Non-Gaussian Filters

Several methods that account for non-Gaussian prior distributions and observation likelihood functions.

Rank Histogram Filter - Anderson 2010

Idea: use a rank histogram to construct a non-parametric form of prior distribution, then use Bayesian approach to update this distribution with observation likelihood.

$\Rightarrow$ Well handles prior distributions that have multiple modes, skewness, kurtosis, or finite bounds.

Consider the update step in observation space:

$$p(y|y^o) = \frac{p(y^o|y)\,p(y)}{\int p(y^o|y)\,p(y)}$$

1. Construct prior distribution $p(y)$

Sort the ensemble members, y_k, $k=1,2,\cdots,N$

Define $p(y)$ on $N+1$ regions, separated by each y_k, and each region has probability mass of $\frac{1}{N+1}$.

$$p(y) = \begin{cases} \mathcal{N}(\mu_1, \sigma_b^2), & y < y_1 \\ \frac{1}{N+1}\frac{1}{y_2 - y_1}, & y_1 \le y < y_2 \\ \vdots & \vdots \\ \frac{1}{N+1}\frac{1}{y_N - y_{N-1}}, & y_{N-1} \le y < y_N \\ \mathcal{N}(\mu_2, \sigma_b^2), & y \ge y_N \end{cases}$$

Left and right "wings" are Gaussian tails, $\sigma_b^2 = \frac{1}{N-1}\sum_{k=1}^{N}(y_k - \bar{y})^2$ and μ_1, μ_2 chosen so that $\int_{-\infty}^{y_1} \mathcal{N}(\mu_1, \sigma_b^2)\,dy = \frac{1}{N+1}$

and $\int_{y_N}^{\infty} \mathcal{N}(\mu_2, \sigma_b^2)\,dy = \frac{1}{N+1}$

2. Observation Likelihood $p(y^o|y) = N(y^o, \sigma_o^2)$
   For now, assume Gaussian distribution, but can be relaxed to other types of distribution too!
   Just need to evaluate $p(y^o|y_k) = \frac{1}{\sqrt{2\pi\sigma_o^2}} \exp\left(-\frac{(y_k - y^o)^2}{2\sigma_o^2}\right)$
   at each y_k, and assume linear function shape between y_k and y_{k+1}, $k = 1, 2, \cdots, N-1$ $\Rightarrow$ to avoid intense computation

3. Find posterior distribution
   Integrate $p(y^o|y)\,p(y)$ for each region to get posterior probability mass. $Z_1 = \int_{-\infty}^{y_1} p(y^o|y)\,p(y)\,dy$

$$Z_{N+1} = \int_{y_N}^{\infty} p(y^o|y)\,p(y)\,dy$$

$$Z_k = \int_{y_k}^{y_{k+1}} p(y^o|y)\,p(y)\,dy = \frac{1}{N+1}\,\frac{1}{y_{k+1}-y_k}\left(p(y^o|y_{k+1}) - p(y^o|y_k)\right)$$

for $k = 1, 2, \cdots, N-1$

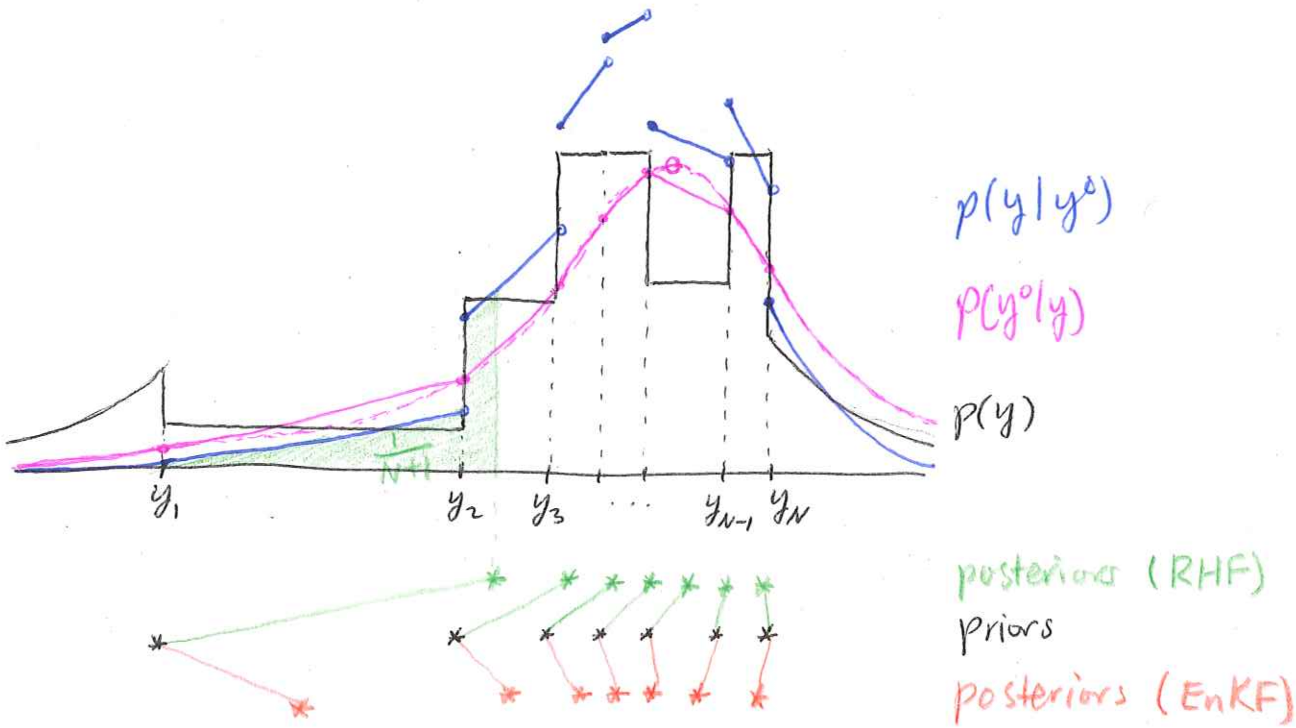
Normalization factor

$$\int_{-\infty}^{\infty} p(y^o|y)\,p(y)\,dy = \sum_{k=1}^{N+1} Z_k \equiv Z$$

final posterior $p(y|y^o) = p(y^o|y)\,p(y)/Z$

4. Find analysis members.
   Integrate $p(y|y^o)$ from left to right and locate the k-th member, so that the cumulative probability mass

$$\int_{-\infty}^{y_k^a} p(y|y^o)\,dy = \frac{k}{N+1}$$



⇒ For serial EnKF, the equation $\delta y_k = \dfrac{\sigma_b^2}{\sigma_b^2+\sigma_o^2}(y^o - y_k)$ is applied to each member $k=1,2,\cdots,N$;

The update is proportional to innovation, and the ratio is fixed.

→ Does not handle outliers well.

⇒ Assumption of tail shape influence behavior of RHF.

Quadratic Filter Hodyss 2011

Idea: account for skewness in posterior distribution by keeping extra terms in its Taylor expansion

Bayesian approach gives posterior distribution $p(x|y^o)$

The analysis is found by $x^a = \int x\, p(x|y^o)\, dx$

prior information about x is that its distribution is $p(x)$ with mean x^b, change of variable $x = x^b + \varepsilon^b$

$$\rightarrow \quad x^a = x^b + \underbrace{\int \varepsilon^b\, p(\varepsilon^b | y^o, x^b)\, d\varepsilon^b}_{f(y^o, x^b)}$$

Define innovation $d_j = y_j^o - H_j x^b$, $j = 1, 2, \cdots, p$

f is the analysis increment function, it can be expressed as

$$f = f_0 + \underbrace{\frac{\partial f}{\partial d}}_{G_1} d + \underbrace{\frac{1}{2}\frac{\partial^2 f}{\partial d^2}}_{G_2} d^2 + \cdots = f_0 + G\hat{d}$$

where $d^2 = d \otimes d$ (Kronecker product)

$d^3 = d \otimes d \otimes d$, and so on.

$$G = (G_1 \ G_2 \cdots G_\infty)$$

$$\hat{d} = \begin{pmatrix} d \\ d^2 \\ \vdots \\ d^\infty \end{pmatrix}$$

Analysis error $\varepsilon^a = x - x^a = \underbrace{(x - x^b)}_{\varepsilon^b} - (f_0 + G\hat{d})$

Determine coefficients in G by taking expectation:

$$E(\varepsilon^a) = E(\varepsilon^b) - (f_0 + G\,E(\hat{d})) = 0$$

$$\Rightarrow \quad f_0 = -G\,E(\hat{d}) \quad \text{gives unbiased analysis.}$$

Define $\hat{d}' = \hat{d} - \mathbb{E}(\hat{d})$, we have $\varepsilon^a = \varepsilon^b - G\hat{d}'$

$$P^a = \mathbb{E}(\varepsilon^a \varepsilon^{aT}) = \mathbb{E}(\varepsilon^b \varepsilon^{bT}) - \mathbb{E}(\varepsilon^b \hat{d}'^T)G^T - G\,\mathbb{E}(\hat{d}'\varepsilon^{bT}) + G\,\mathbb{E}(\hat{d}'\hat{d}'^T)G^T$$

Minimizing analysis error variances gives $\dfrac{\partial\, tr(P^a)}{\partial G} = 0$

$$G = \mathbb{E}(\varepsilon^b \hat{d}'^T)\,\mathbb{E}(\hat{d}'\hat{d}'^T)^{-1}$$

update eqns: $x^a = x^b + G\hat{d}'$

$$P^a = P^b - G\,\mathbb{E}(\hat{d}'\varepsilon^{bT})$$

Assume $\mathbb{E}(d) = 0 \rightarrow \varepsilon^o$ and ε^b unbiased.

$$\mathbb{E}(\hat{d}) = \begin{bmatrix} 0 & \mathbb{E}(d^2)^T & \cdots \end{bmatrix}^T$$

$$\mathbb{E}(\hat{d}'\hat{d}'^T) = \mathbb{E}(\hat{d}\hat{d}^T) - \mathbb{E}(\hat{d})\,\mathbb{E}(\hat{d})^T$$

$$\mathbb{E}(\hat{d}\hat{d}^T) = \begin{pmatrix} HP^bH^T + R & HT^bH^{2T} & \cdots \\ H^2T^{bT}H^T & H^2F^bH^{2T} + A + B + C + R_4 & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

$F^b = \mathbb{E}(\varepsilon^{b2}\varepsilon^{b2T})$

$R_4 = \mathbb{E}(\varepsilon^{o2}\varepsilon^{o2T})$

$T^b = \mathbb{E}(\varepsilon^b \varepsilon^{b2T})$

See A, B, C, values in Hodyss 2011.

$$\mathbb{E}(\hat{d})\,\mathbb{E}(\hat{d})^T = \begin{pmatrix} 0 & 0 & \cdots \\ 0 & \mathbb{E}(d^2)\,\mathbb{E}(d^2)^T & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix}$$

$$\mathbb{E}(\varepsilon^b \hat{d}'^T) = \begin{pmatrix} P^bH^T & T^bH^{2T} & \cdots \end{pmatrix}$$

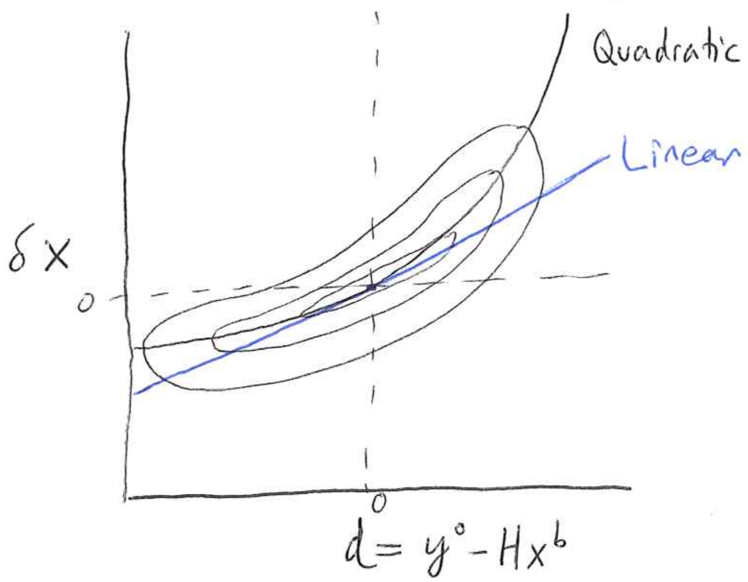
Linear filter truncates at G, d: Kalman filter update

Quadratic filter truncates at $(G_1 \;\; G_2)\begin{pmatrix} d \\ d^2 \end{pmatrix}$:

$$x^a \cong x^b + Kd$$
$$+ (I - KH)T^b H^{2T} \pi^{-1} \left(d^{2\prime} - H^2 T^{bT} H^T (HP^bH^T + R)^{-1} d \right)$$

$$P^a \cong P^b - KHP^b$$
$$- (I - KH) T^b H^{2T} \Pi^{-1} H^2 T^{bT} (I - H^T K^T)$$

where $\Pi = H^2 F^b H^{2T} + A + B + C + R_4$
$$- H^2 T^{bT} H^T (HP^bH^T + R)^{-1} H T^b H^{2T} - \mathbb{E}(d^2)\, \mathbb{E}(d^2)^T .$$



GIGG filter Bishop 2016

Idea: use the correct form of distribution for prior and observation likelihood in Bayesian approach and derive the update eqns for each scenario.

$$p(y|y^o) \propto p(y^o|y)\, p(y)$$

posterior	observation likelihood	prior	
Gaussian	Gaussian	Gaussian	↪ EnKF
Gamma	Inverse-Gamma	Gamma	→ GIG
Inverse-Gamma	Gamma	Inverse-Gamma	→ IGG

GIG: Γ prior distribution with shape $k = P^{-1}$, scale $\theta = \bar{y}^b P$

Γ^{-1} observation likelihood with shape $\alpha = \tilde{R}^{-1} + 1$, scale $\beta = y\tilde{R}^{-1}$

$$p(y) = \frac{1}{\Gamma(k)} \frac{1}{\theta^k} y^{k-1} \exp\left(-\frac{y}{\theta}\right)$$

$$p(y^o|y) = \frac{\beta^\alpha}{\Gamma(\alpha)} (y^o)^{-\alpha-1} \exp\left(-\frac{\beta}{y^o}\right)$$

$\Rightarrow \Gamma$ posterior distribution with shape $k = \Pi^{-1}$, scale $\theta = \bar{y}^a \Pi$

$$P = \frac{var(y^b)}{(\bar{y}^b)^2}, \quad \tilde{P} = \frac{var(y^b)}{var(y^b) + (\bar{y}^b)^2}, \quad R = \frac{var(y^o)}{y^2}, \quad \tilde{R} = \frac{var(y^o)}{var(y^o) + y^2}$$

$$\Pi = \left(\tilde{P}^{-1} + \tilde{R}^{-1}\right)^{-1} = \tilde{P} - \tilde{P}(\tilde{P} + \tilde{R})^{-1}\tilde{P}$$

$$\frac{1}{\bar{y}^a} = \frac{1}{\bar{y}^b} + \frac{\tilde{P}}{\tilde{P} + \tilde{R}}\left(\frac{1}{y^o} - (\tilde{R}+1)\frac{1}{\bar{y}^b}\right)$$

IGG: Γ^{-1} prior with $\alpha = \tilde{P}^{-1} + 1$, $\beta = \tilde{P}^{-1}\bar{y}^b$, Γ obs likelihood with $k = R^{-1}$, $\theta = yR$

$\Rightarrow \Gamma^{-1}$ posterior with $\alpha = \tilde{\Pi}^{-1} + 1$, $\beta = \tilde{\Pi}^{-1}\bar{y}^a$

$$\tilde{\Pi} = (\tilde{P}^{-1} + R^{-1})^{-1}, \quad \bar{y}^a = \bar{y}^b + \frac{\tilde{P}}{\tilde{P} + R}(y^o - \bar{y}^b)$$