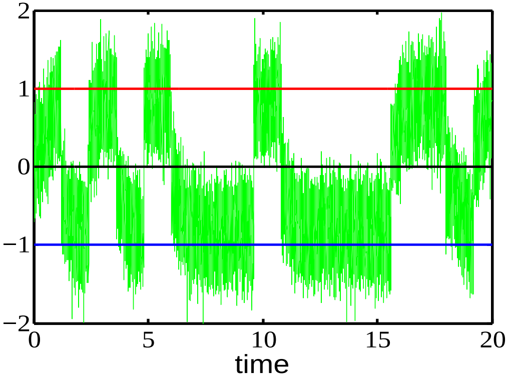
When data fool us...



What Models tell Us about Data, the Non-Gaussian Issue

When data fool us...

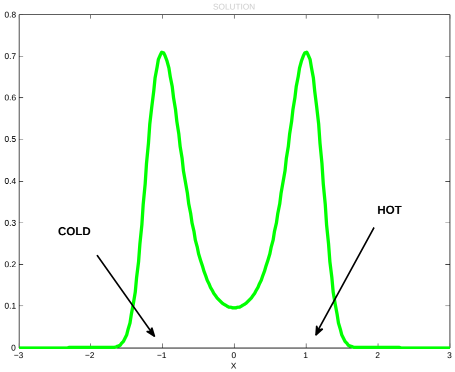
same data, zoomed in



What Models tell Us about Data, the Non-Gaussian Issue

...use our understanding of the dynamics

$$\begin{aligned} dx &= 4x(1-x^2)dt + \kappa dW_t \\ x(0) &= x_0 \end{aligned}$$



Data Assimilation in Climate Studies

Combine information derived from data and models....

Bayes Theorem:

$$P(X|Y) \propto \text{likelihood} \times \text{prior}$$

Least Squares and Kalman Filter/Smoother

$$W(m)x - V = \Theta,$$
$$\Theta \sim \mathcal{N}(0, \sigma).$$

Find $\tilde{x}$, mean, such that $\mathbb{E}(\theta^\top \theta)$ is minimized.
Find the uncertainty $U := \mathbb{E}[(x-\tilde{x})(x-\tilde{x})^\top]$.

Alternatively, can find $\tilde{x}$ and U using a sequential approach, the Kalman Filter/Smoother (RTS Algorithm).

Review in C. Wunsch *The Ocean Inverse Circulation Problem*, Cambridge U. Press

Kalman Filter

Forecast

$$
\begin{aligned}
X^* &= MX(t) + BP(t) \qquad t = 0, 1, \dots, \\
U^* &= MU(t)M^\top
\end{aligned}
$$

Analysis

$$
\begin{aligned}
X(t+1) &= X^* + K(t+1)[Y(t+1) - H(t+1)X^*], \\
U(t+1) &= U^* K(t+1) H(t+1) U^*
\end{aligned}
$$

where the *Kalman Gain Matrix* is

$$
K(t+1) := U^* H(t+1)^\top [H(t+1) U^* H(t+1)^\top]^{-1}
$$

$X(0)$ and $U(0)$ are known.

Review in C. Wunsch *The Ocean Inverse Circulation Problem*, Cambridge U. Press

Nonlinear Non-Gaussian Problems?

Forecast, not much of a problem:

$$X(t+1) = N(X(t), BP(t))$$

But not clear how to propagate uncertainty $U(t+1)$.

Extended Kalman Filter used extensively on nonlinear problems: linearize about $X(t)$ and use closure ideas for moments.

A Nonlinear, Non-Gaussian Example

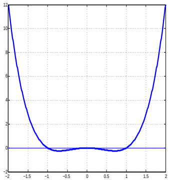
Let $x(t) \in \mathbb{R}^1$, $0 < t \le T$.

Model:

$$
\begin{aligned}
dx &= 4x(1-x^2)dt + \kappa dW_t \\
x(0) &= x_0, \quad \text{known distrib.}
\end{aligned}
$$

Data:

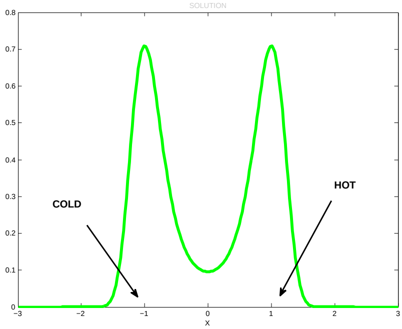
$$
\begin{aligned}
y(t_m) &= x(t_m) + \eta(t_m), \\
\mathbb{E}(\eta_m \eta_l) &= R\delta_{t_m,t_l}, \\
m &= 1, 2, \ldots, M.
\end{aligned}
$$



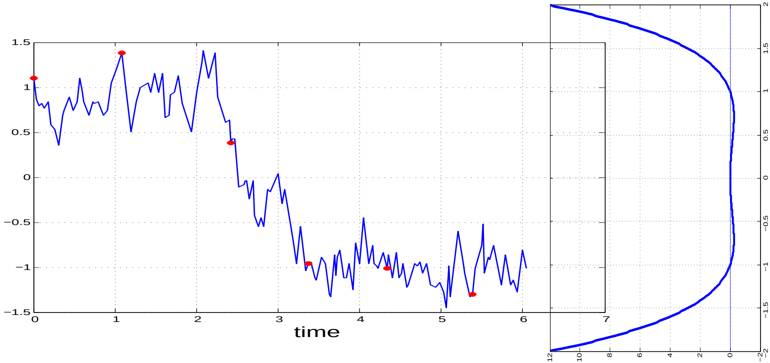
Double-Well Stationary Distribution

$$dx = 4x(1-x^2)dt + \kappa dW_t$$

The Stationary distribution for the double-well problem ($\kappa = 0.5$):



The Observations



A single realization of a random process with known statistics.

Goals of Traditional Data Assimilation

- Find mean history $x_S(t)$ conditioned on observations $\mathbb{E}(x(t)|y_1,...,y_M)$, and the uncertainty $C_s(t) = \mathbb{E}[(x(t)-x_s(t))(x(t)-x_s(t))^T|y_1,...,y_M]$.
- $x_s(t)$ should minimize the $tr(C_s)$
- Guarantee statistical convergence of moments

The EKF Results[1]

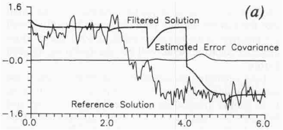


Figure: 10% uncertainty, $\Delta t = 1$.

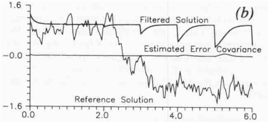


Figure: 20% uncertainty, $\Delta t = 1$.

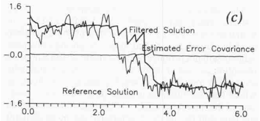


Figure: 20% uncertainty, $\Delta t = 0.25$.

[1] R. Miller, M. Ghil, P. Gauthiez, *Advanced data assimilation in strongly nonlinear dynamical systems*, J. Atmo. Sci. **51** 1037-1056 (1994)

The EKF Results[1]

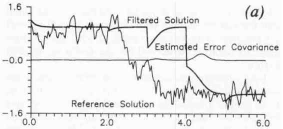


Figure: 10% uncertainty, $\Delta t = 1$.

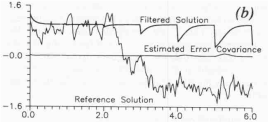


Figure: 20% uncertainty, $\Delta t = 1$.

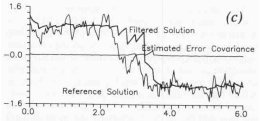


Figure: 20% uncertainty, $\Delta t = 0.25$.

[1] R. Miller, M. Ghil, P. Gauthiez, *Advanced data assimilation in strongly nonlinear dynamical systems*, J. Atmo. Sci. **51** 1037-1056 (1994)

The EKF Results[1]

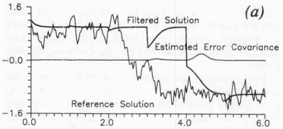


Figure: 10% uncertainty, $\Delta t = 1$.

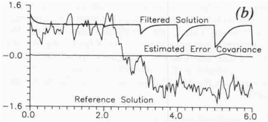


Figure: 20% uncertainty, $\Delta t = 1$.

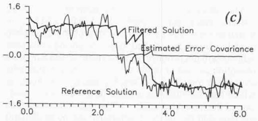


Figure: 20% uncertainty, $\Delta t = 0.25$.

[1] R. Miller, M. Ghil, P. Gauthiez, *Advanced data assimilation in strongly nonlinear dynamical systems*, J. Atmo. Sci. **51** 1037-1056 (1994)

The EKF Results[1]

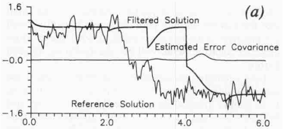


Figure: 10% uncertainty, $\Delta t = 1$.

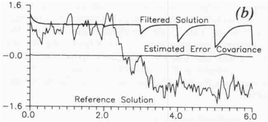


Figure: 20% uncertainty, $\Delta t = 1$.

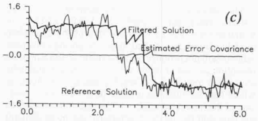


Figure: 20% uncertainty, $\Delta t = 0.25$.

[1] R. Miller, M. Ghil, P. Gauthiez, *Advanced data assimilation in strongly nonlinear dynamical systems*, J. Atmo. Sci. **51** 1037-1056 (1994)

Other Approaches on Nonlinear/Non-Gaussian Problems

- Optimal (variance-minimizer) KSP (Kushner, Stratonovich, Pardoux), early 60's
- 4D-Var/Adjoint (Maximum Likelihood) (Wunsch, McLaughlin, Courtier, late 80's)
- ensemble KF (Evensen, '97)
- Mean Field Variational (Eyink, Restrepo, '01)
- Parametrized Resampling Particle Filter (Kim, Eyink, Restrepo, Alexander, Johnson, '02)
- Path Integral Monte Carlo (Restrepo '07. Alexander, Eyink & Restrepo, '05)
- Diffusion Kernel Filter (Krause, Restrepo, '09)
- Langevin Sampler (A. Stuart, '05)

Other Approaches on Nonlinear/Non-Gaussian Problems

- Optimal (variance-minimizer) KSP (Kushner, Stratonovich, Pardoux), early 60's
- 4D-Var/Adjoint (Maximum Likelihood) (Wunsch, McLaughlin, Courtier, late 80's)
- ensemble KF (Evensen, '97)
- Mean Field Variational (Eyink, Restrepo, '01)
- Parametrized Resampling Particle Filter (Kim, Eyink, Restrepo, Alexander, Johnson, '02)
- Path Integral Monte Carlo (Restrepo '07. Alexander, Eyink & Restrepo, '05)
- Diffusion Kernel Filter (Krause, Restrepo, '09)
- Langevin Sampler (A. Stuart, '05)

KSP: Optimal...so why not use it?

Filter: between t_m and t_{m+1}, solve

$$\begin{aligned} \partial_t \mathscr{P} &= -\partial_x [f(x)\mathscr{P}] + \frac{1}{2}\kappa^2 \partial_{xx}\mathscr{P} \\ \mathscr{P}(x,0) &= P_s(x) \end{aligned}$$

At measurement times t_m, the probability jumps

$$\mathscr{P}(x,t^+) = \frac{1}{N} e^{\frac{1}{R^2}[y_m - \frac{x_m^2}{2}]} \mathscr{P}(x,t^-)$$

Smoother: between t_{m+1} and t_m, solve

$$\begin{aligned} \partial_t \mathscr{A} &= -\partial_x [f(x)] \mathscr{A} + \frac{1}{2} \kappa^2 \partial_{xx} \mathscr{A} \\ \mathscr{A}(x, t_f) &= 1 \end{aligned}$$

At measurement times t_m, the probability jumps

$$\mathscr{A}(x, t^-) = \frac{1}{N} e^{\frac{1}{R^2} [y_m - \frac{x_m^2}{2}]} \mathscr{A}(x, t^+)$$

Finally $P(x, t | y_m, m = 1, 2, , M) = \mathscr{A}(x, t) \mathscr{P}(x, t)$.

KSP Filter and Smoother Results

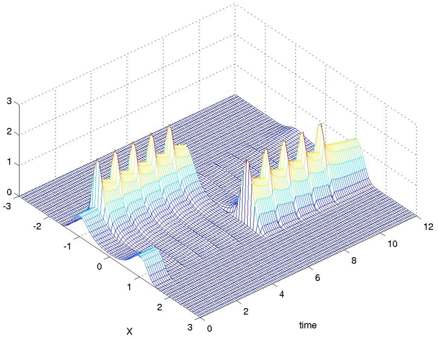


Figure: KSP Filter

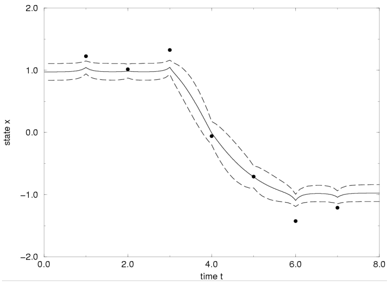


Figure: KSP Smoother

enKF Most Favored in Practice

The enKF (”state-of-the-art”)

- Use model for forecast.
- Update the uncertainty using Monte Carlo.

Pros and Cons:

- Can handle legacy code easily
- Linear (Gaussian) analysis
- Requires full model runs
- Ad-hoc

G. Evensen, Sequential data assimilation with a nonlinear quasigeostrophic model using Monte Carlo methods to forecast error statistics, J. Geophys. Res. **99**, 10143-10162.

PIMC The Path Integral Monte Carlo

- Optimal, on the discretized model
- Simple to implement, *but very subtle*
- Can handle legacy code
- Relies on sampling
- Can yield a variety of different estimators

J. Restrepo, A Path Integral Method for Data Assimilation, Physica D, 2007,
F. Alexander, G. Eyink, J. Restrepo, Accelerated Monte-Carlo for Optimal Estimation of Time Series, J. Stat. Phys., 2005

Bayesian Statement

- $P(x|y) \propto$ Likelihood $\times$ Prior.
- Use data for likelihood.
- Use model for prior.

$$P(x|y) \propto e^{-\mathscr{A}_{model}} e^{-\mathscr{A}_{data}} := e^{-\mathscr{A}(x)}.$$

$\mathscr{A}_{model}$

$$dx = f(x,t)dt + [2D(x,t)]^{1/2}dW$$

is discretized:

$$x_{n+1} = x_n + \Delta t f(x_n, t_n) + [2D(x_n, t_n)]^{1/2}[W_{n+1} - W_n]$$

$$n = 0, 1, ..., T-1$$

$$\mathscr{A}_{model} \approx \sum_{n=1}^{T} [(x_{n+1} - x_n - \Delta t f(x_n, t_n))^\top D(x_n, t_n)^{-1} (x_{n+1} - x_n - \Delta t f(x_n, t_n))],$$

if $\text{Prob}(\Delta W) \propto exp(-\Delta W^2/D)$.

$\mathscr{A}_{data}$

$$y_m = H(x_m) + [2R[x_m, t_m)]^{1/2} \eta_m$$
$$m = 1, 2, \ldots, M.$$

$$\mathscr{A}_{data} = \sum_{m=1}^{M} [(y_m - H(x_m))^\top R(x_n, t_n)^{-1} (y_m - H(x_m))],$$

if $\text{Prob}(\eta) \propto exp(-\eta^2/R)$.

MCMC Samplers

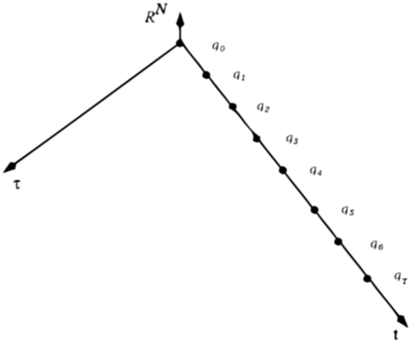
$$P(x|y) \propto e^{-\mathscr{A}_{model}} e^{-\mathscr{A}_{data}} := e^{-\mathscr{A}(x)}.$$

The Path Integral Monte Carlo practicality depends on fast sampling:

- Multigrid (UMC)
- Langevin Sampler (LS)
- Hybrid Monte Carlo (HMC)
- Shadow Hybrid MC (sHMC)
- Riemannian Manifold Hamiltonian Monte Carlo (RM-HMC)
- generalized Hybrid Monte Carlo (gHMC)

(HMC) Hybrid Markov Chain Monte Carlo

- Proposals generated by solving Hamiltonian system in fictitious time τ.
- Accept/reject via Metropolis Hastings



HMC Algorithm

Let $q_n(\tau = 0) = x_n$.

- To each q_n, a conjugate generalized momemtum, p_n, is assigned.
- The momenta p_n give rise to a kinetic contribution $K = \sum_{n=1}^{T} p_n^\top M^{-1} p_n/2$.
- The Hamiltonian of the system $\mathscr{H} = \mathscr{A}(q) + K(p)$. The dynamics are:

$$\begin{aligned} \frac{\partial q_n}{\partial \tau} &= M^{-1} p_n \\ \frac{\partial p_n}{\partial \tau} &= F_n \quad \text{where} \quad F_n = -\mathrm{grad}(\mathscr{A}(q)). \end{aligned}$$

- Solve using Verlet integrator (detailed balance).
- Accept/Reject Metropolis/Hastings.

Why does HMC work? What are good HMC properties?

Write probability $\Pi(q) = \frac{1}{Z_\Pi} e^{-\mathscr{A}(q)}$:

- Sampling $\pi(q,p) = \frac{1}{Z_\pi} e^{-\mathscr{H}(q,p)} \sim \frac{1}{Z} e^{-\mathscr{A}(q)}$ samples $\Pi(q)$.
- Gradient dynamics makes system search through configuration space more efficiently.
- Moves in q_n are linear in p_n, *i.e.*, $\frac{\partial q}{\partial \tau} = M^{-1} p$

$\mathscr{A}(q)$ and grad$(\mathscr{A}(q))$ should be easily evaluated.

Sampler Efficiency Estimates

Sampler Efficiency: key to choosing and tuning sampler

- Computational Cost: $\mathscr{O}(NT)^r n_{\text{method}}(p, L)$
- $p :=< P_{acc} >=< \min\{1, \exp[-\Delta\mathscr{H}]\} > \propto \operatorname{erfc}\left(\frac{1}{2}\delta\tau^m (NT)^{1/2}\right)$.
- $c(L) :=< \mathscr{H}(0)\mathscr{H}(0+L) >$. Depends on problem dimension and state space characteristics.

RM-HMC Algorithm[2]

Hamiltonian replaced by:

$$\mathscr{H} = \mathscr{A}(q) + \frac{1}{2}p^\top G(q)^{-1}p$$

where the *non-degenerate Fisher information matrix* $G := \mathbb{E}\{\nabla\mathscr{A}\nabla\mathscr{A}^\top\}$

Challenges:

- find a time-reversible/volume-preserving discrete integrator for Hamiltonian problem.
- optimize its computational efficiency.

[2] Girolami, Calderhead, Chin, preprint, 2009.

Decrease decorrelation length L: gHMC Algorithm

Hamiltonian dynamics replaced by:

$$\begin{aligned}\frac{\partial q_n}{\partial \tau} &= CM^{-1}p_n \\ \frac{\partial p_n}{\partial \tau} &= C^{\top}F(q_n)\end{aligned}$$

where $C \in \mathbb{R}^{T\times T}$ matrix

Challenge: find C that leads to a significant reduction in the sample decorrelation length.

We used the circulant matrix $C = \mathrm{circ}(1, e^{-\alpha}, e^{-2\alpha}, \ldots, e^{-T\alpha})$.

Sampler Efficiency Comparison

Table: *T is the number of time steps, $(\cdot)$ is the standard deviation on the number of samples, $[\alpha]$ used in C; J is the number of τ time steps.*

$T+1$	HMC (J=1)	HMC (J=8)	UMC	gHMC (J=1)
8	900(125)	170(7)	800(40)	40(8) [0.20]
16	5300(1600)	560(20)	1040(60)	60(10) [0.10]
32	13300(8300)	2700 (140)	1430(100)	200(30) [0.05]
64	30000(7800)	2800(400)	1570(100)	420(70) [0.0245]

The Diffusion Kernel Filter

Collaborator

Paul Krause, UQG U. Arizona

- A particle-filter based method:

$$p(x|y_m,t_m) = \frac{p(y_m|x)\,p(x,t_m)}{p(y_m)}.$$

- Sequential.
- Simple to implement, provided an adjoint exists.
- The *Clustered* cDKF is competitive with EKF.
- Can handle nonlinear/non-Gaussian problems.

P. Krause and J. Restrepo, The Diffusion Kernel Filter Applied to Lagrangian Data Assimilation, Mon. Wea. Rev. 2009,
P. Krause and J. Restrepo, Sequential Estimation with Deterministic Models and Noisy Data, SIAM J. Sci. Comp. 2009

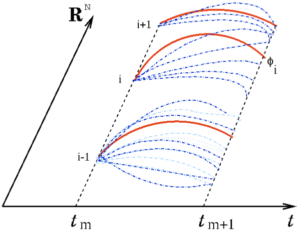
The Diffusion Kernel

$$\Phi(\cdot) = \phi(\cdot) + \Phi'(\cdot),$$

- $\Phi'(\cdot)$ is obtained by applying Duhamel's principle and expectation projections:

$$\Phi'(\xi(t_m, i), t) \sim \int_{t_m}^{t} G\, dw(s - t_m),$$

where



$$G(\xi(t_m, i), t, s) := \nabla\phi(\xi(t_m, i), t)\; g(\Phi(\xi(t_m, i), s))$$

is the *Diffusion Kernel*.

The Uncertainty Norm

The diffusion kernel is

$$G(\xi(t_m, i), t, s) := \nabla \phi(\xi(t_m, i), t)\, g(\Phi(\xi(t_m, i), s))$$

The uncertainty norm $||G||(t)$ for branches of prediction bounds the sup-norm of the covariance matrix of $\Phi'(\xi(t_m, i), t)$:

$$\frac{1}{N}||\mathrm{cov}(\Phi')||_\infty \leq ||G||^2, \quad t_m \leq t \leq t_{m+1}.$$

The key observation is that $||\mathrm{cov}(\Phi')||_\infty$, a measure of entropy of the branch, will be small whenever $||G||$ is small.

The Oceanic Lagrangian Setting

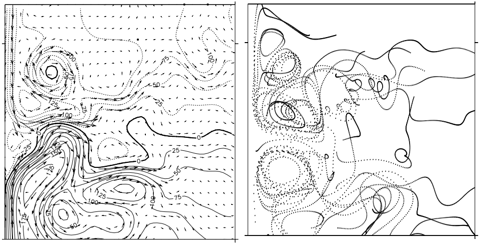


Figure: Synthetic Eulerian Flow. Lagrangian Tracks. From Molcard *et al* J. Atmos Ocean. Tech. (2005).

Figure: Four MLFII's Ready for Hurricane Isidore after Puget Sound Testing in July, 2002. From E. D'Asaro's (Washington) Web Page.

2D Vortex/Drifter Problem

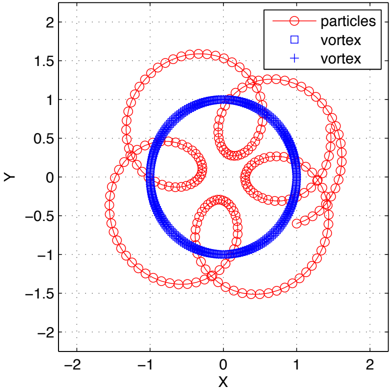
Measure the tracks of N_p passive tracers $\zeta_i(t) := v_m(t) + i w_m(t)$, give an optimal estimate of tracks and of N_v point vortices $z_m(t) := x_m(t) + i y_m(t)$.

Vortices:

$$\frac{dz_m}{dt} = \frac{i}{2\pi} \sum_{l=1, l\neq m}^{N_v} \frac{\Gamma_l}{z_m^* - z_l^*} + \eta^V(t),$$

Drifters:

$$\frac{d\zeta_n}{dt} = \frac{i}{2\pi} \sum_{l=1}^{N_p} \frac{\Gamma_l}{\zeta_n^* - z_l^*} + \eta^P(t).$$



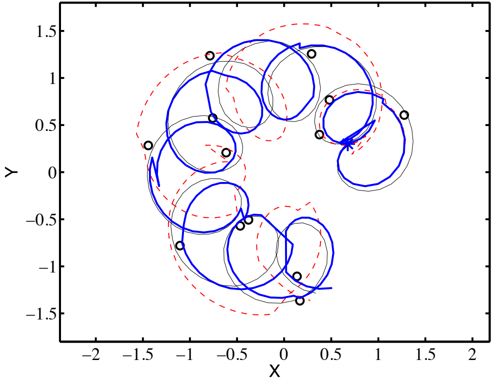
L. Kusnetsov and K. Ide and C.K.R.T. Jones, A method for assimilation of Lagrangian data, Mon. Wea. Rev **131** (2003)

EKF Lagrangian Estimation

Kusnetsov *et al* proposed using *Constrained EKF*:

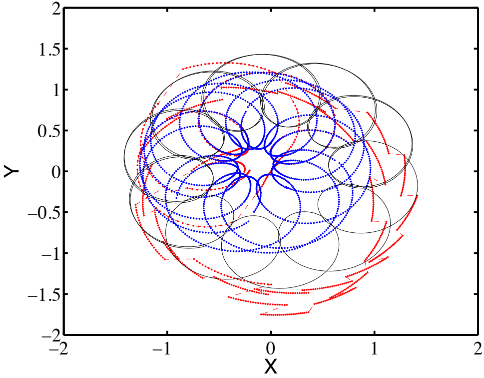
- for reasonable data uncertainties and/or frequent enough observations EKF works reasonably well, based on the computation of distance between the "true" path and the EKF estimate
- if pushed too hard the EKF would fail: saddles in the orbit paths would cause divergences in the forecast stage
- reasonably efficient, robust

Comparison of Benchmark, DKF and EKF



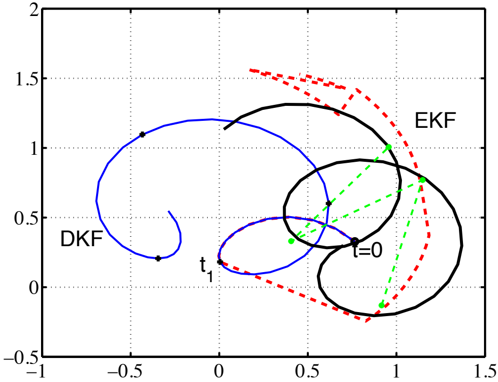
Mean drifter path (v_1, w_1), as estimated by DKF and benchmark and EKF (jagged, dashed). The true path is superimposed (black).

Comparison of Bootstrap, DKF and EKF



Mean drifter path (v_1, w_1), as estimated by DKF and particle filter and EKF (jagged, dashed). The true path is superimposed (black).

(Short-time) Behavior of Bootstrap, DKF and EKF



Mean drifter path (v_1, w_1), as estimated by DKF and particle filter, and EKF (red dashed). "Truth" path (black), Measurements (green, dashed).

Comparison of Benchmark, DKF and enKF

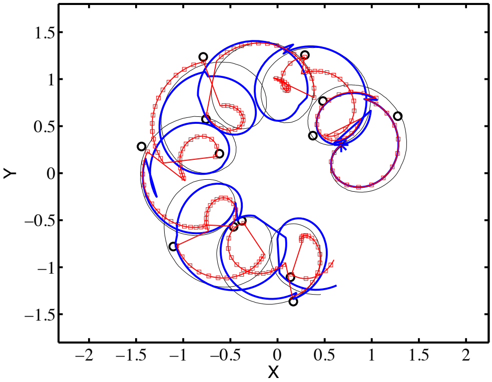
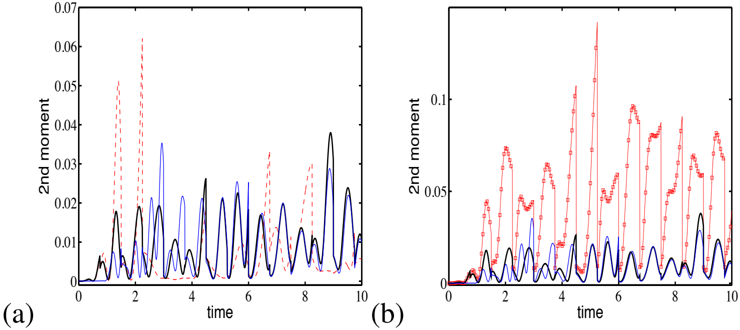


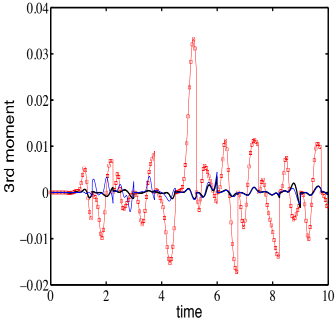
Figure: Mean drifter path (v_1, w_1), as estimated by DKF and benchmark and enKF (jagged, dashed). The true path is superimposed (black).

SECOND MOMENT: Comparison of Benchmark, DKF, enKF and EKF



Comparison of DKF, benchmark filter for $v_1(t)$. Second moment (a) EKF, (b) enKF.

THIRD MOMENT: Comparison of Benchmark, DKF and enKF



Comparison of enKF, DKF, and benchmark filter moment estimates for $v_1(t)$. Third moment. *EKF does not generate a third moment.*

Computational Efficiency, from t_k to t_{k+1}

Bootstrap Filter Cost

$$C\beta\alpha T \times N_x^2 \times I,$$

- I sample paths (5×10^5 in examples).
- N_x is the dimension of state variable.
- T is the number of deterministic time steps.
- αT is the number of times steps taken in the stochastic differential equation integrator, $\alpha \gg 1$, $\beta > 1$. (SODE time step in examples, 10^{-4}, $T = 10^{-2}$).

DKF Cost

$$T \times (C' N_x + C N_x^3) \times I_k \approx T \times C N_x^3 \times I_k.$$

Conditions whereby the cost of the bootstrap filter exceeds that of the DKF:

$$N_x < \alpha\beta I / I_k.$$

Reflects the impact of non-Gaussianity in the cost of the DKF: the less Gaussian, the closer I/I_k is to 1.

EKF vs. DKF Cost

- In the prediction step both methods are comparable.
- In the analysis step EKF is $\min\{N_x^3, N_y^3\}$, for DKF it is $I_k \times N_x^2$.

Computational Efficiency, from t_k to t_{k+1}

Bootstrap Filter Cost

$$C\beta\alpha T \times N_x^2 \times I,$$

- I sample paths (5×10^5 in examples).
- N_x is the dimension of state variable.
- T is the number of deterministic time steps.
- αT is the number of times steps taken in the stochastic differential equation integrator, $\alpha \gg 1$, $\beta > 1$. (SODE time step in examples, 10^{-4}, $T = 10^{-2}$).

DKF Cost

$$T \times (C' N_x + C N_x^3) \times I_k \approx T \times C N_x^3 \times I_k.$$

Conditions whereby the cost of the bootstrap filter exceeds that of the DKF:

$$N_x < \alpha\beta I / I_k.$$

Reflects the impact of non-Gaussianity in the cost of the DKF: the less Gaussian, the closer I/I_k is to 1.

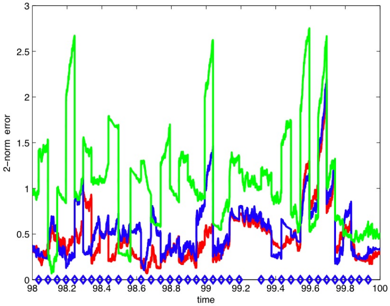
cDKF Cost

Can use $I_k \leq 10^{-r} I$, $r \geq 1$.

In examples, $r = 2$. Conditions whereby the cost of bootstrap exceeds that of rDKF:

$$N_x < \alpha\beta 10^r.$$

In nonlinear problems there is no unique best predictor:

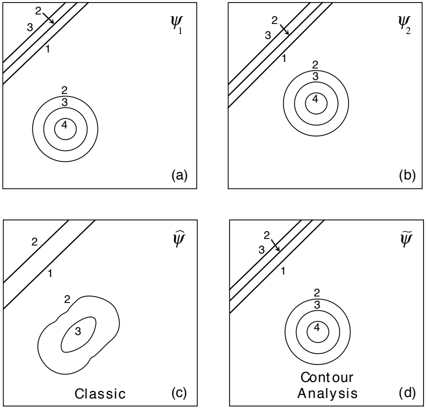


Shown: average, entropy average, likelihood average.

from P. Krause J. M. Restrepo, to be submitted, SIAM J. Sci. Comp, 2010.

Lagrangian Data Blending

Data Blending: Contour Dynamics



A competing approach: B. Beechler, J. Weiss, G. Duane, J. Tribbia, to appear in J. Atmos Sci 2010

The Meteorology vs. Climate Problem

the data assimilation problem will remain dimensionally-challenging:

- Develop statistically-convergent methods
- Increase the availability and quality of the data
- Reduction of the state space:
  - Exploit vastly different degrees of uncertainty.
  - Develop closures and stochastic parameterizations.

all the while, being clear about what it is that we are doing...

The Ensemble Bred Vectors

GOAL:

- Propose an alternative and robust methodology for sensitivity analysis and reduced-space representation
- Provide a mathematical basis for the Bred Vector (BV).

COLLABORATORS:
Nusret Balci (IMA/UMN), Anna Mazzucato (PSU), George Sell (UMN).

The Ensemble Bred Vectors

GOAL:

- Propose an alternative and robust methodology for sensitivity analysis and reduced-space representation
- Provide a mathematical basis for the Bred Vector (BV).

COLLABORATORS:

Nusret Balci (IMA/UMN), Anna Mazzucato (PSU), George Sell (UMN).

The Bred Vectors, Defined

Given a model

$$\begin{aligned} Y(t_n+\delta t_n) &= M(Y(t_n),t_n), \quad n=0,1,2,.., \\ Y(0) &= Y_0, \end{aligned}$$

$Y(t_n) \approx y(t=t_n)$, while $Y_0 \approx y_0$.

BVs computed as follows:

$$\begin{aligned} \delta Y_{n+1} &:= M(Y_n+\delta \mathscr{Y}_n,t_n) - M(Y_n,t_n), \\ \delta \mathscr{Y}_{n+1} &:= R\,\delta Y_{n+1}. \end{aligned}$$

$$R := \frac{\|\delta Y_0\|}{\|\delta Y_{n+1}\|}, \quad \textit{the normalization factor.}$$

See Toth and Kalnay 1993, Bull. Amer. Met. Soc, Toth and Kalnay Mon Wea Rev, 1997

The Lyapunov Vectors, Defined

Let $\delta\Theta$ be the solution of the initial value problem

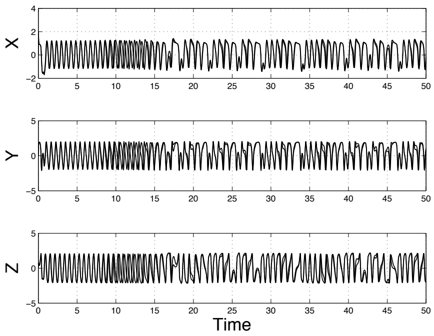
$$\begin{aligned} \delta\Theta(t_n+\delta t_n) &= L_n \delta\Theta(t_n), \quad n = 0,1,2,\ldots \\ \delta\Theta(0) &= \delta\Theta_0, \end{aligned}$$

Take $n \to \infty$.
At each time step the *Tangent Linear Model* is

$$L_n := L(Y_n, t_n) = \left. \frac{\partial G(y,t)}{\partial y} \right|_{y=Y_n, t=t_n}.$$

LV make sense in asymptotic time (the BV is local in time)

The Lorenz63

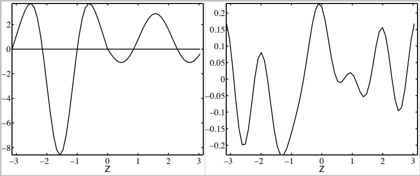


$$\frac{dx}{dt} = \sigma(y-x)$$
$$\frac{dy}{dt} = x(r-z)-y$$
$$\frac{dz}{dt} = xy-bz, \quad t>0. \tag{1}$$

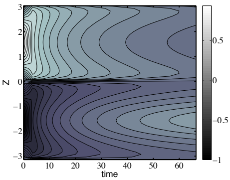
At $t=0$, $x=0.8001$, $y=0.4314$, $z=0.9106$. Here we set $r=28, b=8/3, \sigma=10$.

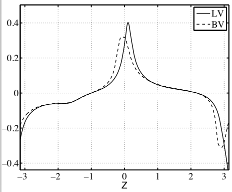
The Cahn Hilliard (Cessi-Young 92 Model)

$$
\begin{aligned}
\frac{\partial S}{\partial t} &= \alpha \frac{\partial^2}{\partial z^2}[f(z) + \mu S(S - \sin(z))^2 + S - \gamma \frac{\partial^2 S}{\partial z^2}], \quad t > 0, z \in [-\pi, \pi] \\
S(z,0) &= S_0(z)
\end{aligned}
$$



Forcing function and perturbation



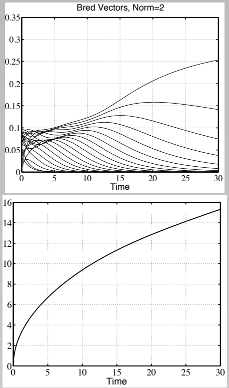


Solution of CY92, and LV vs. BV.

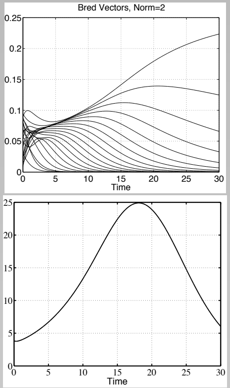
Numerical Instability of Bred Vectors

$$\frac{dx}{dt} = Ax$$

where A is non-normal constant (Jordan) matrix of dim(20):



standard BV computation



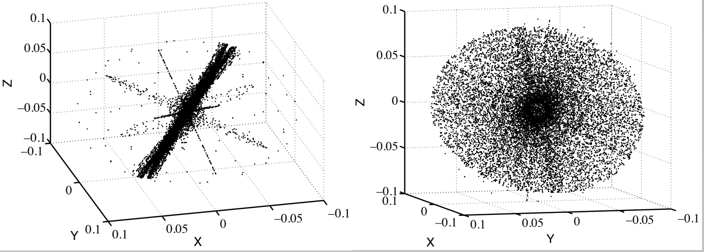
Well-conditioned computation

The Ensemble Bred Vector (EBV), Defined

The Ensemble Bred Vector

- Pick a norm and its size, for an ensemble of initial conditions.
- Advance the ensemble of initial conditions using the BV
- find the element from the ensemble that grew the most and use it to resize the whole ensemble
- repeat

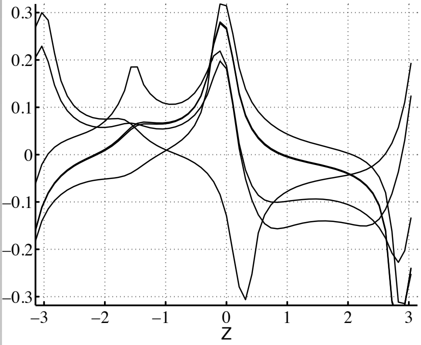
The Lorenz63 using EBV



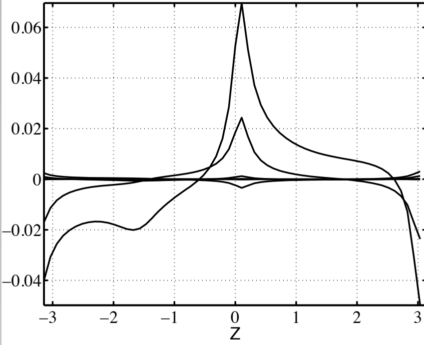
Superposition of an ensemble of 1500 runs, over all times from 0 to 50. (We show 150 of these), normalized to the initial condition 2-norm, set to 0.1.

The Cahn Hilliard CY92, using EBV

BV's



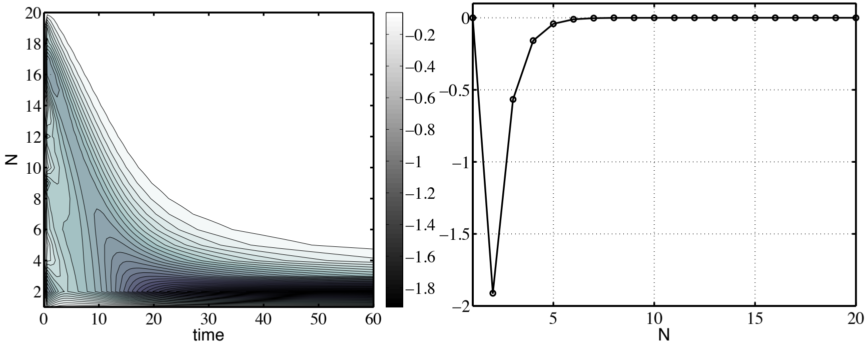
EBV's



The Non-Normal Linear Problem, using EBV

$$\frac{dx}{dt} = Ax,$$

where A is non-normal and dim(20):



Bred Vectors, Some of our Conclusions

- BV are attractive because they can be applied to legacy code
- Established that BV is a nonlinear generalization of the LV, though it is not clear how a finite-time quantity is to be compared to an asymptotically-defined quantity, unless we are discussing a steady state.
- Established that BV is norm dependent
- In applications the norm matters a great deal when considering what is to be analyzed.
- But do they provide any practical information not contained in singular vectors or LV?

Dimension Reduction via Stochastic Parameterization

$$\begin{aligned} dx &= f(x)dt + \text{sochastic term} \\ x(0) &= x_0 \end{aligned}$$

Stochastic Parameterization

GOAL:

- Produce stochastically-based subscale parameterization
- Make greater contact with data (and data assimilation).

COLLABORATORS:
Darin Comeau (U Arizona),
Brad Weir (U Arizona), Jorge Ramírez (U. Nacional de Colombia), J. C. McWilliams (UCLA), M. Banner (UNSW)

Stochastic Parameterization

GOAL:

- Produce stochastically-based subscale parameterization
- Make greater contact with data (and data assimilation).

COLLABORATORS:

Darin Comeau (U Arizona),
Brad Weir (U Arizona), Jorge Ramírez (U. Nacional de Colombia), J. C. McWilliams (UCLA), M. Banner (UNSW)

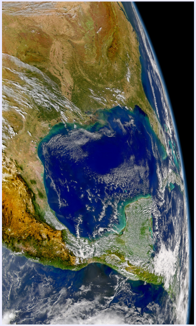
WAVE BREAKING DISSIPATION:

Geophysical Goals

- How does dissipation at wave scales manifest itself at longer time scales?
- What is the correct stochastic parameterization?
- Can we parameterize dissipation in such a way that we use field data more efficiently?

Scale Range of the Model

- 10 secs-months
- 100m-basin scale
- Speed: waves > currents
- $kH \sim 1$
- Applications:
  - climate dynamics (transport)
  - erodible bed dynamics
  - river plume evolution
  - algal/plankton blooms
  - Oil spills/pollution



J. McWilliams and J. M. Restrepo *The Wave-driven OceanCirculation* J. Phys. Oceanogr. (1999)
J. Restrepo, *Wave-Current Interactions and Shore-connected Bars* J. Estuarine Sci. (2001)
J. McWilliams J. M. Restrepo, E. Lane *An asymptotic Theory for the Interaction of Waves and Currents in Coastal Waters* J. Fluid Mechanics (2004)
E. Lane, J. M. Restrepo, J. McWilliams *Wave-Current Interaction: A comparison of radiation-stress and vortex-force representations* J. Phys Oceanogr (2007)

Lagrangian Motion

$$d\mathbf{X} = \mathbf{V}dt$$

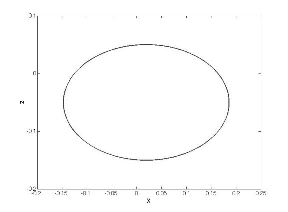


Figure: Deterministic

Lagrangian Motion Under White Capping

$$d\mathbf{X}_t = \mathbf{V}(\mathbf{X},t)dt + d\mathbf{W}_t(\mathbf{X},t)$$

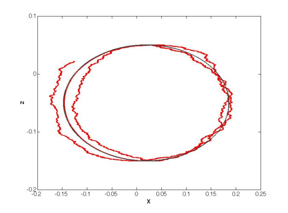


Figure: Stochastic

What Stochastic Model?

$$d\mathbf{X}_t = \mathbf{V}(\mathbf{X},t)dt + d\mathbf{W}_t(\mathbf{X},t)$$

Experiments are needed to determine the right model

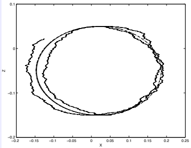


Figure: dW standard white noise

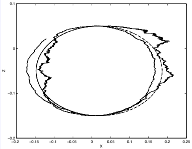


Figure: $d\mathbf{W}$ with added mean-reverting process

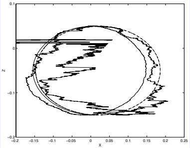
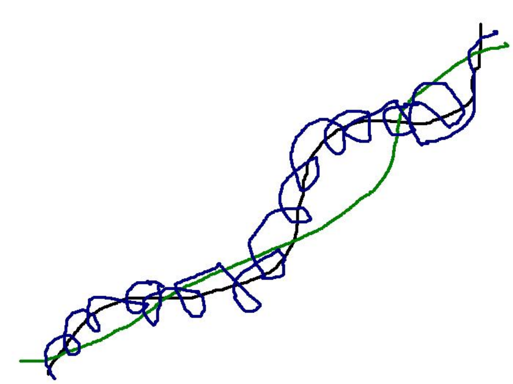


Figure: $d\mathbf{W}$ with added jump process

Lagrangian/Eulerian Projections in Multiscale Setting

BASIC STRATEGY

Multiscale projection of systems of (mostly) hyperbolic equations between the Eulerian and Lagrangian frames.



Current/Waves/Breaking Model

The momentum with breaking-generated stresses and diffusion:

$$\frac{\partial \mathbf{v^c}}{\partial T} = \mathbf{V} \times \mathbf{Z} - \nabla \Phi + \langle (\mathbf{b} \times \mathbf{Z}) + \mathbf{b} \times (\nabla \times \mathbf{b}) + [\mathbf{V} \times \nabla \times \mathbf{b}] - \frac{1}{2} \nabla |\mathbf{b}|^2 \rangle + \nabla \cdot \mathbf{R}$$

where $\mathbf{V} = \mathbf{v^c} + \mathbf{u}^{stokes}$, $\mathbf{Z} = \nabla \times \mathbf{v} + 2\Omega$ and $\Phi = p_0 + \frac{1}{2}|\mathbf{V}|^2$.

The $\mathbf{u}^{stokes}$ is the wave contribution.

The break-stresses modify the vortex force and Bernoulli head.

The diffusion accounts for mixed-layer boundary-layer effects

The tracers obey

$$\frac{\partial C}{\partial T} + \mathbf{V} \cdot \nabla C = -\mathbf{b} \cdot \nabla C + \nabla \cdot \mathbf{Q}.$$

J. M. Restrepo, *Wave Breaking Dissipation in the Wave-Driven Ocean Circulation* J. Phys. Oceanogr. (2007)
J.M. Restrepo, J. M. Ramírez, J.C. McWilliams & M. Banner *Wave Breaking Dissipation and Diffusion in Waves and Currents* , J. Phys. Oceanogr 2010

Wave Breaking Diffusion

The observation is that wave breaking increases the size of the mixing layer and this layer will then create a great deal of dissipation.

$$\mathbf{R}_v \approx \nu \frac{\partial \mathbf{v}_h}{\partial z}, \quad \mathbf{R}_h \approx \nu \nabla \mathbf{v}_h$$
$$Q_v \approx \kappa \frac{\partial C}{\partial z}, \quad \mathbf{Q}_h \approx \kappa \nabla C.$$

We assume that

$$\nu \sim \langle \ell_b \left| w^b \right| \rangle, \qquad \kappa \sim \langle \ell_\theta \left| w^b \right| \rangle.$$

w^b is the vertical component of the velocity associated with breaking, and the mixing length is

$$\ell_b = \gamma \eta(\mathbf{x}, t), \qquad \ell_\theta = \alpha \eta(\mathbf{x}, t).$$

Boundary Conditions

The surface boundary conditions at $z = \eta(\mathbf{x}_h, t)$ are the following:

$$w = \frac{D\eta}{Dt}, \quad p = g\rho_0\eta + p_a, \quad \nu\frac{\partial \mathbf{q}}{\partial z} = \frac{1}{\rho_0}\tau, \quad \kappa\frac{\partial C}{\partial z} = \mathscr{T}.$$

Lead to (at $z = 0$)

$$w^c = \nabla \cdot \mathbf{M} - w^b, \qquad p^c = \eta^c + p_{a0} - P$$

$$\nu\left(\frac{\partial \mathbf{v^c}}{\partial z} + \mathbf{S}\right) = \tau - \nu\frac{\partial \mathbf{b^c}}{\partial z}, \qquad \kappa\frac{\partial C}{\partial z} = \mathscr{T}.$$

where the *wave-induced adjustments* (at $z = 0$) are

$$\mathbf{M} \equiv \left\langle \mathbf{u}^w \eta^w \right\rangle, \quad P \equiv \left\langle p_z^w \eta^w \right\rangle, \quad \mathbf{S} \equiv \left\langle \frac{\partial^2 \mathbf{u}^w}{\partial z^2} \eta^w \right\rangle.$$

How to determine $\mathbf{b}(\mathbf{X}, T)$?

Thus far we have answered the question:

- *if breaking occurs, how do waves and currents get modified, at these large spatio-temporal scales?*

How to determine $\mathbf{b}(\mathbf{X}, T)$?

- *How is the breaking velocity* **b** *determined?*

it can be determined by field data...or

How to determine $\mathbf{b}(\mathbf{X}, T)$?

- *How is the breaking velocity* **b** *determined?*

  it can be determined by field data...or

Steps to determine $\mathbf{b}(\mathbf{X}, T)$:

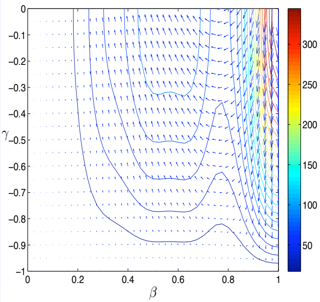
Ingredients in determining **b**, using averaging and breaking kinematics:

- individual breaker $\tilde{\mathbf{b}}$ velocity are obtained parametrically from data.
- breaking events need to be *upscaled*, to yield **b**.
- kinematics of their occurrence can be tied to wave-group dynamics.
- breaking events should be energetically consistent.
- breaking events should *yield* a Poisson process in space-time.

Details in: J.M. Restrepo, J. M. Ramírez, J.C. McWilliams & M. Banner, *Wave Breaking Dissipation and Diffusion in Waves and Currents* , J. Phys. Oceanogr 2010.
B. Weir & J. M. Restrepo, Stability of the Langmuir Circulation in the Presence of Wave Breaking, *in preparation.*

Individual breaker velocity $\tilde{\mathbf{b}}$ are obtained parametrically:

Individual breaker velocity $\tilde{\mathbf{b}}$ given by:



$$\partial_t \tilde{\mathbf{b}} + \tilde{\mathbf{b}} \cdot \nabla \tilde{\mathbf{b}} = \frac{1}{Re} \Delta \tilde{\mathbf{b}} + \mathbf{A},$$

$$\nabla \cdot \tilde{\mathbf{b}} = 0.$$

where

$$\mathbf{A} = k_b \mathscr{X}(x) \mathscr{Y}(y) \mathscr{Z}(z) \mathscr{T}(t) \mathbf{a}$$

cf., Sullivan, McWilliams, Melville, JFM, **507** (2004).

Breaking events need to be *upscaled* to yield **b**

Model $\mathbf{b} := \mathbf{B} + \mathbf{b}'$ as the random sum

$$\mathbf{b}(\mathbf{x}_h, z, T) = \sum_{(\mathbf{X}_h,\tau)\in\Phi} \mathbf{b}_{E(\mathbf{X}_h,\tau)}(\mathbf{x}_h - \mathbf{X}_h, z)\,\delta(\tau - T)$$

where

$$\mathbf{b}_E(\mathbf{x}_h - \mathbf{X}_h, z) := \frac{1}{\tau_E}\int_0^{\tau_E} \tilde{\mathbf{b}}(\mathbf{x}_h - \mathbf{X}_h, z, t)dt$$

The ensemble average at $(\mathbf{x}_h, z, T)$ of some functional $\mathscr{F}$ of the field $\mathbf{b}$ is:

$$\begin{aligned}
\langle \mathscr{F}(\mathbf{b})\rangle (\mathbf{x}_h, z, T)dT \quad &:= \quad \left\langle \sum_{(\mathbf{X}_h,\tau)\in\Phi} \mathscr{F}(\mathbf{b}_{E(\mathbf{X}_h,\tau)}(\mathbf{x}_h - \mathbf{X}_h, z))\delta(T-\tau)\right\rangle \\
&= \quad \int_{\mathbb{R}}\int_{\mathbf{x}_h - \Omega_E} \mathscr{F}(\mathbf{b}_E(\mathbf{x}_h - \mathbf{X}_h, z))\Lambda(\,d\mathbf{X}_h, dT)p(E)dE.
\end{aligned}$$

$$\mathbf{B}(z) = \int_{\mathbb{R}}\left[\int_{\tilde{\Omega}_E} \mathbf{b}_E(\mathbf{X}_h, z, T)d\mathbf{X}_h dT\right]\lambda\, p(E)dE$$

Kinematics of their occurrence can be tied to wave-group dynamics

Waves expressed in terms of a carrier and an envelope:

$$\eta(\mathbf{X}_h, t) = \mathrm{Re}\left\{ e^{i(\bar{\mathbf{k}}_h \cdot \mathbf{x}_h + \bar{\sigma} t)} \rho(\mathbf{x}_h, t) e^{i\theta(\mathbf{x}_h, t)} \right\}.$$

where

$$\mathbb{P}(\rho(\mathbf{x}_h, t) \in d\rho) = \frac{\rho}{2\pi\langle\eta^2\rangle} \exp\left\{ -\frac{\rho^2}{2\langle\eta^2\rangle} \right\}$$

a Rayleigh distribution.

Mean growth rate of wave group energy:

If $\delta := \frac{1}{\langle\sigma\rangle}\frac{D\mu}{Dt} > \delta^* \approx 1.4 \times 10^{-3}$ a breaking wave group.

where the material derivative is taken following the wave group.

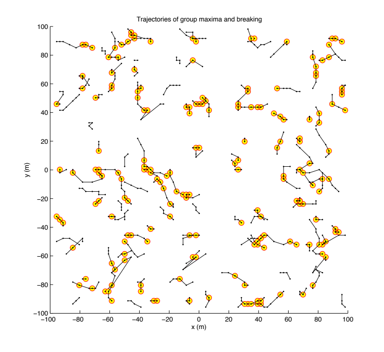
*Local wave energy** $\mu(t) := \eta^2(\mathbf{x}_{\max}, t)\, k^2(\mathbf{x}_{\max}, t)$.

$\mathbf{x}_{\max}(t)$	is maximum of crest point in the group.
$(k, \langle\sigma\rangle)$	wavenumber, mean frequency.

* see Song and Banner, **J Phys Oceanogr**, 2002

2D Example

(Loading breakmovie)



Space-time evolution of breaking events shown in movie.

Breaking events should be energetically consistent:

Recall:

$$\begin{aligned} \partial_t \tilde{\mathbf{b}} + \tilde{\mathbf{b}} \cdot \nabla \tilde{\mathbf{b}} &= \frac{1}{Re} \Delta \tilde{\mathbf{b}} + k_b \mathscr{X}(x) \mathscr{Y}(y) \mathscr{Z}(z) \mathscr{T}(t) \mathbf{a}, \\ \nabla \cdot \tilde{\mathbf{b}} &= 0. \end{aligned}$$

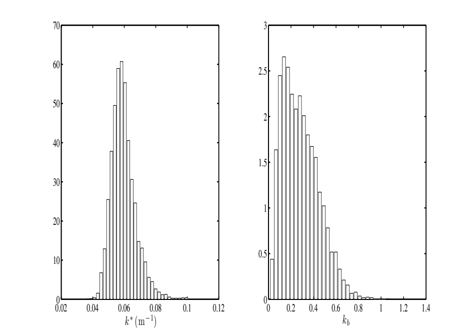
k_b found by equating the energy $E(T_w)$ from the break velocity field with the total energy change of the surface $\rho_0 g \Delta_{\eta^2}$.

$$E(T_w) := \frac{\rho_0 c}{T_w \Omega^b} \int_{\Omega^b[0,T_w]} \mathbf{A} \, d\mathbf{x} ds = \frac{0.55 g \rho_0 k_b}{k^2}.$$

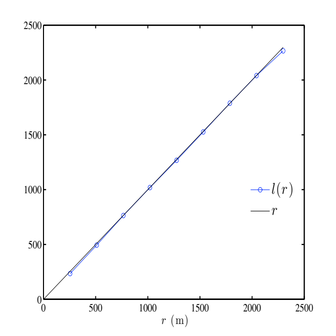
If Δ_{η^2} is total change in $\eta^2(\mathbf{x}_{\max}, \cdot)$ during the breaking event, then Thus

$$k_b = 1.82 k^2 \Delta_{\eta^2}.$$

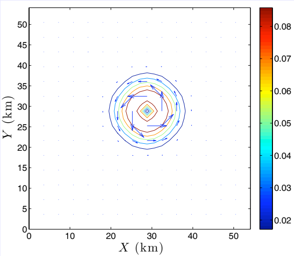
Breaking events should be Poisson process



Histograms of wave group mean wavenumber k^* and breaking strength k_b

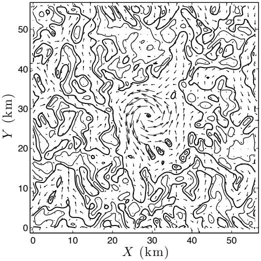


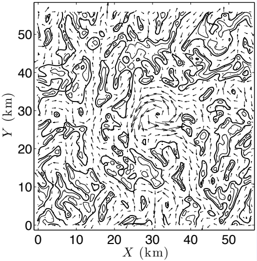
Poisson fit: theoretical indicator function vs. computational $l(r)$.



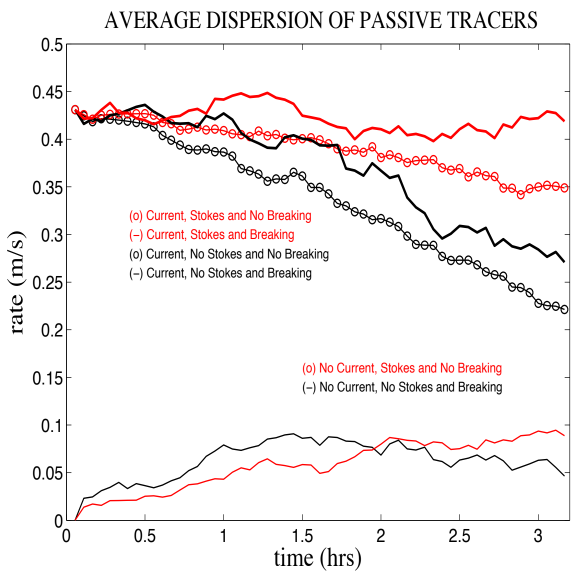
Initial Current, with top speed 0.1 m/sec.

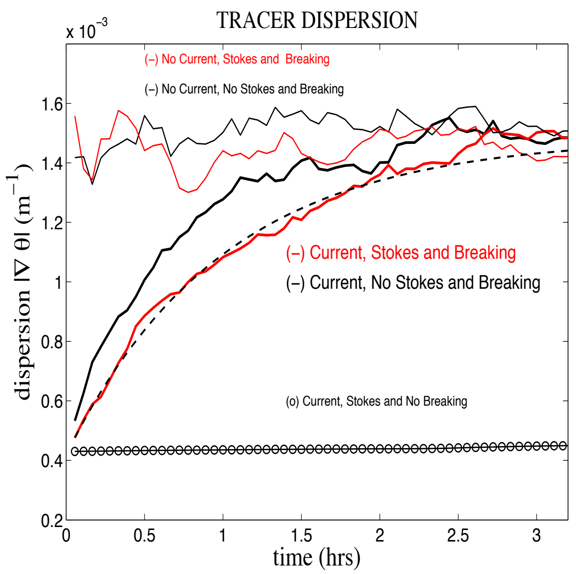
Current, after 6.4 hours (0.1 m/s contours), with waves and breaking due to 15 m/s winds:

(a) 

(b) 

(a) No Stokes drift (b) With Stokes $\mathbf{u}^{stokes} = (0.26, 0)$ m/s.


AVERAGE DISPERSION OF PASSIVE TRACERS
rate (m/s)
time (hrs)
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
0
0.5
1
1.5
2
2.5
3
(o) Current, Stokes and No Breaking
(–) Current, Stokes and Breaking
(o) Current, No Stokes and No Breaking
(–) Current, No Stokes and Breaking
(o) No Current, Stokes and No Breaking
(–) No Current, No Stokes and Breaking


TRACER DISPERSION
x 10^-3
(–) No Current, Stokes and Breaking
(–) No Current, No Stokes and Breaking
(–) Current, Stokes and Breaking
(–) Current, No Stokes and Breaking
(o) Current, Stokes and No Breaking
dispersion |∇ θ| (m^-1)
time (hrs)
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
0.5
1
1.5
2
2.5
3

Closing Comments

- Data assimilation with sparse data forces us to put emphasis on climate model development.
- The ”dimensional-curse” is unavoidable. In the long term, convergent methods allow one to know how much a computational gain is possible at the expense of optimality
- Closure methods and stochastic parameterizations can lead to robust and efficient models and hence dimension reduction...it exploits deep knowledge of the system dynamics.
- Taking advantage of different degrees of uncertainty among the degrees of freedom can be a productive alternative to reduced representations of background error fields

Further Information

Juan M. Restrepo

http://www.physics.arizona.edu/∼restrepo

Uncertainty Quantification Group

http://www.physics.arizona.edu/∼tolwinski