# Lecture 12: Particle Filters and the Rank Histogram Particle Filter

Introduction to Data Assimilation

November 21, 2025

## 1 Basic Particle Filters: Ideas, Derivations, and History

### 1.1 Nonlinear filtering setup

We consider the same discrete-time filtering problem as in previous lectures, but without any Gaussian assumption. Let the state and observation satisfy

$$\boldsymbol{u}_{m+1} = \boldsymbol{f}(\boldsymbol{u}_m) + \boldsymbol{\sigma}_{m+1}, \tag{1}$$

$$\boldsymbol{v}_{m+1} = \boldsymbol{g}(\boldsymbol{u}_{m+1}) + \boldsymbol{\sigma}^o_{m+1}, \tag{2}$$

where

- $\boldsymbol{u}_m \in \mathbb{R}^N$ is the (possibly high-dimensional) state;
- $\boldsymbol{v}_m \in \mathbb{R}^M$ is the observation;
- $\boldsymbol{f}$ is a (possibly strongly) nonlinear model map;
- $\boldsymbol{g}$ is a (possibly nonlinear) observation operator;
- $\boldsymbol{\sigma}_m$ and $\boldsymbol{\sigma}^o_m$ are model and observation noises.

The *filtering problem* is to characterize the posterior distribution

$$p_{m|m}(\boldsymbol{u}) := p(\boldsymbol{u}_m = \boldsymbol{u} \mid \boldsymbol{v}_1, \ldots, \boldsymbol{v}_m),$$

and in practice to approximate its mean, variance, and other statistics.

In continuous time, the model is often written as an SDE

$$d\boldsymbol{u} = \boldsymbol{f}(\boldsymbol{u}, t)\, dt + \boldsymbol{\sigma}(\boldsymbol{u}, t)\, d\boldsymbol{W}(t), \tag{3}$$

where $\boldsymbol{W}(t)$ is a vector of independent Wiener processes ($\boldsymbol{\sigma}$ is nonlinear noise term here). The corresponding prior density $p(\boldsymbol{u}, t)$ satisfies the Fokker–Planck equation

$$\partial_t p = -\nabla \cdot (\boldsymbol{f} p) + \tfrac{1}{2}\nabla^2\big((\boldsymbol{\sigma} \otimes \boldsymbol{\sigma}^\top) p\big), \tag{4}$$

with initial condition $p(\boldsymbol{u}, t_0) = p_{0|0}(\boldsymbol{u})$.

For high-dimensional turbulent systems, solving this PDE is infeasible. Particle filters replace the exact density $p_{m|m}$ by a random empirical measure supported on a finite ensemble of *particles*.

## 1.2 Empirical representation of the posterior

At time $t_m$, suppose the posterior distribution $p_{m|m}(\boldsymbol{u})$ is approximated by a weighted ensemble

$$\{\boldsymbol{u}_{m|m}^{(k)},\, w_m^{(k)}\}_{k=1}^K,$$

with normalized nonnegative weights

$$w_m^{(k)} \geq 0, \qquad \sum_{k=1}^K w_m^{(k)} = 1.$$

The posterior density is represented as

$$p_{m|m}(\boldsymbol{u}) \approx \sum_{k=1}^K w_m^{(k)}\, \delta(\boldsymbol{u} - \boldsymbol{u}_{m|m}^{(k)}), \tag{5}$$

where $\delta$ is the Dirac measure. This is a random measure; the approximation error is a sampling error which becomes severe as dimension grows ("curse of dimensionality").

The posterior mean and covariance are approximated by

$$\bar{\boldsymbol{u}}_{m|m} \approx \sum_{k=1}^K w_m^{(k)}\, \boldsymbol{u}_{m|m}^{(k)}, \tag{6}$$

$$\mathrm{Cov}(\boldsymbol{u}_m \mid \mathcal{F}_m) \approx \sum_{k=1}^K w_m^{(k)} \big(\boldsymbol{u}_{m|m}^{(k)} - \bar{\boldsymbol{u}}_{m|m}\big)\big(\boldsymbol{u}_{m|m}^{(k)} - \bar{\boldsymbol{u}}_{m|m}\big)^\top. \tag{7}$$

## 1.3 Propagation step: prior particles

The first step at time $t_{m+1}$ is to approximate the *prior* density

$$p_{m+1|m}(\boldsymbol{u}) := p(\boldsymbol{u}_{m+1} = \boldsymbol{u} \mid \boldsymbol{v}_1, \ldots, \boldsymbol{v}_m).$$

Formally, $p_{m+1|m}$ is obtained by solving the Fokker–Planck equation from $t_m$ to $t_{m+1}$ with initial condition $p_{m|m}$. Alternatively, one can solve the discrete model (1).

In the particle filter, instead of evolving the PDE, we propagate the particles under the SDE (3):

$$\boldsymbol{u}_{m+1|m}^{(k)} = \Phi(\boldsymbol{u}_{m|m}^{(k)}, \boldsymbol{\omega}_{m+1}^{(k)}),$$

where $\Phi$ is the numerical flow map for (3), and $\boldsymbol{\omega}_{m+1}^{(k)}$ are independent realizations of the driving noise increment.

If we assume *all posterior particles at $t_m$ are equally likely*, that is $w_m^{(k)} = 1/K$, we obtain an equally weighted prior ensemble at $t_{m+1}$:

$$\boldsymbol{u}_{m+1|m}^{(k)} = \Phi(\boldsymbol{u}_{m|m}^{(k)}, \boldsymbol{\omega}_{m+1}^{(k)}), \qquad k = 1, \ldots, K, \tag{8}$$

$$p_{m+1|m}(\boldsymbol{u}) \approx \frac{1}{K} \sum_{k=1}^K \delta(\boldsymbol{u} - \boldsymbol{u}_{m+1|m}^{(k)}). \tag{9}$$

## 1.4 Update step: Bayes formula and weight update

At time $t_{m+1}$, we receive an observation $\boldsymbol{v}_{m+1}$ with likelihood

$$p(\boldsymbol{v}_{m+1} \mid \boldsymbol{u}_{m+1}) =: \mathcal{L}(\boldsymbol{v}_{m+1}; \boldsymbol{u}_{m+1}),$$

which may have non-Gaussian errors and nonlinear dependence on $\boldsymbol{u}_{m+1}$.

Bayes' theorem gives the exact posterior density

$$p_{m+1|m+1}(\boldsymbol{u}) = \frac{p(\boldsymbol{v}_{m+1} \mid \boldsymbol{u})\, p_{m+1|m}(\boldsymbol{u})}{\displaystyle\int p(\boldsymbol{v}_{m+1} \mid \tilde{\boldsymbol{u}})\, p_{m+1|m}(\tilde{\boldsymbol{u}})\, d\tilde{\boldsymbol{u}}}. \tag{10}$$

Substituting the empirical prior (9) into (10) yields

$$p_{m+1|m+1}(\boldsymbol{u}) \approx \sum_{k=1}^{K} w_{m+1}^{(k)}\, \delta(\boldsymbol{u} - \boldsymbol{u}_{m+1|m}^{(k)}), \tag{11}$$

$$w_{m+1}^{(k)} = \frac{\mathcal{L}(\boldsymbol{v}_{m+1}; \boldsymbol{u}_{m+1|m}^{(k)})}{\displaystyle\sum_{\ell=1}^{K} \mathcal{L}(\boldsymbol{v}_{m+1}; \boldsymbol{u}_{m+1|m}^{(\ell)})}. \tag{12}$$

Thus the *weights* carry the observational information; the particle locations are unchanged in the basic scheme.

## 1.5 Weight degeneracy and resampling

In high-dimensional or strongly chaotic systems, iterating the "propagate–weight" cycle ((9) and (12)) leads to *weight degeneracy*:

$$w_{m+1}^{(k)} \approx 0 \text{ for most } k, \quad w_{m+1}^{(k^*)} \approx 1 \text{ for some } k^*.$$

This is the particle-filter manifestation of the *curse of dimensionality*: the effective sample size collapses to 1, and the empirical posterior collapses to a single particle.

A standard remedy is *sequential importance resampling* (SIR):

(a) Compute weights $w_{m+1}^{(k)}$ by (12).

(b) Resample $K$ new particles $\{\boldsymbol{u}_{m+1|m+1}^{(k)}\}$ by drawing indices $I_k$ from the discrete distribution $\mathbb{P}(I = k) = w_{m+1}^{(k)}$ and setting $\boldsymbol{u}_{m+1|m+1}^{(k)} = \boldsymbol{u}_{m+1|m}^{(I_k)}$.

(c) Reset weights to $w_{m+1}^{(k)} = 1/K$.

This avoids catastrophic weight collapse but introduces *duplicate* particles which then diverge only through model noise. For deterministic models (no model noise) the support can quickly become low rank, and for high-dimensional systems (even with stochastic forcing) enormous ensemble sizes are required for accurate sampling.

## 1.6 Historical remarks

- The basic bootstrap particle filter (SIR) was proposed in the early 1990s in the statistics and engineering communities as a nonlinear, non-Gaussian generalization of the Kalman filter.
- Rigorous analysis of particle filters for nonlinear SDEs and Markov processes was developed by Del Moral and co-authors in the 1990s and 2000s; Doucet, de Freitas, and Gordon popularized "sequential Monte Carlo" as a general framework.
- For *high-dimensional geophysical systems*, work by Bickel, Bengtsson, Snyder, van Leeuwen and others showed that standard importance sampling ideas fail catastrophically: the required number of particles grows essentially exponentially in an effective dimension related to the attractor or observation space.
- This motivated alternative ensemble methods (EnKF, ETKF, EAKF, hybrid filters) and new particle-filter ideas designed specifically for high-dimensional, weakly observed, turbulent systems, such as localization, nudging, and marginal-based filters like the rank histogram filter.

## 1.7 Comparison with Kalman-type filters (high-level)

Kalman-type filters (KF, EKF, EnKF, ETKF, EAKF) represent the posterior by its mean and covariance, typically assuming approximate Gaussianity:

- **KF:** exact for linear–Gaussian systems; closed analytic formulas.
- **EKF:** local linearization of nonlinear dynamics and observations.
- **EnKF:** ensemble representation of mean/covariance under Gaussian prior and linear (or linearized) observations.
- **ETKF/EAKF:** ensemble square-root filters that deterministically adjust anomalies to match the Kalman posterior covariance.

In contrast, a particle filter tries to approximate the *full posterior density* $p_{m|m}(\boldsymbol{u})$, allowing:

- strongly nonlinear dynamics and observations,
- multi-modal and heavy-tailed distributions,
- arbitrary (non-Gaussian) likelihoods.

The price is severe sampling error in high dimension with small ensemble size. Designing practical particle filters that remain skillful in realistic turbulent regimes with $K \sim 10$ is still a major research challenge.

# 2 Rank Histogram Particle Filter (RHF)

## 2.1 Motivation and overview

Standard particle filters suffer from weight degeneracy when the state is high-dimensional and observations are sparse or precise. One way to mitigate this is to avoid representing the full joint posterior $p(\boldsymbol{u} \mid \boldsymbol{v})$ explicitly and instead work with *marginal* posterior distributions in each coordinate (or in low-dimensional subspaces).

The **rank histogram particle filter** (RHF), due to Jeff Anderson, is a particle filter designed to:

- use small ensemble sizes ($K \sim 3$–$10$),
- avoid weight collapse by keeping particles equally weighted,
- update particles based on marginal rank information (rank histograms),
- be applicable even for deterministic dynamical models.

The central idea is:

1. For each state component $u_j$ (or projected coordinate), construct a piecewise-constant approximation of the marginal prior density from the sorted particles.
2. Approximate the likelihood along that coordinate.
3. Use Bayes' theorem to construct a marginal posterior density.
4. Sample new equally weighted particles from that posterior by enforcing equal-mass quantiles.

This is done *coordinate-wise* and can be parallelized. The RHF thus replaces importance weights by a deterministic transformation of particle locations.

## 2.2 Notation and setup

Assume we have an equally weighted prior ensemble at time $t_{m+1}$:

$$\{\boldsymbol{u}^{(k)}_{m+1|m}\}_{k=1}^{K}, \qquad p_{m+1|m}(\boldsymbol{u}) \approx \frac{1}{K}\sum_{k=1}^{K} \delta(\boldsymbol{u} - \boldsymbol{u}^{(k)}_{m+1|m}).$$

We focus on a single component $u_j$ of the state:

$$u_j^{(k)} := \left(\boldsymbol{u}^{(k)}_{m+1|m}\right)_j, \qquad j = 1, \dots, N.$$

For clarity we suppress the $j$ index and write $u^{(k)}$ instead of $u_j^{(k)}$ in this section.

### 2.2.1 Step 1: Construct a marginal prior $p_{m+1|m,j}(u)$

First, sort the particles in coordinate $j$:

$$u^{(1)} < u^{(2)} < \cdots < u^{(K)}.$$

The RHF assumes that these sorted particles partition the real line into $K+1$ intervals each carrying equal prior probability mass $1/(K+1)$. On the *interior* intervals $[u^{(k)}, u^{(k+1)}]$ the prior density is taken to be constant:

$$p_{m+1|m,j}(u) = \frac{1}{K+1} \sum_{k=1}^{K-1} \chi_{[u^{(k)},u^{(k+1)}]}(u) \, \frac{1}{u^{(k+1)} - u^{(k)}} + \text{tails},$$

where $\chi_{[a,b]}(u)$ is the characteristic function of the interval $[a, b]$:

$$\chi_{[a,b]}(u) = \begin{cases} 1, & a \le u \le b, \\ 0, & \text{otherwise.} \end{cases}$$

For the two *unbounded* intervals $(-\infty, u^{(1)}]$ and $[u^{(K)}, \infty)$ RHF uses Gaussian tails. Let $\mu_j$ and $\sigma_j^2$ be the empirical mean and variance of the prior particles in coordinate $j$:

$$\mu_j = \frac{1}{K} \sum_{k=1}^{K} u^{(k)}, \tag{13}$$

$$\sigma_j^2 = \frac{1}{K-1} \sum_{k=1}^{K} (u^{(k)} - \mu_j)^2. \tag{14}$$

Define $p_G(u \mid \mu, \sigma)$ to be the Gaussian pdf with mean $\mu$ and standard deviation $\sigma$. RHF chooses means $\mu_1$ and $\mu_K$ such that the Gaussian tails carry mass $1/(K+1)$:

$$\int_{-\infty}^{u^{(1)}} p_G(u \mid \mu_1, \sigma_j) \, du = \frac{1}{K+1}, \tag{15}$$

$$\int_{u^{(K)}}^{\infty} p_G(u \mid \mu_K, \sigma_j) \, du = \frac{1}{K+1}. \tag{16}$$

With these choices, the full marginal prior density is

$$\begin{aligned} p_{m+1|m,j}(u) = {} & \frac{1}{K+1} \sum_{k=1}^{K-1} \chi_{[u^{(k)},u^{(k+1)}]}(u) \, \frac{1}{u^{(k+1)} - u^{(k)}} \\ & + p_G(u \mid \mu_1, \sigma_j) \, \chi_{(-\infty,u^{(1)}]}(u) + p_G(u \mid \mu_K, \sigma_j) \, \chi_{[u^{(K)},\infty)}(u). \end{aligned} \tag{17}$$

This is a piecewise continuous PDF with $K+1$ equal-probability regions.

### 2.2.2 Step 2: Approximate the marginal likelihood

Suppose the observation $\boldsymbol{v}_{m+1}$ has $j$-th component

$$v_{m+1,j} = \big(\boldsymbol{v}_{m+1}\big)_j$$

and the likelihood along coordinate $j$ is $p(v_{m+1,j} \mid u)$, viewed as a function of the scalar $u$. RHF approximates this likelihood on each interior bin by a constant equal to the average of the likelihood at the bin end points, and on the tails by its value at the nearest particle:

$$\tilde{p}(v_{m+1,j} \mid u) = \sum_{k=1}^{K-1} \frac{1}{2}\Big(p(v_{m+1,j} \mid u^{(k)}) + p(v_{m+1,j} \mid u^{(k+1)})\Big)\chi_{[u^{(k)},u^{(k+1)}]}(u)$$
$$+\, p(v_{m+1,j} \mid u^{(1)})\, \chi_{(-\infty,u^{(1)}]}(u) + p(v_{m+1,j} \mid u^{(K)})\, \chi_{[u^{(K)},\infty)}(u). \tag{18}$$

This yields a *piecewise constant* approximation of the likelihood in $u$.

#### 2.2.3 Step 3: Marginal posterior via Bayes formula

Apply Bayes' theorem in this one-dimensional coordinate:

$$p(u \mid v_{m+1,j}) = \frac{\tilde{p}(v_{m+1,j} \mid u)\, p_{m+1|m,j}(u)}{\displaystyle\int_{\mathbb{R}} \tilde{p}(v_{m+1,j} \mid \tilde{u})\, p_{m+1|m,j}(\tilde{u})\, d\tilde{u}}. \tag{19}$$

Because both the prior marginal (17) and the approximate likelihood (18) are piecewise constant on the interior bins, the posterior $p(u \mid v_{m+1,j})$ is also piecewise constant there, with discontinuities only at the original particle locations $u^{(k)}$ and possibly at the tails.

#### 2.2.4 Step 4: Equal-mass posterior quantiles and new particles

The key design choice in RHF is to produce a new ensemble of equally weighted particles whose marginal distribution matches $p(u \mid v_{m+1,j})$ as well as possible.

Let $\tilde{u}^{(1)} < \tilde{u}^{(2)} < \cdots < \tilde{u}^{(K)}$ denote the *sorted* new particles in coordinate $j$. RHF enforces the equal-mass constraints

$$\int_{-\infty}^{\tilde{u}^{(1)}} p(u \mid v_{m+1,j})\, du = \frac{1}{K+1}, \tag{20}$$

$$\int_{\tilde{u}^{(k)}}^{\tilde{u}^{(k+1)}} p(u \mid v_{m+1,j})\, du = \frac{1}{K+1}, \quad k = 1, \dots, K-1. \tag{21}$$

That is, each of the $K+1$ intervals

$$(-\infty, \tilde{u}^{(1)}], \quad [\tilde{u}^{(1)}, \tilde{u}^{(2)}], \dots, [\tilde{u}^{(K)}, \infty)$$

carries mass $1/(K+1)$ under the marginal posterior.

Since $p(u \mid v_{m+1,j})$ is piecewise constant on each interior interval $[u^{(k)}, u^{(k+1)}]$, the integrals in (20)–(21) are *linear* functions of $\tilde{u}^{(k)}$ whenever $\tilde{u}^{(k)}$ lies inside a single bin. Thus the equations can be solved incrementally:

- Find $\tilde{u}^{(1)}$ by inverting the cumulative distribution function in the left tail; here one typically uses the Gaussian tail and a standard normal CDF.
- For $k = 1, \dots, K-1$, given $\tilde{u}^{(k)}$, find $\tilde{u}^{(k+1)}$ inside some interval $[u^{(r)}, u^{(r+1)}]$ so that (21) holds, which reduces to solving a linear equation.

In practice this is implemented by scanning through bins and accumulating mass until the desired quantile is reached.

### 2.2.5 Step 5: Undoing the sorting and reassembling the state

The construction so far is one-dimensional and uses the sorted order of the original particles in coordinate $j$. To recover $K$ *labeled* particles in $\mathbb{R}^N$:

1. Keep track of the permutation $\pi_j$ that sorts the prior particles in coordinate $j$:

$$u^{(\pi_j(1))} < u^{(\pi_j(2))} < \cdots < u^{(\pi_j(K))}.$$

2. Assign the new components by reversing the sorting:

$$u^{\text{new}}_{j,\pi_j(k)} := \tilde{u}^{(k)}, \quad k = 1, \ldots, K.$$

3. Repeat this procedure for all coordinates $j = 1, \ldots, N$ (or for a chosen set of low-dimensional projections).

After processing all coordinates, we obtain new particles

$$\boldsymbol{u}^{(k)}_{m+1|m+1}, \qquad k = 1, \ldots, K,$$

which are again *equally weighted*.

## 2.3 Algorithmic summary of RHF

For each analysis time $t_{m+1}$, the RHF proceeds as follows:

**Step 1. Forecast:** Propagate $\{\boldsymbol{u}^{(k)}_{m|m}\}$ to $\{\boldsymbol{u}^{(k)}_{m+1|m}\}$ by the dynamical model (deterministic or stochastic).

**Step 2. For each component** $j = 1, \ldots, N$**:**

- (a) Extract and sort the prior values $u^{(k)}_j$;
- (b) Construct the piecewise-constant marginal prior density $p_{m+1|m,j}(u)$ with Gaussian tails as in (17);
- (c) Approximate the likelihood $\tilde{p}(v_{m+1,j} \mid u)$ as in (18);
- (d) Form the marginal posterior $p(u \mid v_{m+1,j})$ by Bayes' theorem (19);
- (e) Compute equal-mass posterior quantiles $\tilde{u}^{(k)}$ by enforcing (20)–(21);
- (f) Undo the sorting to assign new components $u^{\text{new}}_{j,\pi_j(k)} = \tilde{u}^{(k)}$.

**Step 3. Reassemble particles:** The collection of updated components gives the new posterior ensemble $\{\boldsymbol{u}^{(k)}_{m+1|m+1}\}$ with equal weights $1/K$.

## 2.4 Remarks and practical issues

**Advantages.**

- RHF avoids weight degeneracy by construction: particles remain equally weighted after each update.

- The method is applicable to both stochastic and deterministic dynamical models, since resampling does not duplicate identical particles in state space in the same way as SIR.
- The marginal updates are one-dimensional and can be parallelized over coordinates and particles.
- For modest state dimension and small ensemble size, RHF can capture non-Gaussian posterior features better than Gaussian-based Kalman-type filters.

**Drawbacks.**

- RHF operates *marginally* in each coordinate; it does not directly enforce the correct multivariate posterior covariance or preserve dynamical balances among components. This may lead to dynamically unbalanced states.
- When all prior particles in a coordinate lie in the tails of the likelihood function, numerical issues arise in evaluating the denominator in (19) (dynamic range problem).
- Like other particle filters, RHF can still suffer from high-dimensionality issues if the state dimension is very large and the ensemble size is fixed and small.

**Relation to maximum entropy particle filter (MEPF).** The RHF uses discontinuous piecewise-constant marginal priors and likelihoods. The *maximum entropy particle filter* (MEPF) replaces these by smooth (e.g. cubic) densities that maximize entropy subject to moment constraints, improving robustness in dynamic-range-limited regimes. Conceptually, MEPF and RHF share the same philosophy: work with marginal distributions and equal-mass quantile updates, rather than importance weights.

## 2.5 Comparison with Kalman-type ensemble filters

We summarize the qualitative differences between particle filters (RHF) and Kalman-type filters (KF, EKF, EnKF, ETKF, EAKF).

| Method | Representation of uncertainty | Assumptions / regime |
|---|---|---|
| KF | Mean + covariance; exact Gaussian posterior | Linear dynamics and observations, Gaussian noises |
| EKF | Mean + covariance via local linearization | Mild nonlinearity; near-Gaussian posterior |
| EnKF | Ensemble of states; sample mean/cov | Approx. Gaussian posterior; works well for weak/moderate nonlinearity |
| ETKF/EAKF | EnKF + deterministic square-root transform | Same as EnKF but improved covariance control; small $K$ |
| RHF (particle filter) | Ensemble of equally weighted particles; marginal posterior via rank histograms | Non-Gaussian, nonlinear; high-dimensional but designed for small $K$ |

Key contrasts:

- **Non-Gaussianity.** Kalman-type filters are fundamentally Gaussian (or linearized Gaussian); RHF can represent skewness, heavy tails, and multi-modal marginals.
- **Update mechanism.** KF/EKF/EnKF/ETKF/EAKF use a linear update

$$\boldsymbol{u} \leftarrow \boldsymbol{u} + \boldsymbol{K}(\boldsymbol{v} - \boldsymbol{g}(\boldsymbol{u})),$$

  whereas RHF uses a nonlinear rearrangement of particle locations based on rank and posterior quantiles.
- **Weights.** Ensemble Kalman methods keep equal weights and update anomalies; SIR particle filters update weights; RHF keeps equal weights but updates positions via marginal posterior sampling.
- **High-dimensional behavior.** KF/EnKF/ETKF/EAKF are designed specifically to be feasible in very high dimension with small $K$ (thanks to Gaussian structure and localization). RHF partially addresses weight degeneracy but still faces challenges in very high-dimensional turbulent systems.

In practice, modern geophysical data assimilation systems often rely on EnKF/ETKF/EAKF as workhorses, potentially augmented by ideas from particle filters (e.g. marginal resampling, localized particle filters) in strongly nonlinear, non-Gaussian regimes. Particle filters like RHF and MEPF provide a bridge between fully general Bayesian filtering and computationally tractable ensemble Kalman methods.