Forecasting and Updating Traffic Flow.

Thomas Bengtsson

Department of Statistics

University of California - Berkeley

NCAR

June 16, 2005.

This work is supported by the National Science Foundation

Outline

1. Sequential DA using ensemble Kalman filter.

2. ensemble Kalman filter $\longrightarrow$ “full” Bayes DA.

3. Traffic example: DA in non-linear & non-Gaussian system.

4. DA in high-dimensional systems: what is reasonable?

Some to live by...

1. "Better to have the approximate solution to the correct problem than the exact solution to the wrong problem" -*J. Tukey*

Some to live by...

1. "Better to have the approximate solution to the correct problem, than the exact solution to the wrong problem" -*J. Tukey.*

2. "'Flying is like milk, everybody needs it' -*D.Nychka*.

Some to live by...

1. “Better to have the approximate solution to the correct problem than the exact solution to the wrong problem” -*J. Tukey*

2. “’Flying is like milk, everybody needs it’ -*D.Nychka*.

3. “It is easier to solve a problem if you know a lot about it” -*G.W. Bush*.

Where are we?

1. Sequential DA using ensemble Kalman filter.

2. ensemble Kalman filter $\longrightarrow$ "full" Bayes DA.

3. Traffic example: DA in non-linear & non-Gaussian system.

4. DA in high-dimensional systems: what is reasonable?

Approximation to reality:

Weather observations $\longrightarrow$ $\mathbf{y}_t = H(\mathbf{x}_t) + \boldsymbol{\epsilon}_t$

Atmospheric State $\longrightarrow$ $\mathbf{x}_t = G(\mathbf{x}_{t-1}) + \boldsymbol{\eta}_t$

$\mathbf{y}_t$, data

$\mathbf{x}_t$, unobserved

H maps state to observation (linear or non-linear)

G highly nonlinear (chaotic, approximate, known)

$\boldsymbol{\eta}_t$ (parameterized) model error, stochastic forcing

$\boldsymbol{\epsilon}_t$ (gaussian) observation error, $cov(\boldsymbol{\epsilon}_t) = \mathbf{R}$

Goal: Real-time sequential assimilation and forecasting:

$$p(\mathbf{x}_t|\mathbf{Y}^{t-1}), \mathbf{y}_t \xrightarrow{\text{Bayes}} p(\mathbf{x}_t|\mathbf{Y}^t) \xrightarrow{G(\cdot)} p(\mathbf{x}_{t+1}|\mathbf{Y}^t), \mathbf{y}_{t+1} \xrightarrow{\text{Bayes}} p(\mathbf{x}_{t+1}|\mathbf{Y}^{t+1})$$

Ensemble Kf algorithm

- Let $\mathbf{x}_{t,i}^f \sim p(\mathbf{x}_t|\mathbf{Y}^{t-1})$ $(i = 1, \ldots, M)$ be a sample from the prior.

  - EnKF: With $\hat{\mathbf{P}}_t^f$ the sample covariance of $\{\mathbf{x}_{t,i}^f\}$, generate the posterior by

$$\mathbf{x}_{t,i}^a = \mathbf{x}_{t,i}^f + \hat{\mathbf{K}}_t\Big(\mathbf{y}_t + \mathbf{e}_{t,i} - \mathbf{H}_t\mathbf{x}_{t,i}^f\Big), \quad \mathbf{e}_{t,i} \sim (\mathbf{0}, \mathbf{R}).$$

  - EnKf asymptotically optimal if $p(\mathbf{x}_t|\mathbf{Y}^{t-1})$ and $p(\mathbf{y}_t|\mathbf{x_t})$ Gaussian;

- Common misconceptions about EnKf:

  1. Won't work if $p(\mathbf{x}_t|\mathbf{Y}^{t-1})$ is non-Gaussian or $\mathbf{G}$ non-linear;
     - will provide BLUP as $M \to \infty$.
     - EnKf "respects" non-Gaussian properties in prior *sample*,
  2. Must have $M \sim \mathcal{O}(\dim(\mathbf{x}_t))$;
     - sample error depends on spectrum of $\mathbf{P}_t^f$;
     - localization/tapering and square-root Kfs effectively remove errors due to sampling (*Furrer & Bengtsson, 2005).*

EnKF applied to Lorenz 96

- Atmospheric system with variables as k longitudes: $z_1, \ldots, z_{40}$. (Subscript denotes spatial location.)
- Equations: for $j = 1, \ldots, 40$,

$$\dot{z}_j = z_{j-1}(z_{j+1} - z_{j-2}) - z_j + F,$$

  where F represents forcing.

- The equations contain quadratic nonlinearities mimicking advection:

  $\dot{u}_i \ \propto \ u_i \frac{\partial u_i}{\partial x} \ \approx \ u_i(u_{i'} - u_{i^\star})/\delta x.$

- F is chosen so that phase space is bounded and the system exhibits chaotic behavior.
- *Simulations:* $m = 10$, 'short' lead time ($\delta_t = .05$),
  'observe' $z_1, z_3, \ldots, z_{39}$: $y_j = z_j + \epsilon_j$, $\epsilon_j \sim N(0, 4)$,

Where are we?

1. Sequential DA using ensemble Kalman filter.

2. ensemble Kalman filter $\rightarrow$ "full" Bayes DA.

3. Traffic example: DA in non-linear & non-Gaussian system.

4. DA in high-dimensional systems: what is reasonable?

Particle Filter Approximation - I

- Consider the general state-space model

$$\begin{aligned} \text{Observation:} &\quad \mathbf{y}_{t+1} \sim p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}) &\qquad (\mathbf{y}_t = H(\mathbf{x}_t) + \boldsymbol{\epsilon}_t) \\ \text{State evolution:} &\quad \mathbf{x}_{t+1} \sim p(\mathbf{x}_{t+1}|\mathbf{x}_t) &\qquad (\mathbf{x}_t = G(\mathbf{x}_{t-1}) + \boldsymbol{\eta}_t) \end{aligned}$$

- (The numerator of) Bayes theorem in the sequential DA setting:

$$p(\mathbf{y}_{t+1}, \mathbf{x}_{t+1}) \propto p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}) \int p(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{Y}^t) p(\mathbf{x}_t|\mathbf{Y}^t) d\mathbf{x}_t.$$

- With $\mathbf{x}^a_{t,i} \sim p(\mathbf{x}_t|\mathbf{Y}^t)$, the numerator (and posterior) is approximated by

$$p(\mathbf{y}_{t+1}, \mathbf{x}_{t+1}) \propto p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}) \frac{1}{M} \sum_{j=1}^{M} p(\mathbf{x}_{t+1}|\mathbf{x}^a_{t,i}).$$

Particle Filter Approximation - II

(continued) $p(\mathbf{x}_{t+1}|\mathbf{Y}^{t+1}) \propto p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1})\frac{1}{M}\sum_{j=1}^{M} p(\mathbf{x}_{t+1}|\mathbf{x}^a_{t,i})$

- When the densities on RHS are Gaussian, this "yields" the EnKf.

- Implements Kf recursion as $M \to \infty$

- Generalization to non-Gaussian case:

  Draw $\mathbf{x}^f_{t+1,i} \sim p(\mathbf{x}_{t+1}|\mathbf{x}^a_{t,i})$, and let $w_i \propto p(\mathbf{y}_{t+1}|\mathbf{x}^f_{t+1,i})$
  - We could:

  1. Accept $\mathbf{x}^f_{t+1,i}$, as a draw from posterior, with probability w_i; or,
  2. Approximate $p(\mathbf{x}_{t+1}|\mathbf{Y}^{t+1}) \approx \sum_{i=1}^{M} w_i \delta(\mathbf{x}_t - \mathbf{x}^f_{t+1,i})$; or,
  3. Develop further to produce: $w_{t,i} \to w_{t+1,i}$ (*particle filter*).

  - Implements Bayes theorem as $M \to \infty$

Particle Filter Approximation - III

- Particle filters/rejection/importance sampling algorithms are problematic in high-dimensions:

  - manifestation of the *curse-of-dimensionality*

- A particular *remedy* - The Auxiliary PF:

$$
\begin{aligned}
p(\mathbf{x}_{t+1}|\mathbf{y}_{t+1}) &\propto p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}) \int p(\mathbf{x}_{t+1}|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{Y}^t)d\mathbf{x}_t \\
&\approx p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}) \sum_j p(\mathbf{x}_{t+1}|\mathbf{x}^a_{t,j}) \\
&= \sum_j \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1})}{p(\mathbf{y}_{t+1}|\mu_{t+1,j})} p(\mathbf{y}_{t+1}|\mu_{t+1,j}) p(\mathbf{x}_{t+1}|\mathbf{x}^a_{t,j}) \\
&= \sum_j \frac{p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1})}{p(\mathbf{y}_{t+1}|\mu_{t+1,j})} g_{t,j} p(\mathbf{x}_{t+1}|\mathbf{x}^a_{t,j})
\end{aligned}
$$

- Here, $p(\mathbf{y}_{t+1}|\mu_{t+1,j})$ is a "high-density" area of the likelihood.
- Will not "solve" problem of uneven weights.

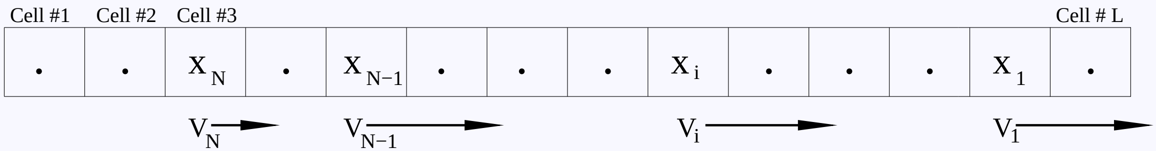
Where are we?

1. Sequential DA using ensemble Kalman filter.
2. ensemble Kalman filter $\longrightarrow$ "full" Bayes DA.
3. Traffic example: DA in non-linear & non-Gaussian system.
4. DA in high-dimensional systems: what is reasonable?

State-Space Dynamics

- We use the simple model of Nagel and Schrekenberg (1992) as described in Helbing (2001):

  A road stretch:



  - Let there be L cells numbered left to right.
  - Vehicles at locations $x_1, \ldots, x_N$, with velocities $v_1, \ldots, v_N$.
  - Location of the lead vehicle is x_1; location of last vehicle is x_N.
  - The state of the system is $\mathbf{x} = \{N, x_1, \ldots, x_N, v_1, \ldots, v_N\}$.
  - The $v_i \in \{0, 1, \ldots, 5\}$ and they satisfy $x_i + v_i \leq x_{i-1} - 1$.

State-Space Dynamics

The state transition mechanism $\mathbf{x} \longrightarrow \mathbf{x}'$ is as follows:

1. Change velocities:

$$v_i \longrightarrow v_i' = \max\{0, \min(v_i + 1, x_{i-1} - x_i - 1, 5) - \xi_i\},$$

where $\xi_i \overset{iid}{\sim} \text{Bernoulli}(p)$.

2. Move vehicles:

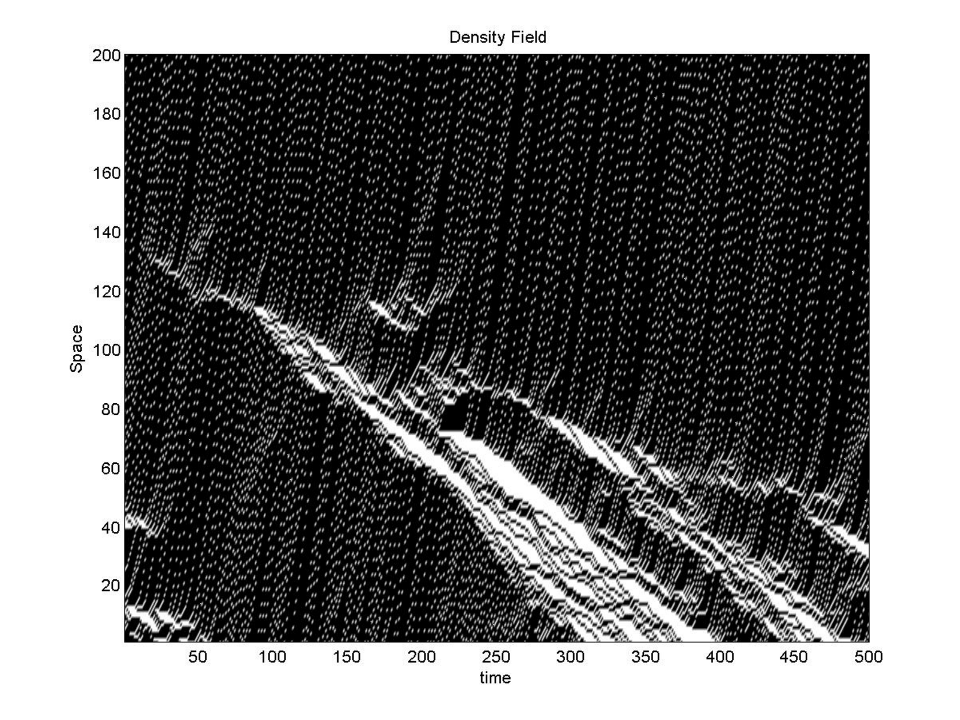
$$x_i \longrightarrow x_i' = x_i + v_i'$$

3. Adjust N: Remove lead vehicle and/or add new vehicle: e.g.,
   - If $x_1 + v_1' > L$, the lead car is removed w.p. p_{exit}.
   - a new car is added with probability p_{new} at location x_{N+1} chosen uniformly in $\{1, 2, \ldots, \min(5, x_N' - v_N' - 1)\}$.

- The above defines $p(\mathbf{x}_{t+1} | \mathbf{x}_t, p)$.

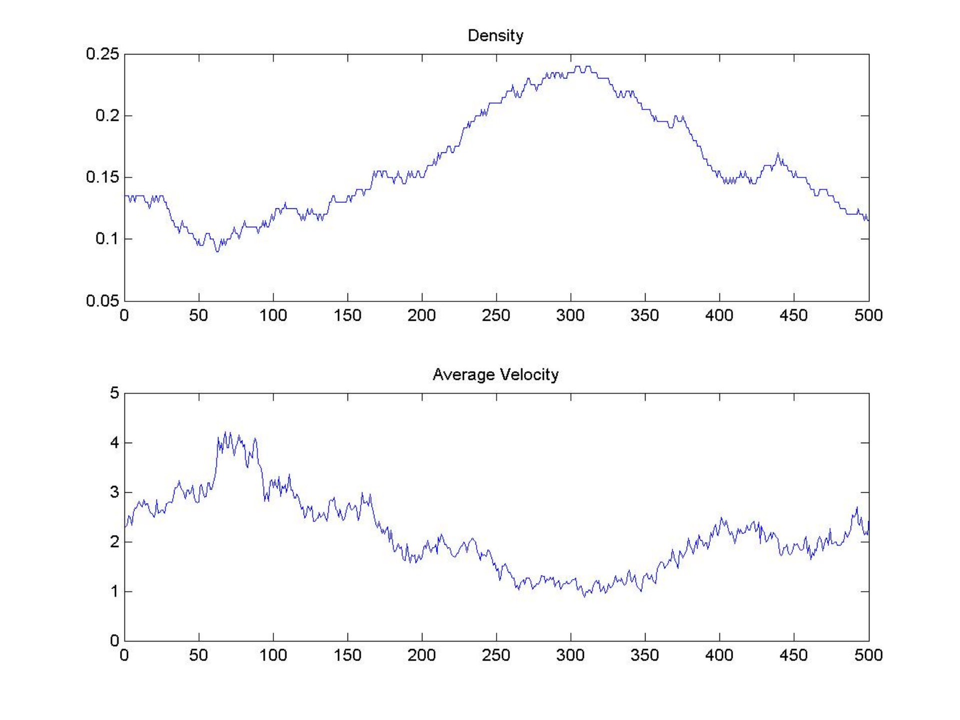
Density Field

Illustration using $L = 200,\ N(\text{start}) = 50,\ p = .5,\ T = 500.$



Density and Average Velocity

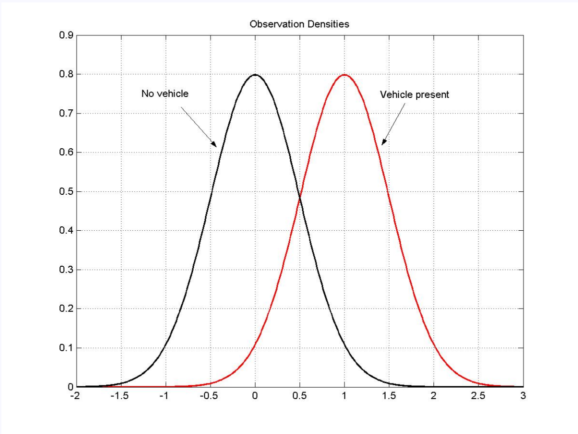
Illustration using $L = 200, N(\text{start}) = 50,\ p = .5,\ T = 500.$



Observation Model

- A simple model for the observations: with $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$,

$$Y^i = \begin{cases} 1 + \epsilon_i, & \text{cell } i \text{ occupied;} \\ \epsilon_i & \text{cell } i \text{ not occupied.} \end{cases}$$



– Assumption of independence (and normality) can be relaxed, but yields more complicated updating mechanism.

- The above defines $p(\mathbf{y}_{t+1} | \mathbf{x}_t, \sigma^2)$.

Particle Filter Approach

- A sequential importance sampler (e.g., particle filter) is obtained by a recursion on the weights $w^i_{t-1} \times p(\mathbf{y}_t|\mathbf{x}^i_t) \rightarrow w^i_t$:

$$\overbrace{\{\mathbf{z}^i_t, w^i_{t-1}\}}^{prior} \text{ and } \overbrace{p(\mathbf{y}_t|\mathbf{x}_{t,i})}^{likelihood} \quad \xrightarrow{Bayes} \quad \overbrace{\{\mathbf{x}_{t,i}, w^i_{t-1} \times p(\mathbf{y}_t|\mathbf{x}_{t,i})\}}^{Posterior}$$

  – This produces a *likelihood filter*.

- We will adapt the PF to include observations from one-step ahead.

- The particle approach:

$$p(\mathbf{x}_{t+1}|\mathbf{Y}^{t+1}) \propto p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}) \int p(\mathbf{x}_{t+1}|\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{Y}^t) \mathrm{d}\mathbf{x}_t$$
$$\approx \frac{1}{M} \sum_{i=1}^{M} p(\mathbf{x}_{t+1}|\mathbf{x}_t^i) p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1}) \quad (1)$$

– Because of the system properties, we can sample from (1) directly, without using a rejection method or importance sampling.

Trick: Multiplying and dividing (1) by $p(\mathbf{y}_{t+1}|\mathbf{z}_t^i)$

$$p(\mathbf{x}_{t+1}|\mathbf{Y}^{t+1}) \propto \frac{1}{M} \sum_{i=1}^{M} p(\mathbf{y}_{t+1}|\mathbf{x}_t^i) \frac{p(\mathbf{x}_{t+1}|\mathbf{x}_t^i) p(\mathbf{y}_{t+1}|\mathbf{x}_{t+1})}{p(\mathbf{y}_{t+1}|\mathbf{x}_t^i)}$$
$$= \frac{1}{M} \sum_{j=1}^{M} p(\mathbf{y}_{t+1}|\mathbf{x}_t^i) p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \mathbf{y}_{t+1}) \quad (2)$$

– Both densities in (2) are computable.

Sampling Procedure

Want: $p(\mathbf{x}_{t+1}|\mathbf{Y}^{t+1}) \approx \frac{1}{M}\sum_{j=1}^{M} p(\mathbf{y}_{t+1}|\mathbf{x}_t^i)p(\mathbf{x}_{t+1}|\mathbf{x}_t^i, \mathbf{y}_{t+1})$

- Assume a random sample $\mathbf{x}_t^i \sim p(\mathbf{x}_t|\mathbf{Y}^t)$ and generate a draw from posterior:

  1. Sample $\tilde{\mathbf{x}}_t^i = \mathbf{x}_t^i$ with probability proportional to $p(\mathbf{y}_{t+1}|\mathbf{x}_t^i)$
  2. Drawing $\mathbf{x}_{t+1}^i \sim p(\mathbf{x}_{t+1}|\tilde{\mathbf{x}}_t^i, \mathbf{y}_{t+1})$

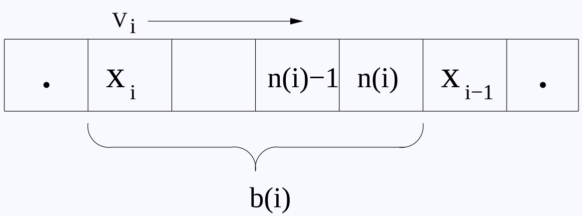
  – We do this M times to obtain updated particles $\{\mathbf{x}_{t+1}^i, \frac{1}{M}\}$.

Evaluating $p(\mathbf{y}_{t+1}|\mathbf{x}_t^j)$: Strategy

- In our setting, there are two aspects of the state-transition dynamics that drastically simplify simulation and particle filter approximations:

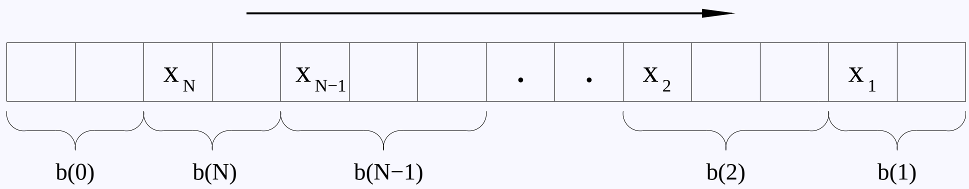
  1. vehicles are moved independently of one another; and, each vehicle can only move to one of two possible positions
  2. dependence of blocks of the y_i's (measurement on cell i) on vehicle locations is very simple.

- Illustration: for particle j, let $b(i)$ be the set of possible locations for vehicle i at time $t+1$



- Allows evaluation of $p(\mathbf{y}_{t+1}|\mathbf{x}_t^j) = \prod_i p(y_{t+1}^{b(i)}|\mathbf{x}_t^j)$

Evaluating $p(\mathbf{y}_{t+1}|\mathbf{x}_t^j)$: Specifically



- Let $b(i)$ index the data which depends on vehicle i. We want to evaluate

$$p(\mathbf{y}_{t+1}|\mathbf{x}_t^j) = \prod_{i=1}^{N} p(y_{t+1}^{b(i)}|\mathbf{x}_t^j).$$

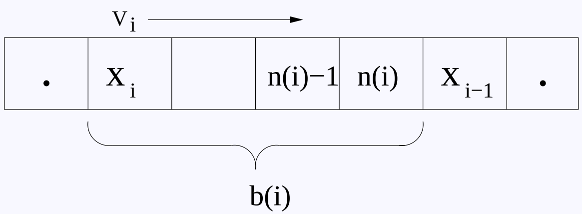
  – After much cancelation,

$$p(\mathbf{y}_{t+1}|\mathbf{x}_t^j) \propto \frac{p(y_{t+1}^{b(1)}|\mathbf{x}_t^j)}{\prod_{k\in b(1)} \phi\left(\frac{y_{t+1}^k}{\sigma}\right)} \prod_{i=2}^{N} \left[(1-p)\exp\left(\frac{y_{t+1}^{n(i)}}{\sigma^2}\right) + p\exp\left(\frac{y_{t+1}^{n(i)-1}}{\sigma^2}\right) \right].$$

- Draw state $\mathbf{x}_t^j$ with probability $\frac{p(\mathbf{y}_{t+1}|\mathbf{x}_t^j)}{\sum_k p(\mathbf{y}_{t+1}|\mathbf{x}_t^k)}$.

Moving vehicles according to $\mathbf{x}_{t+1}^j \sim p(\mathbf{x}_{t+1}|\tilde{\mathbf{x}}_t^j, \mathbf{y}_{t+1})$

- Consider drawing $\mathbf{x}_{t+1}^j \sim p(\mathbf{x}_{t+1}|\tilde{\mathbf{x}}_t^j, \mathbf{y}_{t+1})$. This can again be done vehicle by vehicle:



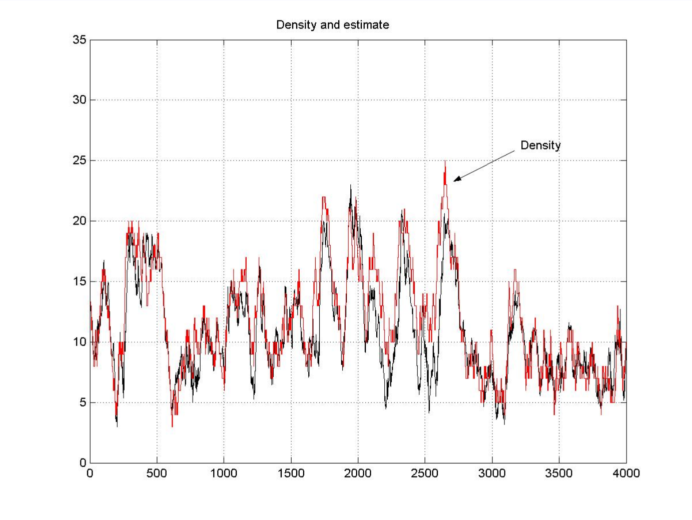
– for vehicle i, we randomly choose the move corresponding to $n(i)$ or $n(i)-1$, by evaluating the ratio

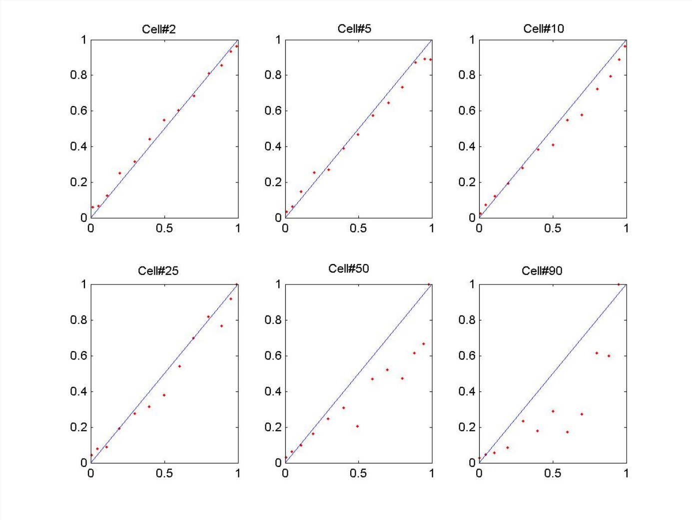
$$\alpha_i = \frac{p(y_{t+1}^{b(i)}|\xi_i = 1)p(\xi_i = 1)}{p(y_{t+1}^{b(i)}|\xi_i = 0)p(\xi_i = 0)} = \frac{p \exp\left(\frac{y_{t+1}^{n(i)-1}}{\sigma^2}\right)}{(1-p)\exp\left(\frac{y_{t+1}^{n(i)}}{\sigma^2}\right)},$$

and ξ_i is then chosen to be 1 with probability $\alpha_i/(1+\alpha_i)$.

Filter Performance

- Left: Density and estimated density.
- Right: Probability forecast verification:





What now?

- Remains to be done

  -Initialization.

  -Recursive parameter estimation.

- For realistic application:

  -Extend to correlated measurement errors.

  -How general is sampling scheme when model is more complex?