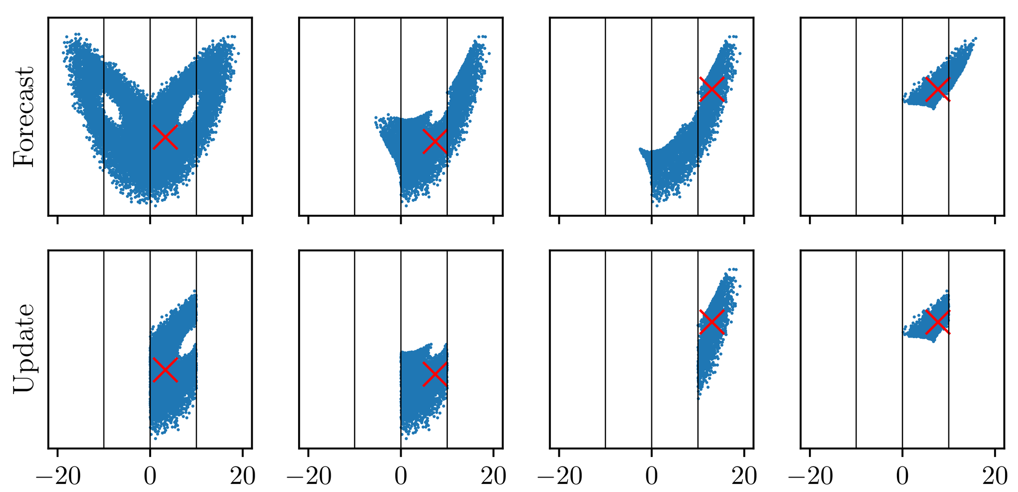
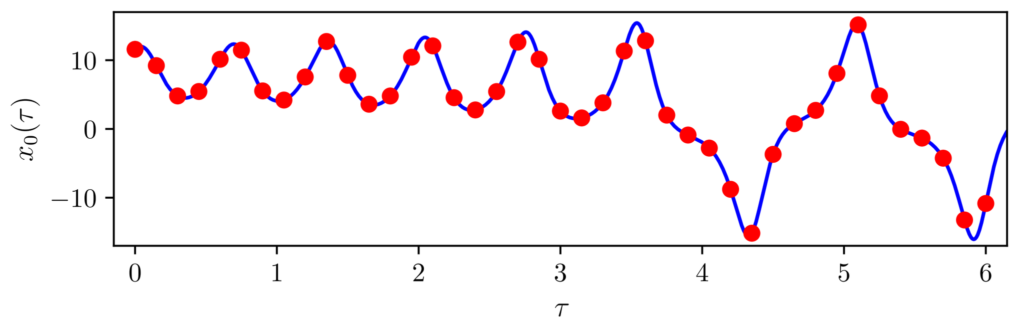
Entropy, Cross Entropy and Data Assimilation

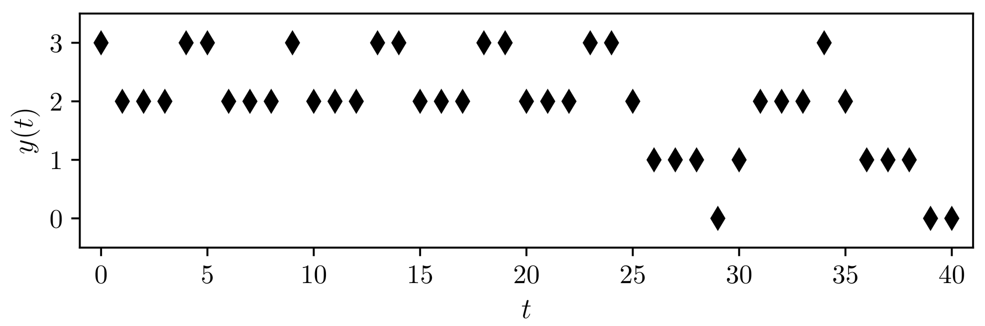
Andrew M. Fraser andy@FraserPhysics.com

ODE → Observations

$$\dot{x} = F(x)$$
$$x[0:10000] = \text{integrate}(F, n_t = 10000, \Delta_\tau = 0.15)$$
$$\text{bins} = [-10, 0, 10]$$
$$y = G(x)$$
$$y[0:10000] = \text{digitize}(x[0:10000, 0], \text{bins})$$



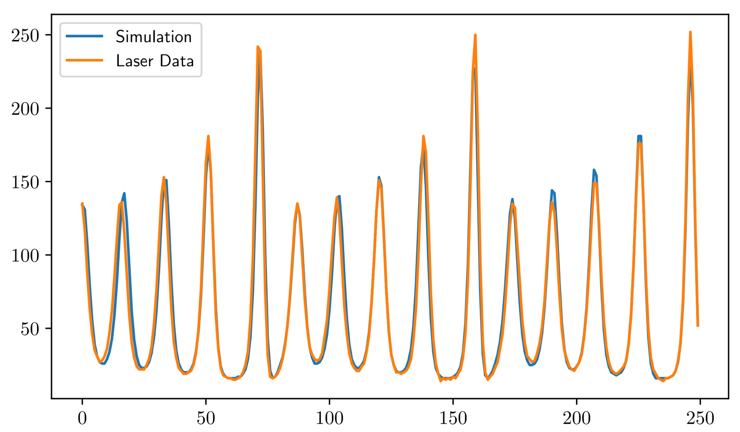


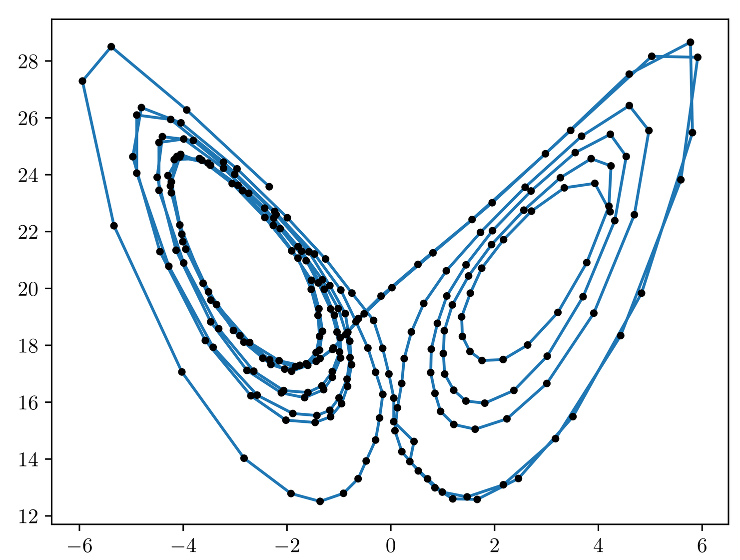


Data Assimilation: Observations → States ∈ $\mathbb{R}^3$

Laser data from Tang and Weiss
Extended Kalman smoothing for state space trajectory estimate

$$\hat{x} = \text{argmax}\, P(x \mid y)$$





HMMs & Data Assimilation: Observations → States ∈ $\mathbb{Z}$

Parameters, θ, of Hidden Markov Model (HMM) with states $s \in \mathbb{Z}$ and observations $y \in \mathbb{Z}$:

$P_{S \leftarrow S}$: State to state transition probabilities
$P_{Y \leftarrow S}$: Conditional observation probabilities

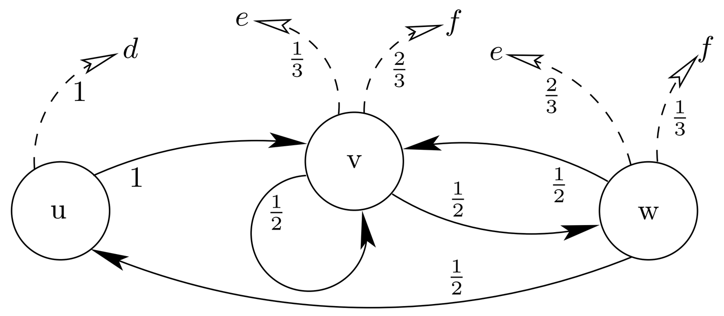
Estimation algorithms:

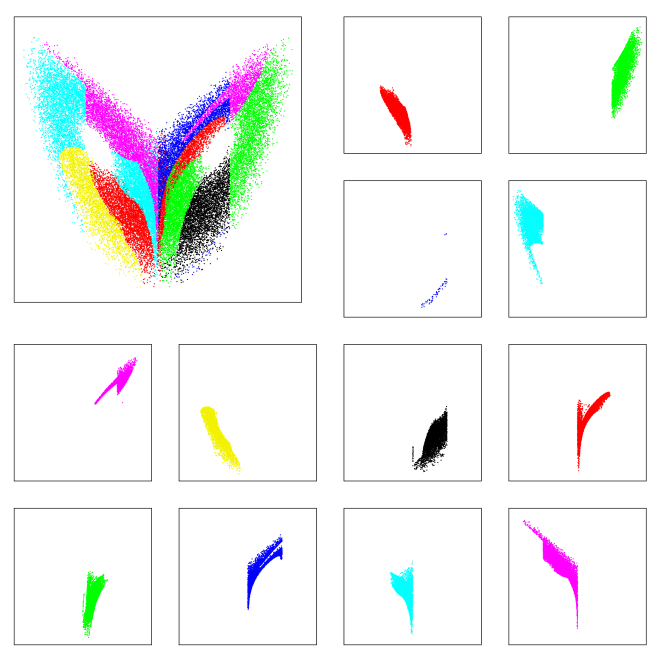
Forward Filter: Conditional probability of states $P(x[t] \mid y[0:t+1], \theta)$
MLE Parameters: (Forward-Backward, also known as Baum-Welch)

$$\hat{\theta} = \underset{\theta}{\text{argmax}}\, P(y[0:1000] \mid \theta)$$

MAP States: (Viterbi)

$$\hat{s}[0:1000] = \underset{s[0:1000]}{\text{argmax}}\, P(s[0:1000] \mid y[1000], \theta)$$





Variations on a Theme

Forward data assimilation alternates between **update** & **forecast**.

Update:
$$\alpha(x,t) \equiv P(X[t] = x \mid y[0:t+1])$$
$$\alpha(x,t) \propto \alpha(x,t) P_{Y \leftarrow X}(y[t] \mid x)$$

One must evaluate $\int \alpha(x,t)dx$ to normalize the update.

Forecast:
$$a(x,t) \equiv P(X[t] = x \mid y[0:t])$$
$$a(x,t) = \int \alpha(\chi, t-1) P_{X \leftarrow X}(x \mid \chi) d\chi$$

Kalman Filter $P_{X \leftarrow X}$ and $P_{Y \leftarrow X}$ are linear with Gaussian residuals.
Extended Kalman Filter Kalman filter for nonlinear functions with local linear approximations.
Hidden Markov Model State and observation spaces are finite sets.
Particle Filter Monte-Carlo for integrals. Probabilities represented by clouds of points.

Entropy and Lyapunov Exponents

Entropy for true model P_μ

$$h(\mu) \equiv \lim_{n \to \infty} -\frac{1}{n} \mathbb{E}_\mu \left[\log(P_\mu(y[0:n]))\right]$$

For $\forall y[0:n] \in A_\epsilon^{(n)}$, the typical or plausible set

$$\frac{-\log(P_\mu(y[0:n]))}{n} = h \pm \epsilon \qquad \text{definition } h \text{ is the rate that prob} \to 0.$$
$$\Pr\left\{A_\epsilon^{(n)}\right\} > 1 - \epsilon$$
$$\left|A_\epsilon^{(n)}\right| \leq e^{n(h+\epsilon)} \qquad h \text{ is the rate that } A_\epsilon^{(n)} \text{ grows.}$$

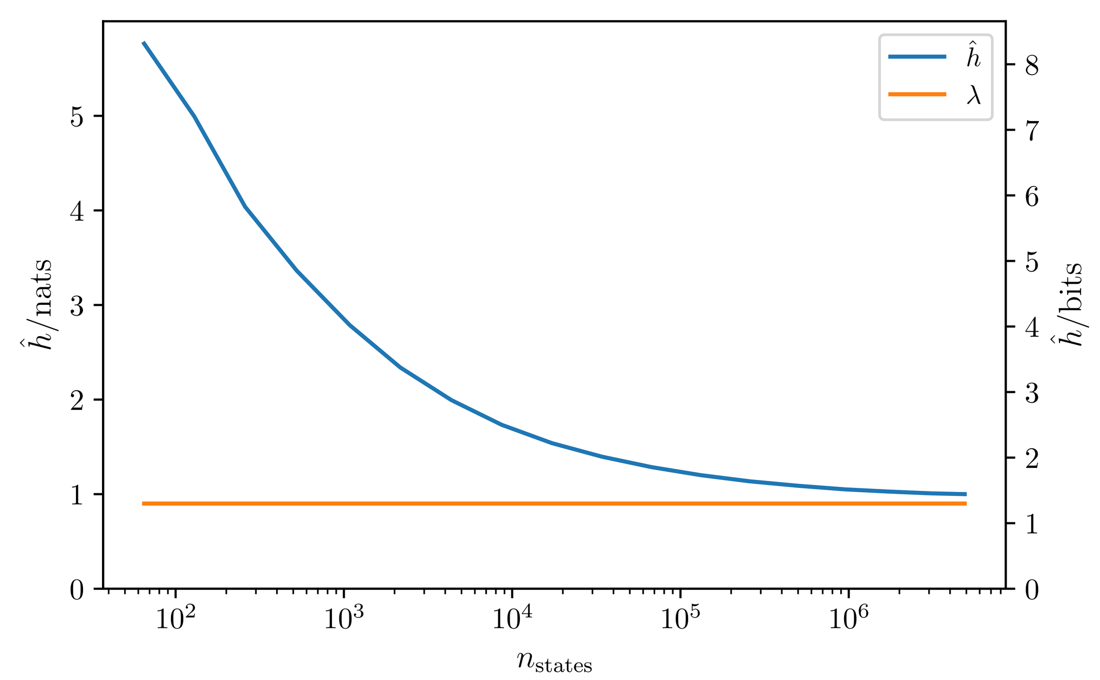
Cross Entropy of other model θ wrt true μ

$$h(\mu||\theta) \equiv \lim_{n \to \infty} -\frac{1}{n} \mathbb{E}_\mu \left[\log(P_\theta(y[0:n]))\right]$$
$$h(\mu||\theta) - h(\mu) \geq 0 \quad \text{equality} \rightarrow \mu = \theta \text{ almost everywhere}$$

Lyapunov exponents, λ_i characterize the exponential rates that trajectories converge or diverge. Estimate them numerically with Benettin's procedure that requires integrating tangent equation. Work of Ruelle, Pesin, Ledrappier, Young says that for the Lorenz system the largest exponent is equal to the entropy, ie,

$$h = \lambda_0 \approx 0.906.$$

So **0.906** is a lower **bound** for the cross entropy of a model of time series from the Lorenz system.



Use HMMs with many states to approach the bound.

Extended Kalman Filter

Given a model for states x and observations y in which

$$x[t+1] = f(x[t]) + \eta[t]$$
$$y[t] = g(x[t]) + \epsilon[t]$$

where f and g are differentiable but perhaps nonlinear and η and ϵ are iid Gaussian noise, *extended Kalman filtering* (EKF) is the practice of using Gaussians to model conditional distributions of states and observations. One propagates the means with the functions f and g and uses the derivatives of those functions to calculate covariances.

For Figures 1–3 I added draws from Gaussians with scales σ_η and σ_ϵ to the states and observations respectively of Lorenz simulations and applied EKFs.

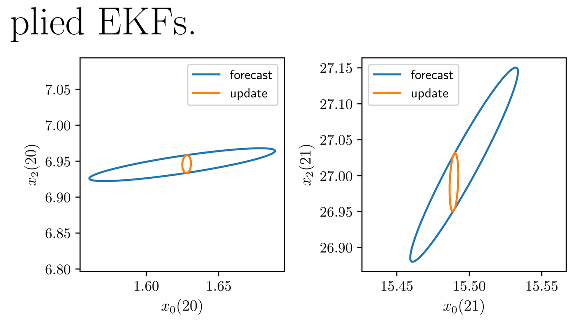


Figure 1: Level sets of conditional Gaussians illustrate forecast and update distributions.

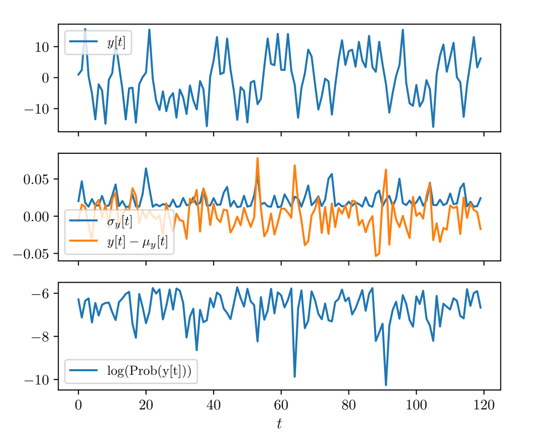


Figure 2: Time series of observations and characterizations of the forecast errors.

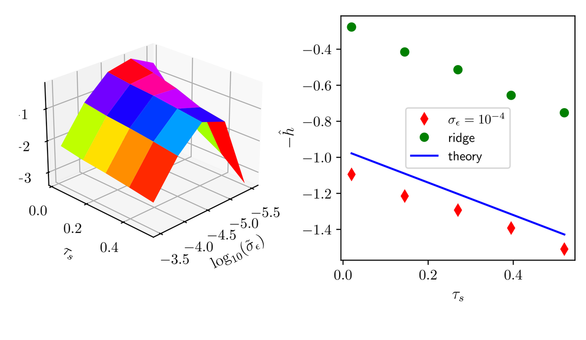


Figure 3: Dependence of cross entropy on state noise, σ_η, observation noise, σ_ϵ, and the time interval between samples, τ_s. While the slopes on the right match the true entropy, the intercepts are negative because the models are not optimal.

Particle Filter

To find something better than an HMM with zillions of states:

- Cover attractor with boxes, ie, particles
- Assign a uniform probability density in each box
- Use numerical ODE integration of Lorenz system and its tangent to move boxes forward in time
- When boxes get too big subdivide them
- When boxes overlap and get too numerous, random resample to decimate