IEEE Oceanic Engineering Society (OES)

Distinguished Lecture

University of New Orleans

New Orleans, LA

March 3, 2021

Ocean Acoustic Signal Processing:

A Bayesian Approach

James V. Candy

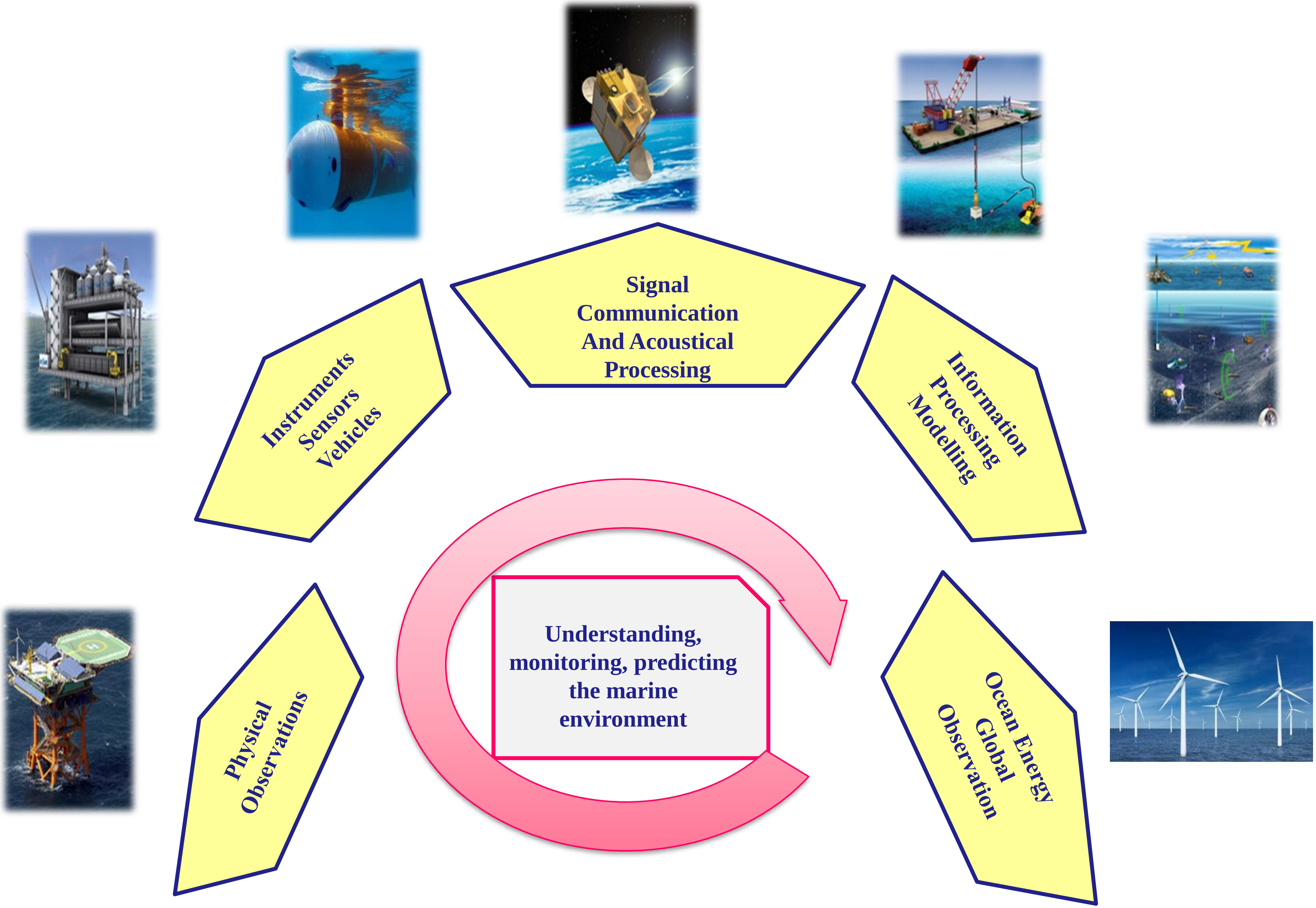
IEEE Fellow *ASA Fellow*

Lawrence Livermore National Laboratory
Chief Scientist for Engineering
Distinguished Member of Technical Staff

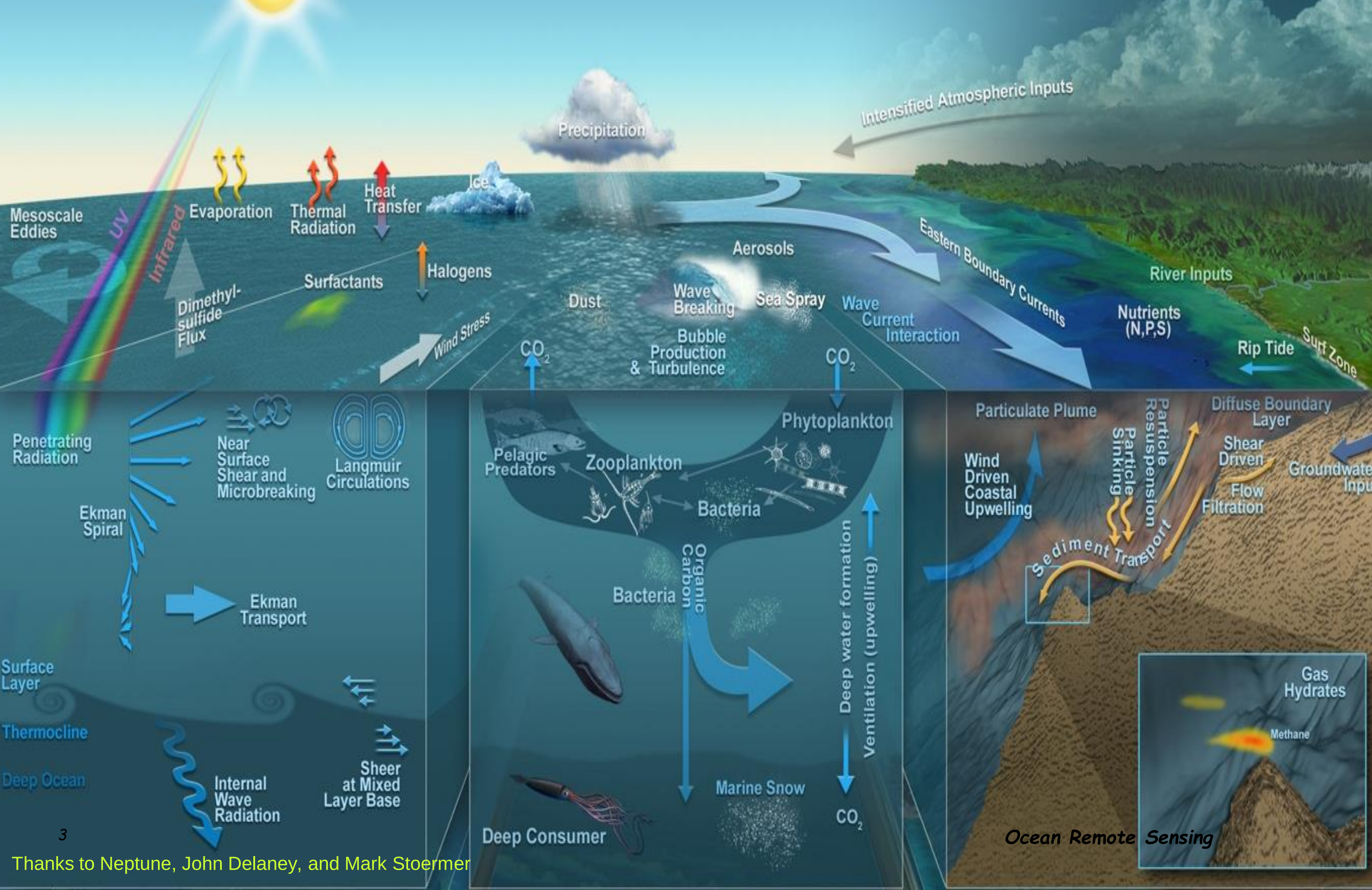
Adjunct Professor UC Santa Barbara

This work performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract DE-AC52-07NA27344 Document Reference: LLNL-PRES-748788 *2019*

IEEE OES


Signal
Communication
And Acoustical
Processing
Instruments
Sensors
Vehicles
Information
Processing
Modelling
Understanding,
monitoring, predicting
the marine
environment
Physical
Observations
Ocean Energy
Global
Observation

The Earth System is Complex

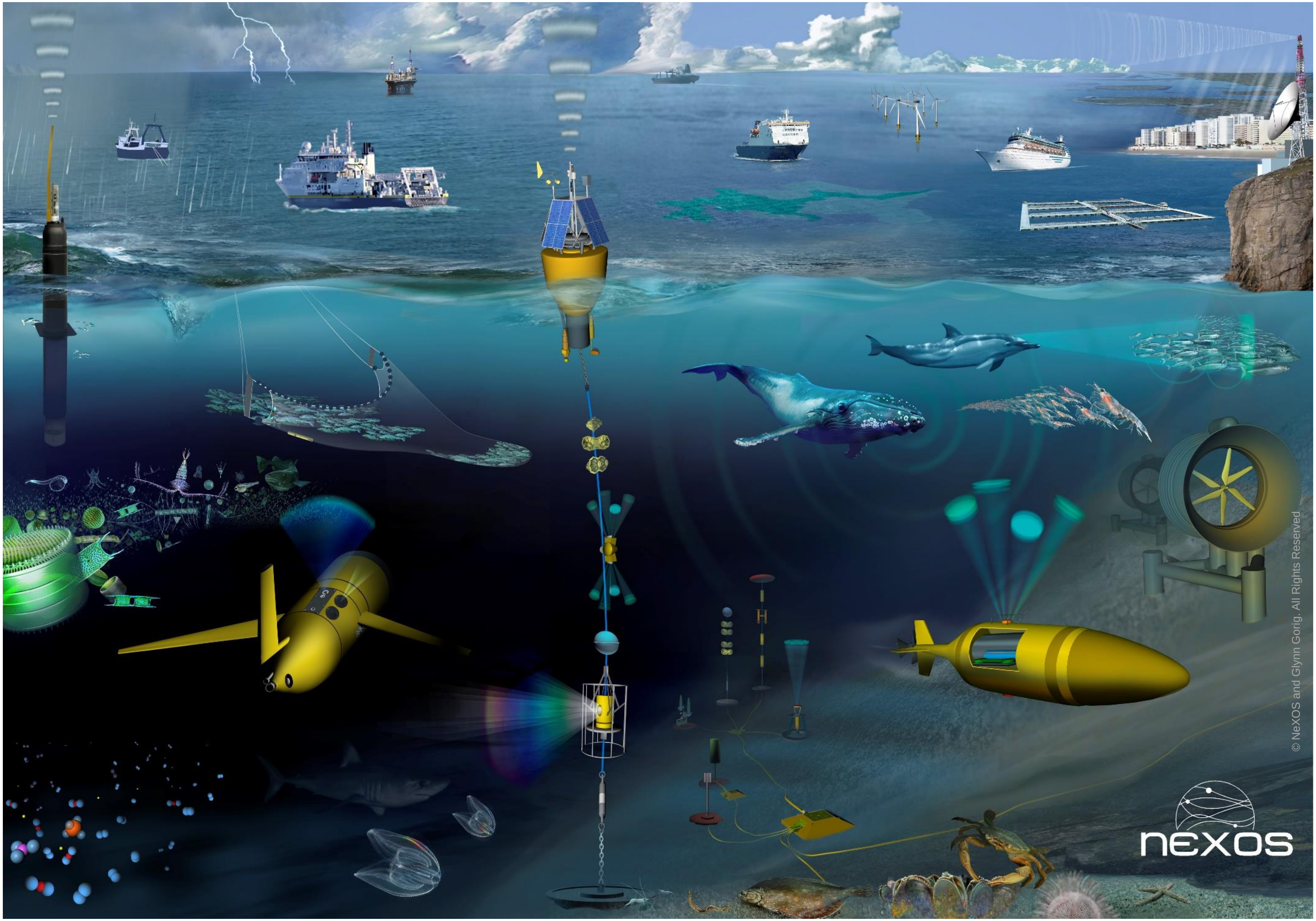


3

Ocean Remote Sensing

Thanks to Neptune, John Delaney, and Mark Stoermer

IEEE OES is involved in all aspects of Oceanic Engineering



ORGANIZATION

- INTRODUCTION
- BAYESIAN APPROACH
- SEQUENTIAL BAYESIAN PROCESSOR
- PARTICLE FILTERS (STATE-SPACE)
- OCEAN ACOUSTIC APPLICATION

BAYESIAN APPROACH

Bayesian Model-Based Signal Processing:

- Statistical signal processing is simply the processing of uncertain data
- When the underlying processes are non-linear and noise (uncertainty) is non-Gaussian, then a Bayesian approach enables a potential solution to the processing problem
- The incorporation of sophisticated mathematical models into the processor enables the extraction of the desired information
- Bayesian model-based signal processing is primarily concerned with the estimation of the underlying *posterior distribution* governing the problem incorporating physics-based mathematical models

Bayesian processing is based on PDF estimation using Bayes' rule, specifically, it:

- is concerned with the estimation of the underlying posterior probability distribution (of X) based on all of the data (Y) available

$$\boxed{\hat{\Pr}[X \mid Y]}$$

- Applies Bayes' rule to perform the posterior estimation:

$$\Pr\left[X \mid Y\right] = \frac{\Pr\left[Y \mid X\right] \times \Pr\left[X\right]}{\Pr\left[Y\right]}$$

- extracts statistics from the posterior (inference) to solve a variety of problems (signal enhancement, detection, parameter estimation, etc.).

- enables estimates like the conditional mean which is simply performed by:

$$\hat{\mathrm{X}} := \mathrm{E}\left\{X \mid Y\right\} \quad \Rightarrow \quad \hat{\mathrm{X}} = \int X \times \hat{\Pr}[X \mid Y]\, dX$$

Bayesian techniques use BAYES' RULE:

Posterior

Likelihood

Prior

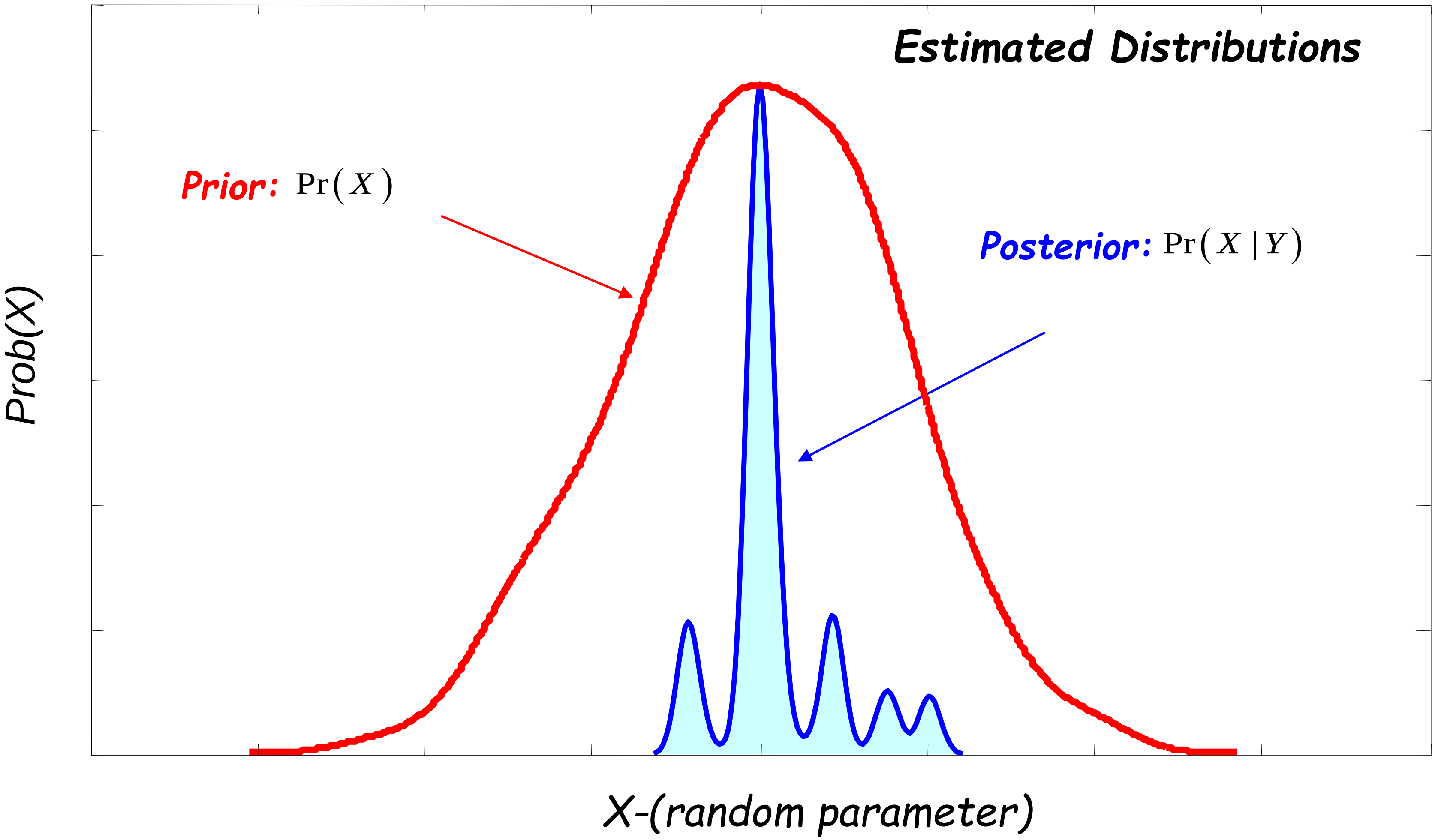
Bayes' Rule:

$$\Pr\left[X \mid Y = y\right] = \frac{\Pr[Y \mid X = x] \times \Pr[X = x]}{\Pr[Y = y]}$$

Evidence/Normalization

This "simple relationship" is the principal foundation of Bayesian signal processing both theoretically (derivations) and pragmatically (implementations) relying on numerical integration or Monte Carlo sampling techniques

Bayesian techniques can be thought of as converting the prior → posterior



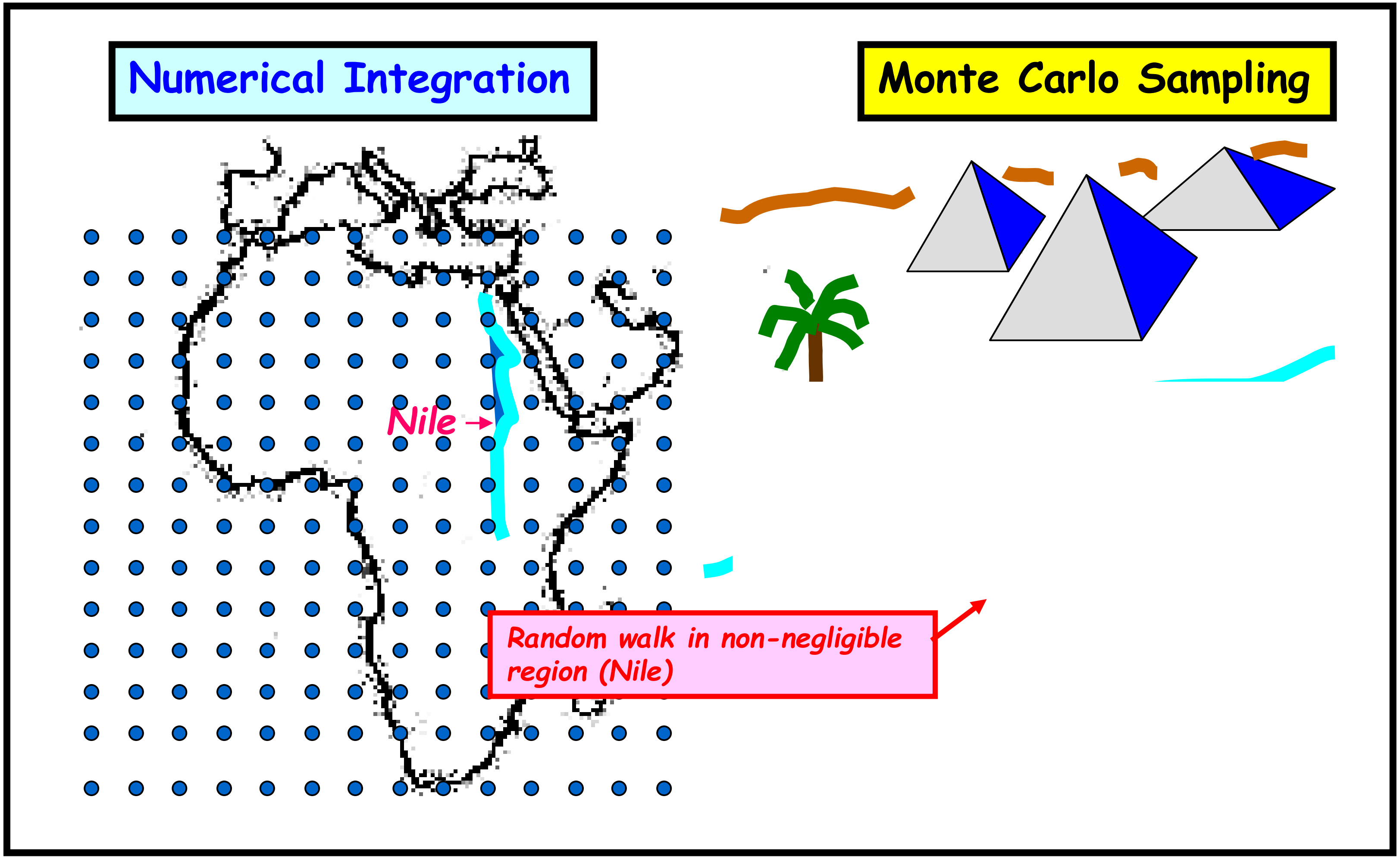
The Bayesian approach to signal processing problem solving represents:

- an alternative simulation-based numerical approach to finding solutions to complex mathematical problems that cannot easily be solved otherwise
- a powerful means for generating random samples used in estimating "posterior" probability distributions required for statistical estimation and therefore signal processing
- a class of stochastic MONTE CARLO (MC) computations to simulate the dynamics of a physical or mathematical systems capturing their inherent uncertainties
- a set of MC techniques that have "recently" evolved in the signal processing area and are high interest especially in Bayesian model-based processor (BMBP) problems

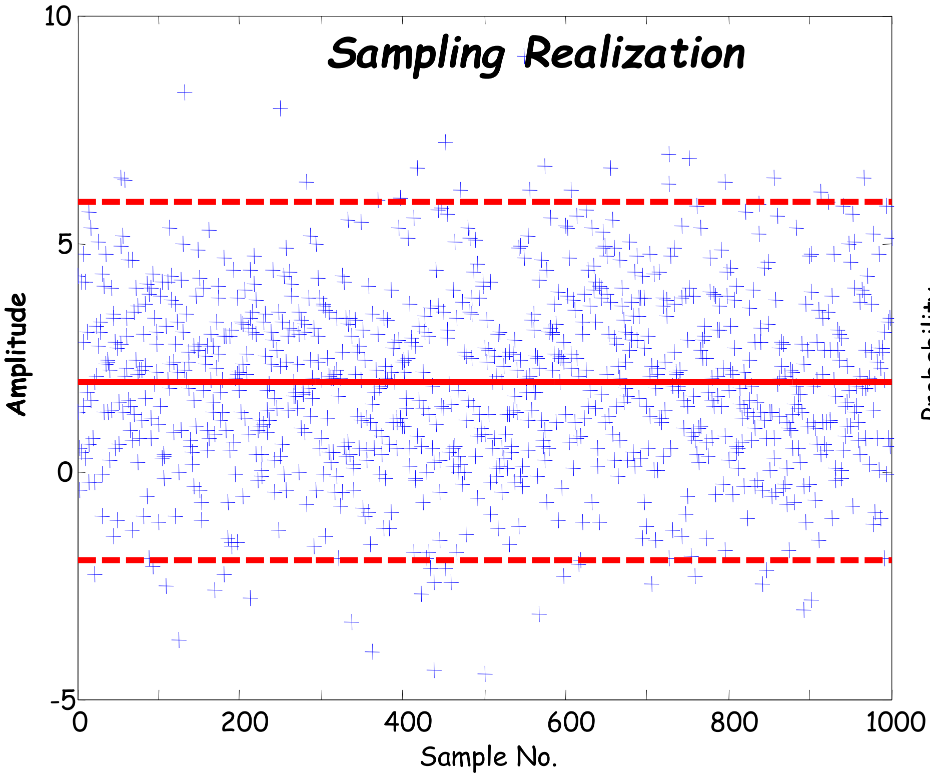
The Monte Carlo method provides the foundation for "simulation-based" Bayesian signal processing

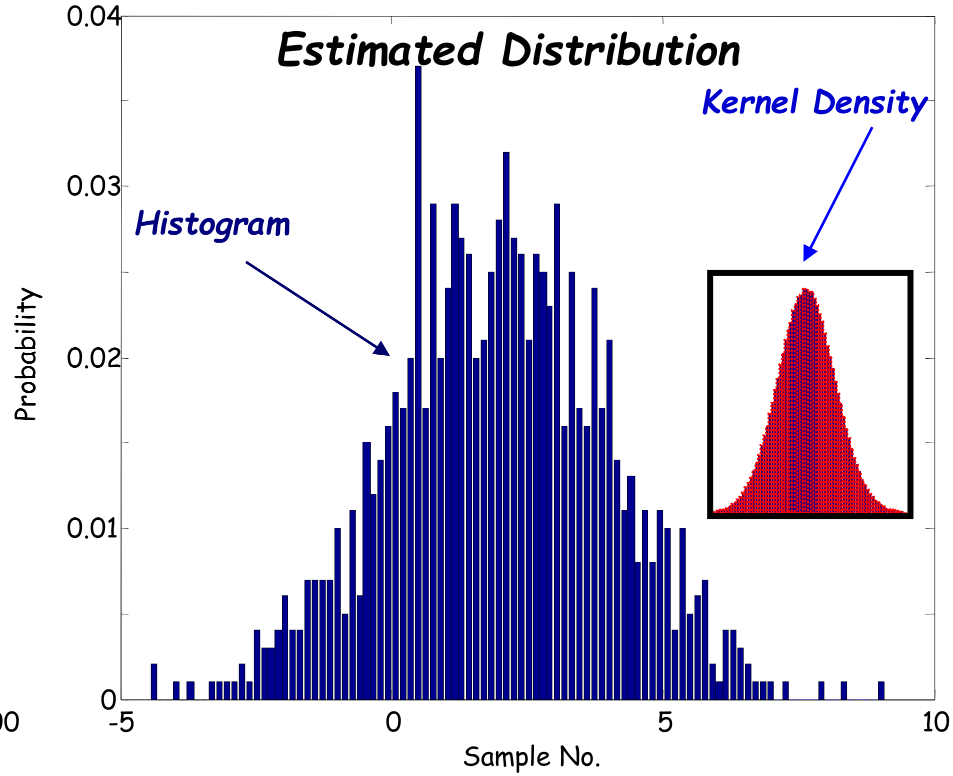
- The MC method is a stochastic computational technique capable of efficiently simulating complex systems
- MC method evolved in the mid-1940's
- It was conceived by Fermi (1930) and Ulam (1945) with the advent of ENIAC computer, coined and developed by Metropolis, Ulam and von Neumann (1947)
- It has been applied in many areas: computational physics and biology, chemistry, mathematics, engineering, materials and finance to name a few
- It solves problems in simulation, integration, optimization, inversion and learning

"Monte Carlo sampling" provides an approximate method of integration (e.g. average Nile depth) [Frenkel '04]

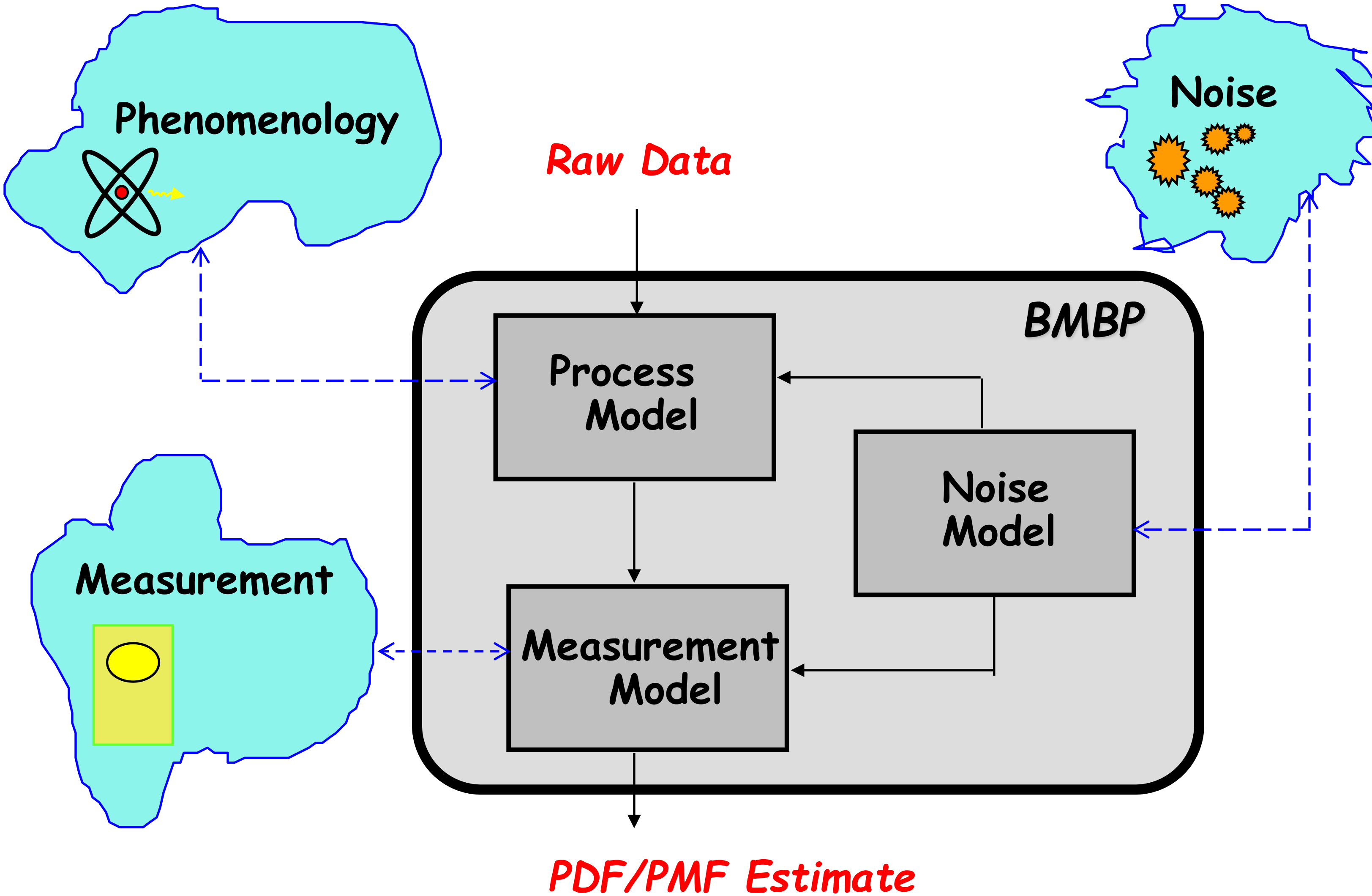


<u>Monte Carlo methods</u> rely on samples generated from the sampling distribution to estimate statistics (mean, variance, etc.) as demonstrated in this Gaussian example

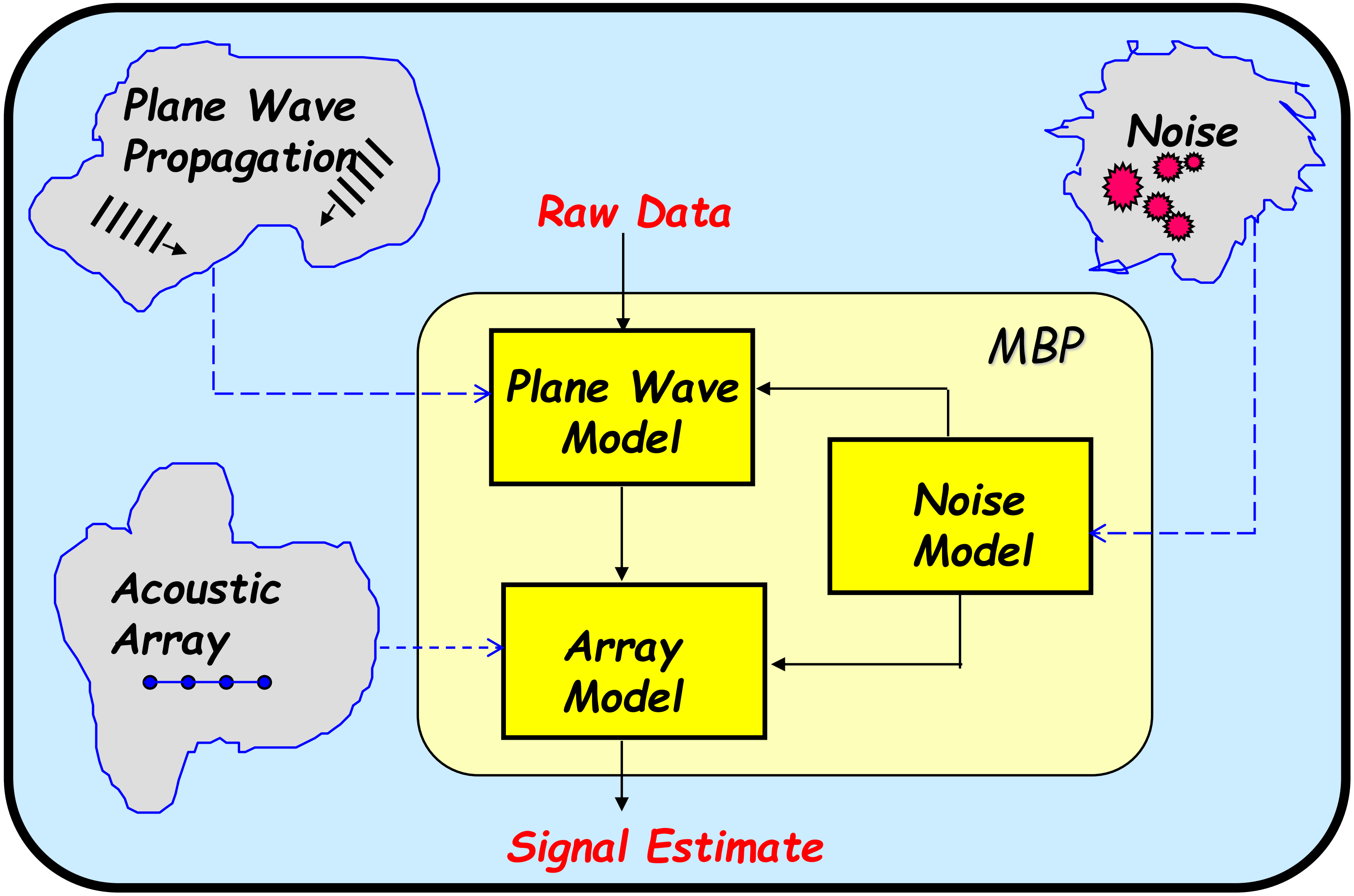




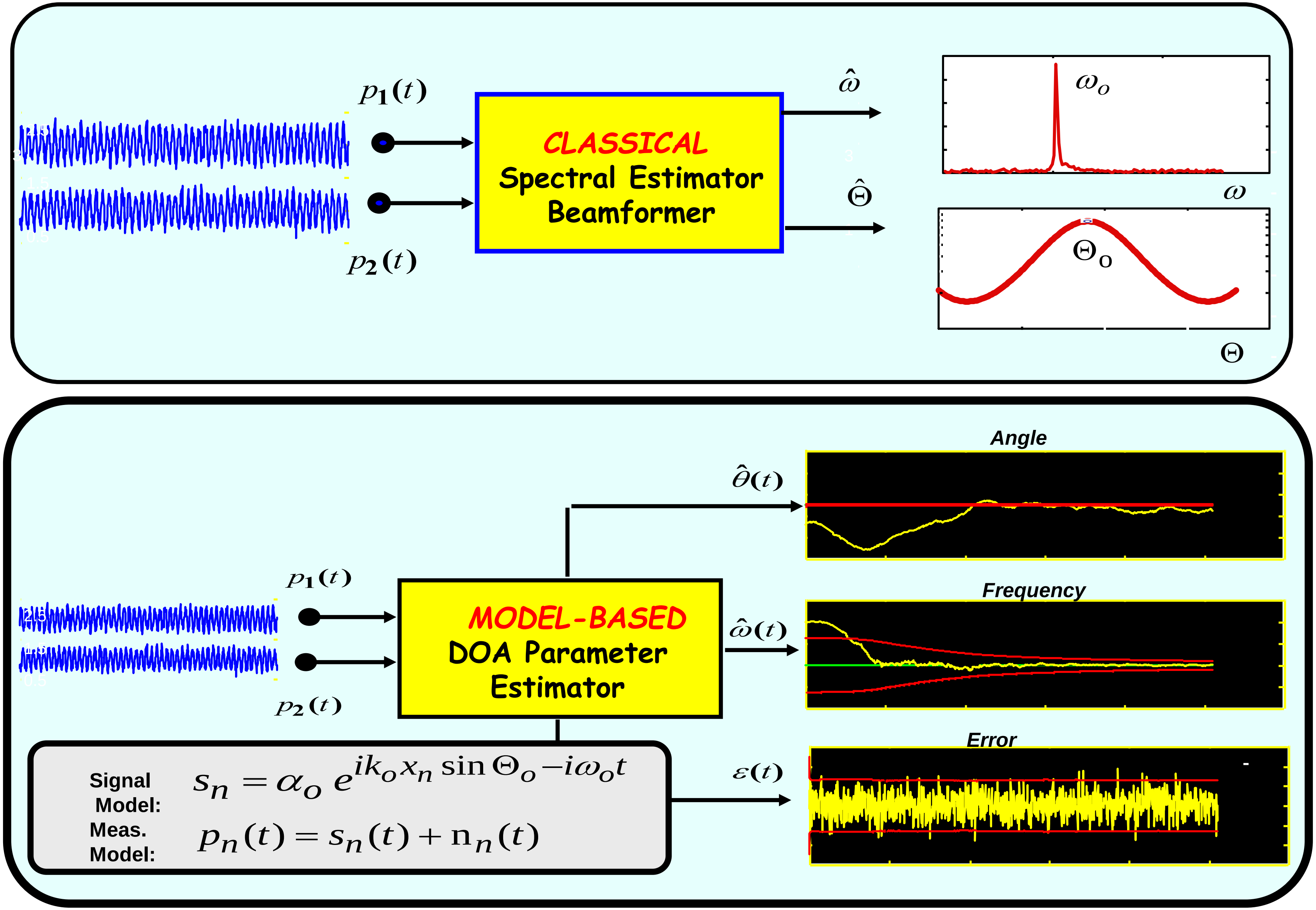
Bayesian Model-Based Processing (BMBP) techniques incorporate "a priori" knowledge of the phenomenology into a processing scheme to estimate the posterior distribution and extract the desired signal



The model-based approach to the signal (plane wave) enhancement and estimation problems can be cast as:



PW-DOA Est.— Classical vs. Model-Based



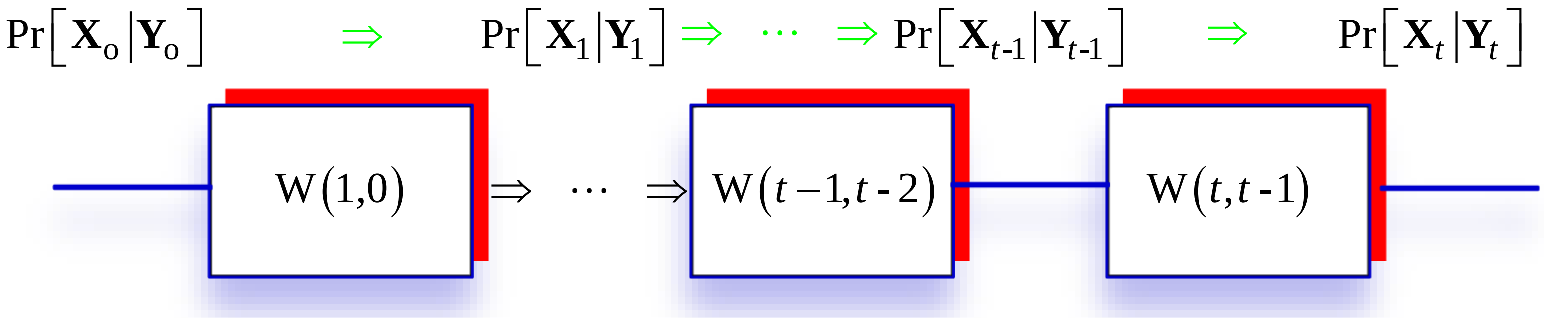
SEQUENTIAL BAYESIAN PROCESSORS

(PARTICLE FILTERS)

SEQUENTIAL BAYESIAN PROCESSING: Particle Filters (PFs):

- Monte Carlo techniques obtain random sample-based representations of the posterior probability distributions
- When "real-time" operations are required or the underlying statistics are "nonstationary", then sequential MC methods must be employed to solve the problem
- A PF is a "sequential" (Monte Carlo) technique in which the underlying posterior distribution of interest is characterized by a set or "cloud" of random samples (particles)
- In this sense, a PF is a non-parametric representation of the posterior in discrete form (probability mass function)
- PFs represent the next generation of "processors" that are not constrained to linear models or Gaussian distributions

The posterior distribution can be estimated using the sequential Bayesian processor (SBP):



$$\Pr[\mathbf{X}_t|\mathbf{Y}_t] = W(t,t-1) \times \Pr[\mathbf{X}_{t-1}|\mathbf{Y}_{t-1}]$$

where the Bayes' operator is defined at each stage by

$$\boxed{W(t,t-1) := \frac{\Pr[\mathbf{y}(t)|\mathbf{x}(t)] \times \Pr[\mathbf{x}(t)|\mathbf{x}(t-1)]}{\Pr[\mathbf{y}(t)|\mathbf{Y}_{t-1}]}; \quad t = 1, \cdots, N}$$

A PF is a "sequential" (Monte Carlo) technique in which the underlying posterior distribution of interest is characterized by a set or "cloud" of random samples---the particles

IT is an algorithm that (sequentially) propagates and updates the random samples (particles) drawn from the previous stage to obtain a set of samples approximately distributed from the next stage

$$\Pr\left[X_t \middle| \mathbf{Y}_t\right] = W_i\left(t, t-1; X_i(t)\right) \times \Pr\left[X_{t-1} \middle| \mathbf{Y}_{t-1}\right]$$

i-th Particle

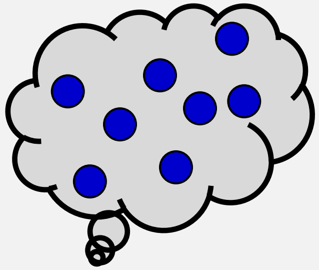
where $W_i\left(t, t-1; X_i(t)\right)$ is the weight (Bayes' operator) and

$X_i(t)$ is the i^{th} – particle at stage (time) t

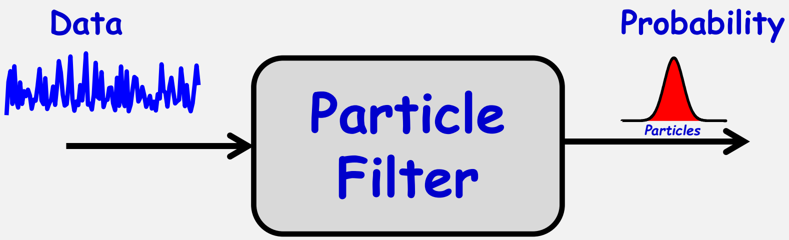
Next Stage

Previous Stage

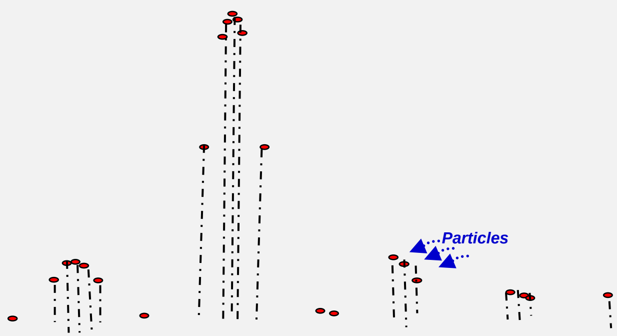
A "particle" is a random sample



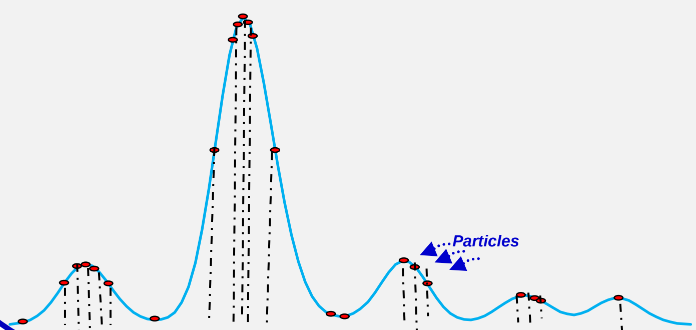
A "swarm" or "cloud" is a group of particles



A particle filter is a processor that has data on input and estimates the "posterior distribution" on output

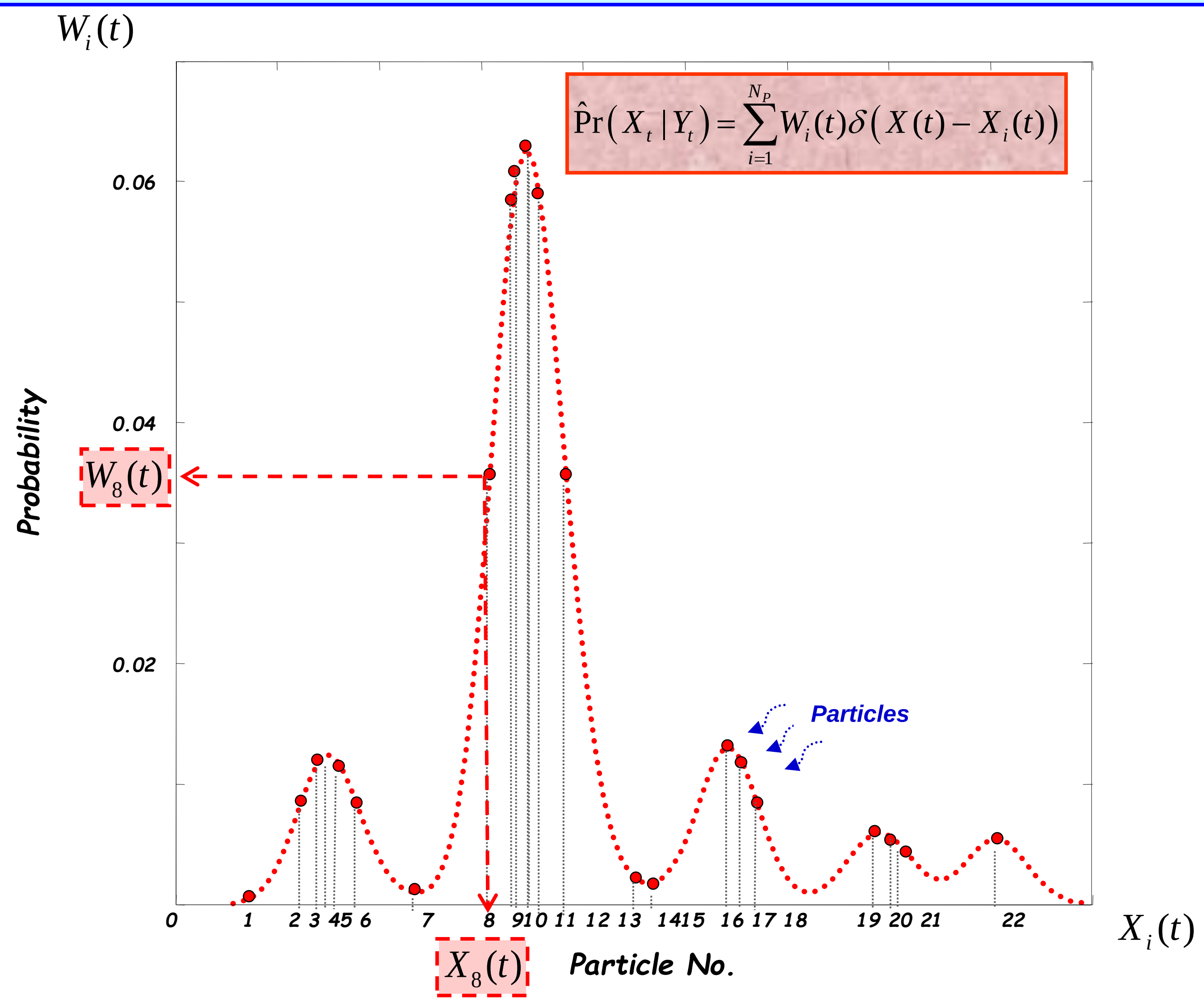


The particles are the "location" parameters along with their associated weights that gather in "highest probability regions" to provide a non-parametric estimate of the empirical posterior distribution

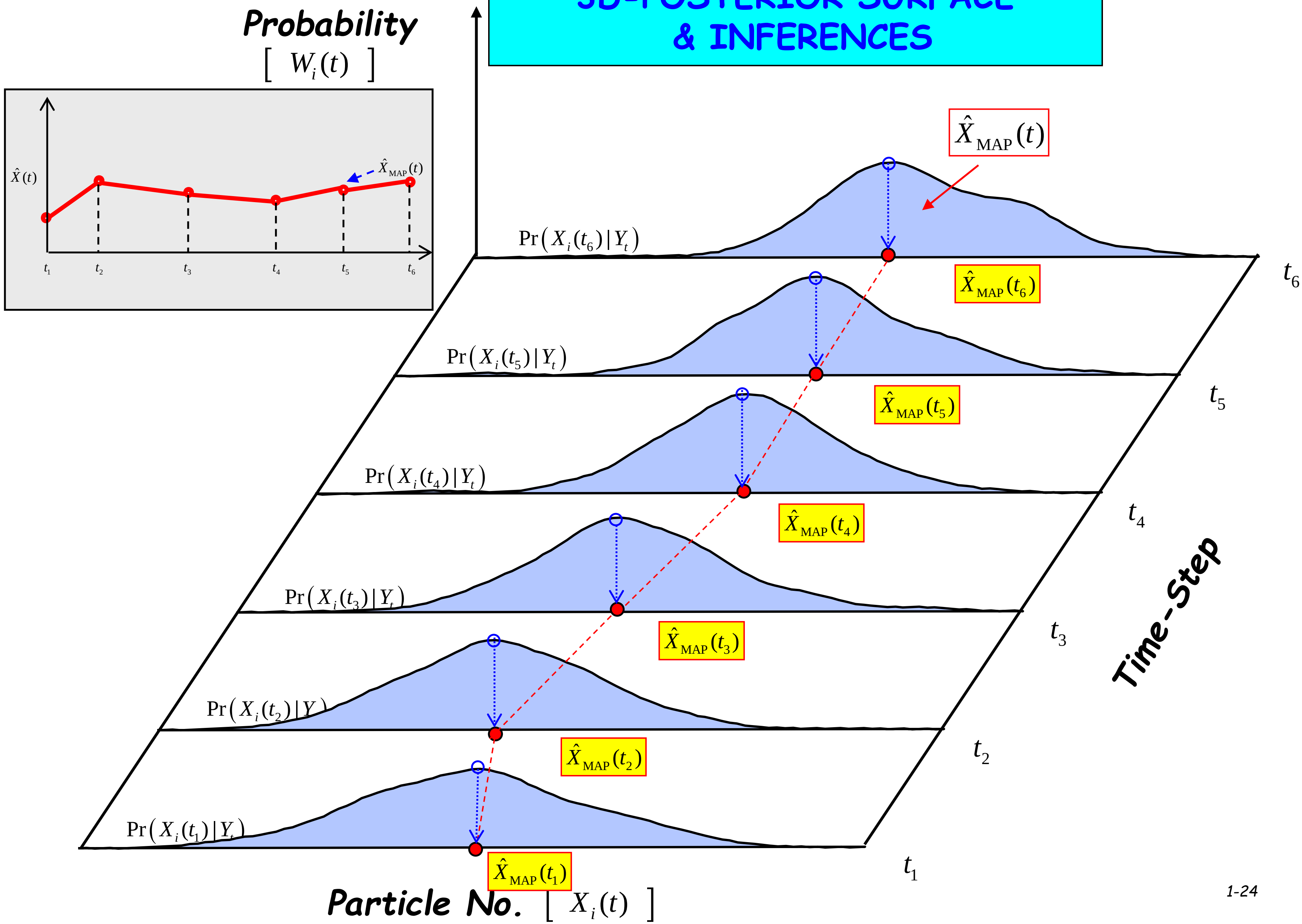


The resulting "posterior distribution" is observed through probability mass function estimation (histogram, kernel density)

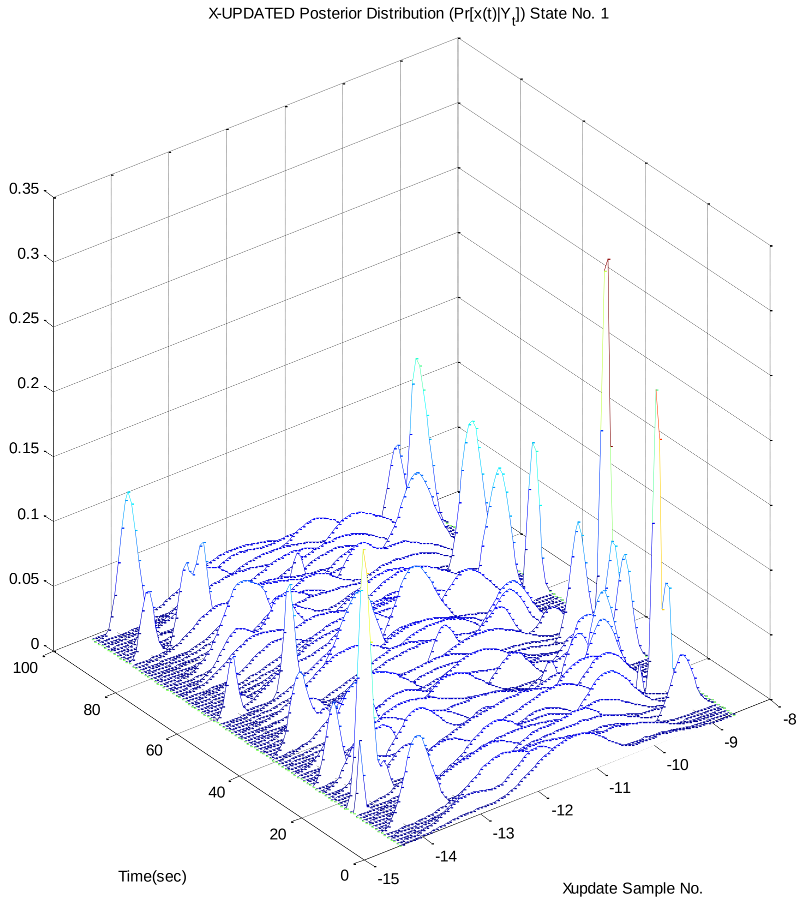
PARTICLE FILTERS are *sequential* MC techniques in which the underlying *posterior distribution* of interest is represented by a "cloud" of random samples (particles) in the state/parameter space

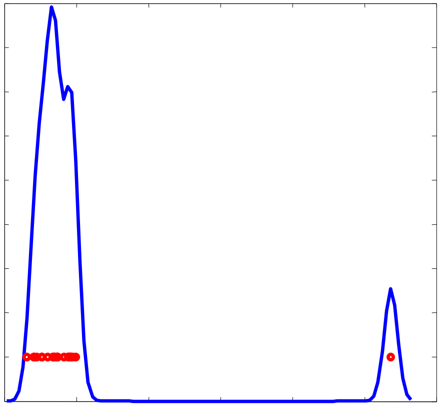


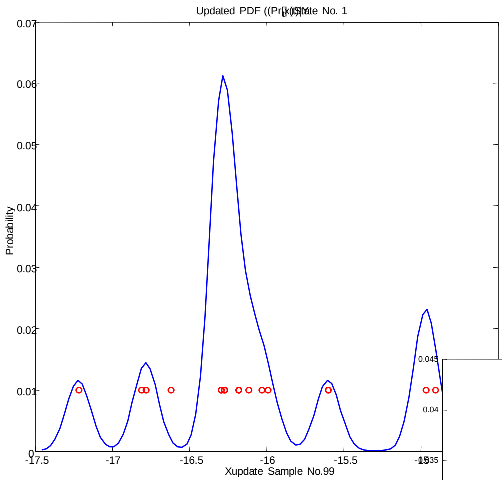
DYNAMIC PARTICLES LEAD TO A 3D-POSTERIOR SURFACE & INFERENCES

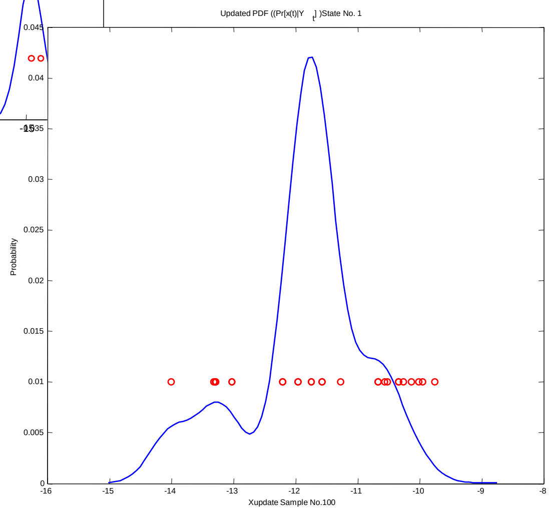


The PF provides a 3D-posterior in the dynamic case especially for multi-modal problems









Particle filters have applicability in many areas:

- **Signal processing**
  - Image processing and segmentation
  - Model selection
  - Tracking and navigation
- **Communications**
  - Channel estimation
  - Blind equalization
  - Positioning in wireless networks

- **Applications**
  - Biology & Biochemistry
  - Chemistry
  - Economics & Business
  - Genomics
  - Geosciences
  - Immunology
  - Materials Science
  - Physics/Optics
  - Pharmacology & Toxicology
  - Psychiatry/Psychology
  - Social Sciences

PFs: advantages and disadvantages:

ADVANTAGES

- Ability to represent arbitrary densities
- Adaptive focusing on highly probable regions
- Dealing with multi-modal PDFs (non-Gaussian) noise
- The framework also enables the inclusion of multiple models

DISADVANTAGES

- High computational complexity
- It is difficult to determine optimal number of particles
- Number of particles increase with increasing model dimension
- Potential problems: degeneracy and loss of diversity
- The choice of proposal density is crucial

SEQUENTIAL BAYESIAN

STATE-SPACE

PROCESSORS

Bayesian approach to the state-space: definitions

$$\begin{aligned} x(t) &= \mathcal{A}\left(x(t-1), u(t-1), w(t-1)\right) \\ y(t) &= \mathcal{C}\left(x(t), u(t), v(t)\right) \end{aligned}$$

where w and v are the respective process and measurement noise sources with u a known input. Here $\mathcal{A}(\cdot)$ is the nonlinear (or linear) dynamic state transition function and $\mathcal{C}(\cdot)$ the corresponding measurement function. Both conditional probabilistic distributions embedded within the Bayesian framework are *completely* specified by these functions and the underlying noise distributions: $\Pr(w(t-1))$ and $\Pr(v(t))$. That is, we have the equivalence

$$\begin{aligned} \mathcal{A}\left(x(t-1), u(t-1), w(t-1)\right) &\Rightarrow \Pr\left(x(t)|x(t-1)\right) \Leftrightarrow \mathcal{A}\left(x(t)|x(t-1)\right) \\ \mathcal{C}\left(x(t), u(t), v(t)\right) &\Rightarrow \Pr\left(y(t)|x(t)\right) \Leftrightarrow \mathcal{C}\left(y(t)|x(t)\right) \end{aligned}$$

Bayesian approach to state-space: posteriors

the prediction recursion characterized by the *Chapman-Kolmogorov equation* replacing transition probability with the implied model-based conditional, that is,

$$\Pr(x(t)|Y_{t-1}) = \int \overbrace{\mathcal{A}(x(t)|x(t-1))}^{\text{Embedded Process Model}} \times \Pr(x(t-1)|Y_{t-1})dx(t-1)$$

Next we incorporate the model-based likelihood into the posterior equation with the understanding that the process model has been incorporated into the prediction

$$\Pr(x(t)|Y_t) = \overbrace{\mathcal{C}(y(t)|x(t))}^{\text{Embedded Measurement Model}} \times \Pr(x(t)|Y_{t-1}) \Big/ \Pr(y(t)|Y_{t-1})$$

State-space Bayesian processors based on sequential importance samplers follow easily as:

$$W(t) = W(t-1) \times \frac{\Pr(y(t)|x(t)) \times \Pr(x(t)|x(t-1))}{q(x(t)|X_{t-1}, Y_t)}$$

Now let us recall the general state-space characterization representing the transition and likelihood probabilities as:

$$\begin{aligned} \Pr\left(x(t)|x(t-1)\right) &\Leftrightarrow \mathcal{A}\left(x(t)|x(t-1)\right) \\ \Pr\left(y(t)|x(t)\right) &\Leftrightarrow \mathcal{C}\left(y(t)|x(t)\right) \end{aligned}$$

Assuming this is true, then the *SSPF* recursion becomes

$$\begin{aligned} x_i(t) &\sim q(x(t)|x(t-1), y(t)) \\ W_i(t) &= W_i(t-1) \times \boxed{\frac{\mathcal{C}(y(t)|x_i(t)) \times \mathcal{A}(x_i(t)|x_i(t-1))}{q(x_i(t)|x_i(t-1), y(t))}} \\ \mathcal{W}_i(t) &= \frac{W_i(t)}{\sum_{i=1}^{N_p} W_i(t)} \end{aligned}$$

and the filtering *posterior* is estimated by

$$\hat{\Pr}(x(t)|Y_t) \approx \sum_{i=1}^{N_p} \mathcal{W}_i(t) \times \delta(x(t) - x_i(t))$$

Note that as N_p becomes large, in the limit, we have

$$\lim_{N_p \to \infty} \hat{\Pr}(x(t)|Y_t) \longrightarrow \Pr(x(t)|Y_t)$$

The "generic" state-space particle filtering method is given by:

<u>INITIALIZE</u>:

$$x_i(0) \longrightarrow \Pr(x(0)); \qquad W_i(0) = \frac{1}{N_p}; \quad i = 1, \cdots, N_p \qquad \text{[sample]}$$

<u>IMPORTANCE SAMPLING</u>:

$$x_i(t) \sim \mathcal{A}\left(x(t)|x_i(t-1)\right) \qquad \text{[state transition]}$$

State-space transition model

$$\mathcal{A}\left(x(t)|x_i(t-1)\right) \equiv A\left(x(t-1), u(t-1), w_i(t-1)\right); \quad w_i \sim \Pr(w_i(t)) \qquad \text{[transition]}$$

Weight Update:

$$\boxed{W_i(t) = W_i(t-1) \times \frac{\mathcal{C}\left(y(t)|x_i(t)\right) \times \mathcal{A}\left(x(t)|x_i(t-1)\right)}{q\left(x(t)|x(t-1), y(t)\right)} \qquad \text{[weights]}}$$

Measurement likelihood model

$$\mathcal{C}\left(y(t)|x_i(t)\right) \equiv C\left(x(t), u(t), v(t)\right); \quad v_i \sim \Pr(v(t)) \qquad \text{[likelihood]}$$

Weight normalization

$$\mathcal{W}_i(t) = \frac{W_i(t)}{\sum_{i=1}^{N_p} W_i(t)}$$

<u>DISTRIBUTION</u>:

$$\hat{\Pr}(x(t)|Y_t) \approx \sum_{i=1}^{N_p} \mathcal{W}_i(t)\delta(x(t) - x_i(t)) \qquad \text{[posterior distribution]}$$

Importance distributions provide the key: "transition prior" (Gordon et. al. '93)

Another choice for an importance distribution is the *transition prior*. This *prior* is defined in terms of the state-space representation by $\mathcal{A}(x(t)|x(t-1)) \rightarrow \mathcal{A}(x(t-1), u(t-1), w(t-1))$ which is dependent on the known excitation and process noise statistics. It is given by

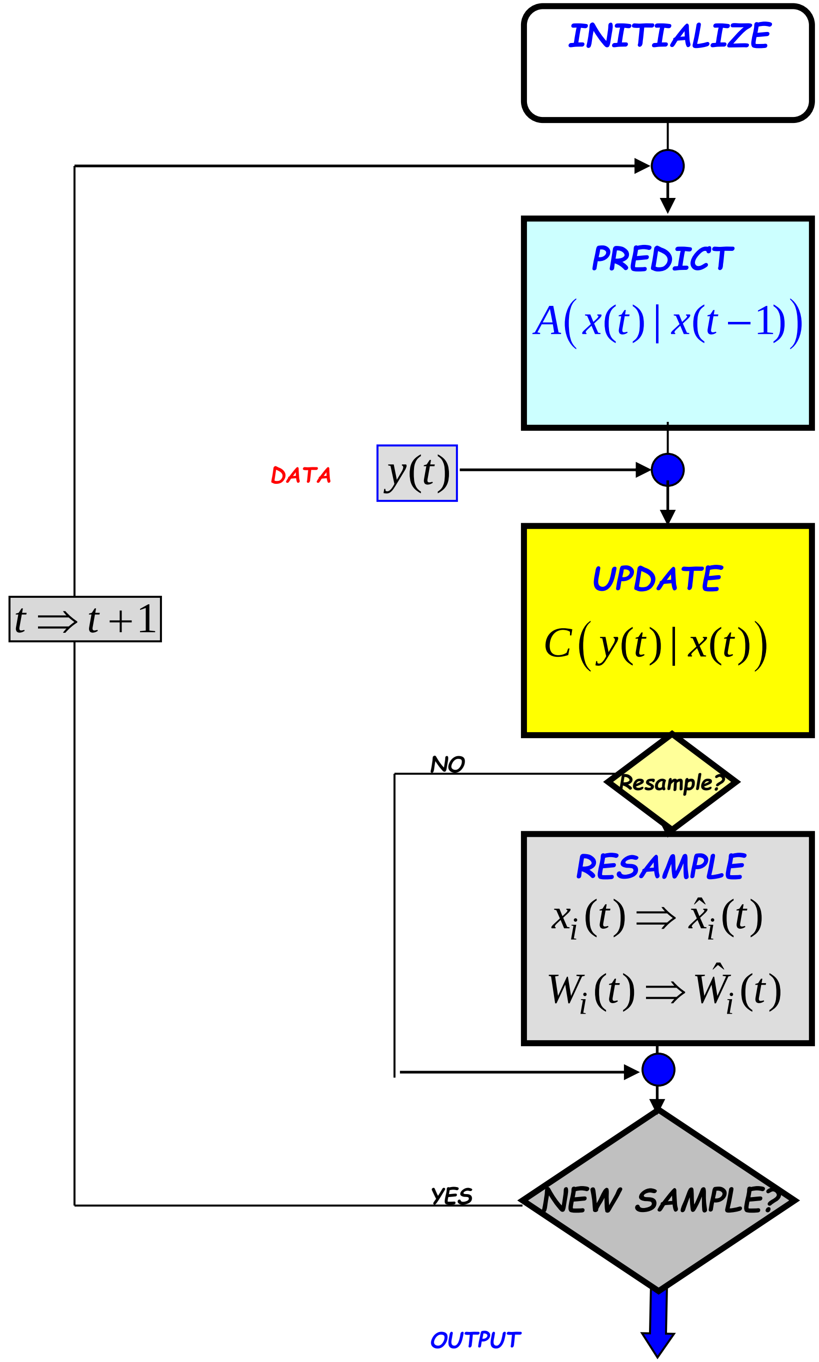
$$\boxed{q_{prior}(x(t)|x(t-1), Y_t) \longrightarrow \Pr(x(t)|x(t-1))}$$

Substituting this choice into the weights gives

$$W_i(t) = W_i(t-1) \times \frac{\Pr(y(t)|x_i(t)) \times \Pr(x(t)|x_i(t-1))}{q_{prior}(x(t)|x_i(t-1), Y_t)} = \boxed{W_i(t-1) \times \Pr(y(t)|x_i(t))}$$

BOOTSTRAP ESTIMATOR

STATE-SPACE SIR ALGORITHM



$\left\{\hat{x}_i(t-1), \hat{W}_i(t-1)\right\} \Rightarrow \Pr\left[x(t-1)\middle|Y_{t-1}\right]$

$\left\{\hat{x}_i(t), \hat{W}_i(t)\right\} \Rightarrow \Pr\left[x(t)\middle|Y_{t-1}\right]$

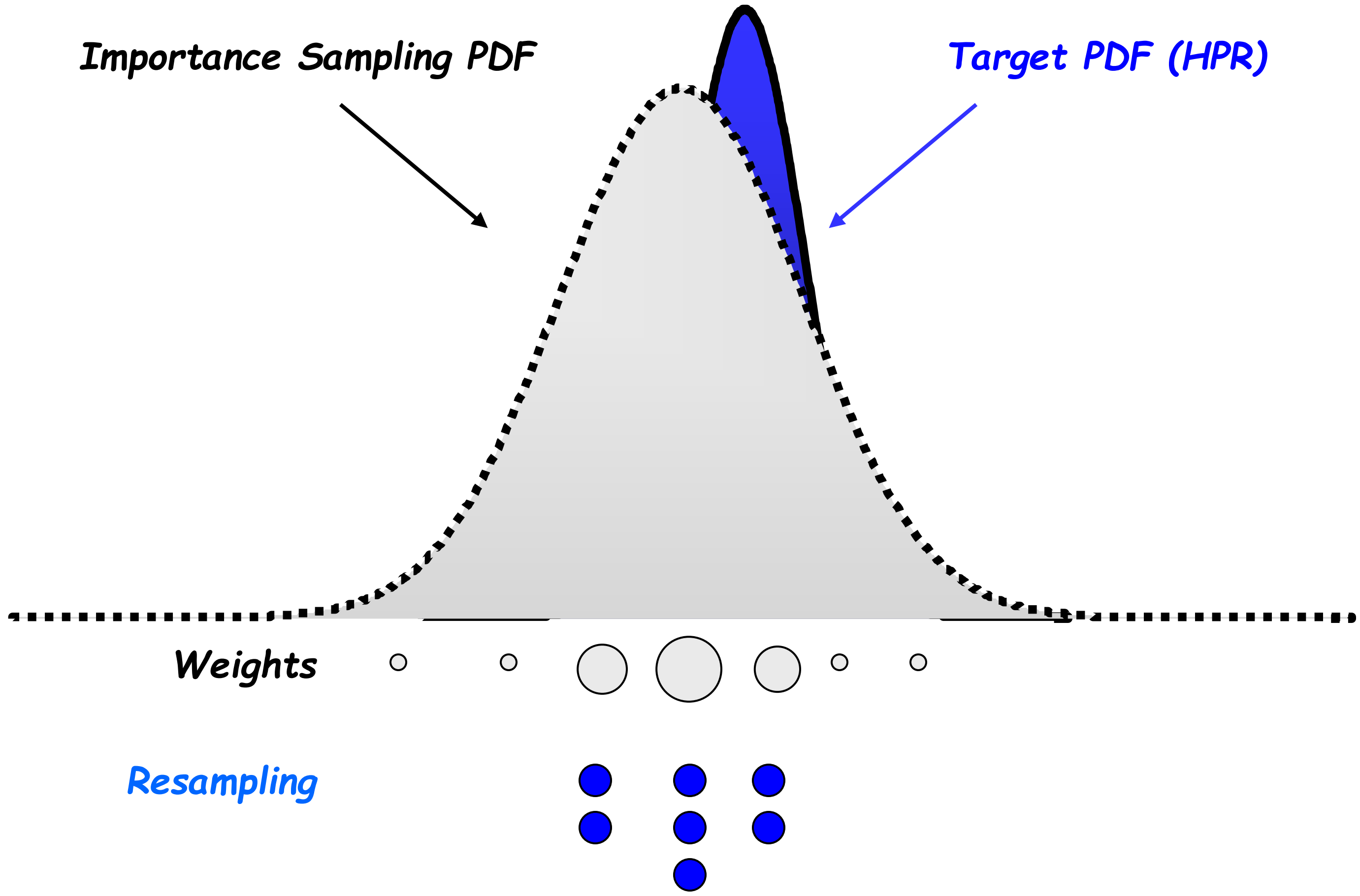
$\left\{x_i(t), W_i(t)\right\} \Rightarrow \Pr\left[x(t)\middle|Y_t\right]$

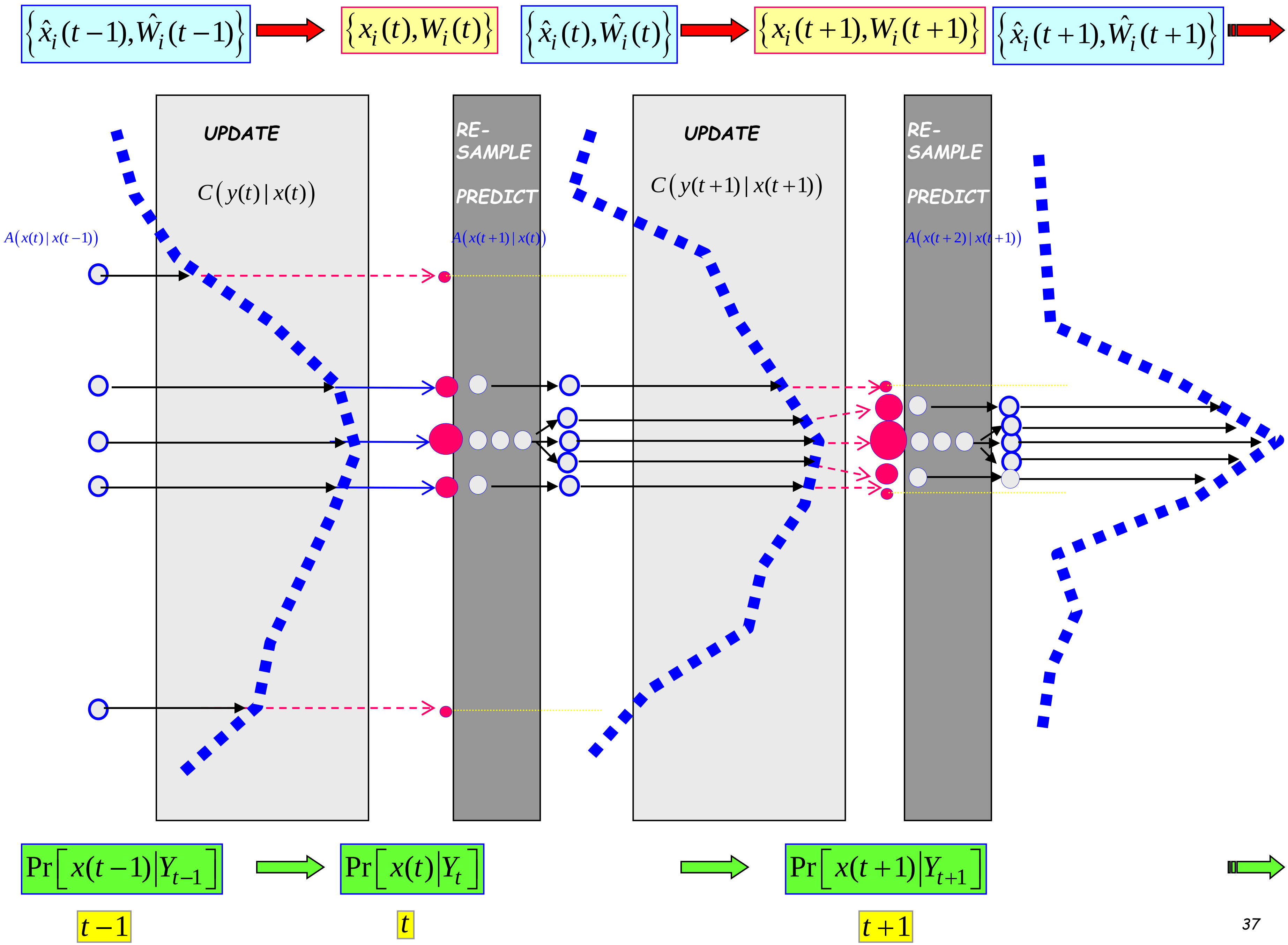
$\left\{x_i(t), W_i(t)\right\} \Rightarrow \Pr\left[x(t)\middle|Y_t\right]$

PROBLEM: Particles deplete in number (degenerate) to a single particle due to the increased variance in each step; therefore,

- The particles must be "rejuvenated" or equivalently resampled
- Resampling inhibits the depletion problem, but increases the uncertainty (weight variance)
- If not implemented properly, it can also increase computational time extensively (non-parallel)
- Resampling is essentially a process that attempts to preserve particles with large weights (acceptance probabilities) while discarding those with small weights.

Resampling is accomplished by a variety of techniques all with the same purpose: to generate more particles in the high probability regions (large weights) and remove the particles with small weights




$\left\{\hat{x}_i(t-1),\hat{W}_i(t-1)\right\}$
$\left\{x_i(t),W_i(t)\right\}$
$\left\{\hat{x}_i(t),\hat{W}_i(t)\right\}$
$\left\{x_i(t+1),W_i(t+1)\right\}$
$\left\{\hat{x}_i(t+1),\hat{W}_i(t+1)\right\}$
UPDATE
$C\left(y(t)\,|\,x(t)\right)$
RE-SAMPLE
PREDICT
$A\left(x(t+1)\,|\,x(t)\right)$
UPDATE
$C\left(y(t+1)\,|\,x(t+1)\right)$
RE-SAMPLE
PREDICT
$A\left(x(t+2)\,|\,x(t+1)\right)$
$A\left(x(t)\,|\,x(t-1)\right)$
$\Pr\left[x(t-1)\middle|Y_{t-1}\right]$
$\Pr\left[x(t)\middle|Y_t\right]$
$\Pr\left[x(t+1)\middle|Y_{t+1}\right]$
$t-1$
t
$t+1$

BOOTSTRAP PF ALGORITHM:

INITIALIZE:

$$x_i(0) \sim \Pr(x(0)) \qquad W_i(0) = \frac{1}{N_p} \quad i = 1, \cdots, N_p \qquad \text{[sample]}$$

IMPORTANCE SAMPLING:

$$x_i(t) \sim \mathcal{A}\left(x(t)|x_i(t-1)\right) \rightarrow A\left(x(t-1), u(t-1), w_i(t-1)\right); \quad w_i \sim \Pr(w_i(t)) \qquad \text{[state transition]}$$

Weight Update

$$W_i(t) = \boxed{\mathcal{C}\left(y(t)|x_i(t)\right)} \rightarrow C\left(x(t), u(t), v(t)\right); \quad v \sim \Pr(v(t)) \qquad \text{[weights/likelihood]}$$

Weight normalization

$$\mathcal{W}_i(t) = \frac{W_i(t)}{\sum_{i=1}^{N_p} W_i(t)}$$

RESAMPLING:

$$\hat{x}_i(t) \Rightarrow x_i(t)$$

DISTRIBUTION:

$$\hat{\Pr}(x(t)|Y_t) \approx \sum_{i=1}^{N_p} \mathcal{W}_i(t)\delta(x(t) - \hat{x}_i(t)) \qquad \text{[posterior distribution]}$$

For the PF problem, the Kullback-Leibler divergence metric is:

For our problem suppose $\Pr(\mathbf{x}(t)|Y_t)$ is the *true* posterior *PMF* and $\hat{\Pr}(\mathbf{x}_i(t)|Y_t)$ is the estimated (particle) distribution, then the *Kullback-Leibler (KL)* information quantity of the true distribution relative to the estimated is defined by

$$\mathcal{I}_{KL}\left(\Pr(\mathbf{x}(t)|Y_t);\hat{\Pr}(\mathbf{x}_i)|Y_t)\right) := E_{\mathbf{x}}\left\{\ln\frac{\Pr(\mathbf{x}(t)|Y_t)}{\hat{\Pr}(\mathbf{x}_i(t)|Y_t)}\right\} = \sum_{i=1}^{N_p}\ln\frac{\Pr(\mathbf{x}(t)|Y_t)}{\hat{\Pr}(\mathbf{x}_i(t)|Y_t)}\times\Pr(\mathbf{x}(t)|Y_t)$$

The *KL* possesses some very useful properties. It satisfies, perhaps its most important property from a distribution comparison viewpoint—when the true distribution and its estimate are *close* (or identical), then the information quantity is

$$\mathcal{I}_{KL}\left(\Pr(\mathbf{x}(t)|Y_t);\hat{\Pr}(\mathbf{x}_i(t)|Y_t)\right) = 0 \Leftrightarrow \Pr(\mathbf{x}(t)|Y_t) = \hat{\Pr}(\mathbf{x}_i(t)|Y_t) \quad \forall i$$

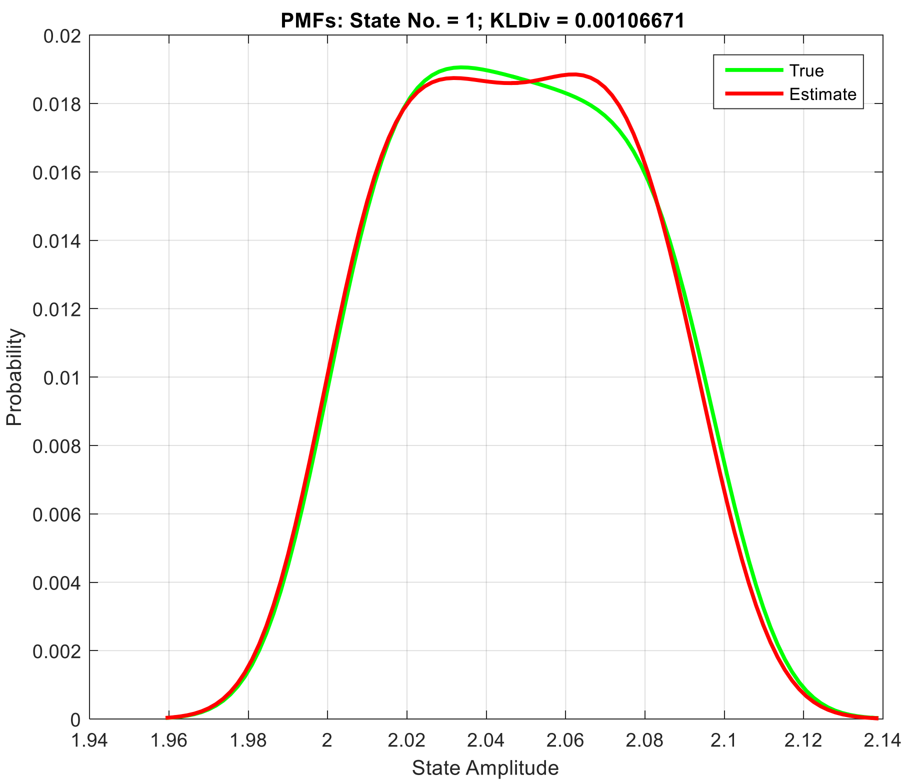
This property infers that as the *estimated* posterior distribution approaches the *true* distribution, then the value of the *KL* approaches *zero* (minimum).

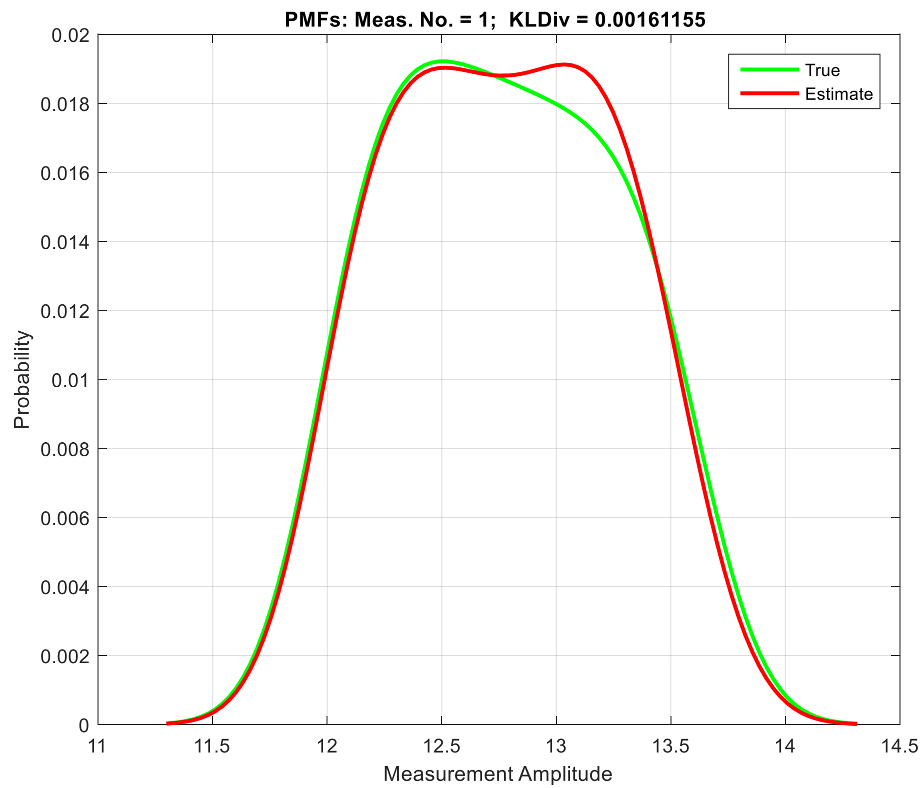
Our interest lies in comparing two probability distributions to determine "how close" they are to one another. Even though, the $\mathcal{I}_{KL}$ does quantify the *difference* between the true and estimated distribution, it is not a distance metric to answer this question due to its lack of symmetry. However, the *Kullback-Leibler divergence* (*KD*) defined by

$$\mathcal{J}_{KD}\left(\Pr(\mathbf{x}(t)|Y_t);\hat{\Pr}(\mathbf{x}_i(t)|Y_t)\right) := \mathcal{I}_{KL}\left(\Pr(\mathbf{x}(t)|Y_t);\hat{\Pr}(\mathbf{x}_i(t)|Y_t)\right) + \mathcal{I}_{KL}\left(\hat{\Pr}(\mathbf{x}_i(t)|Y_t);\Pr(\mathbf{x}(t)|Y_t)\right)$$

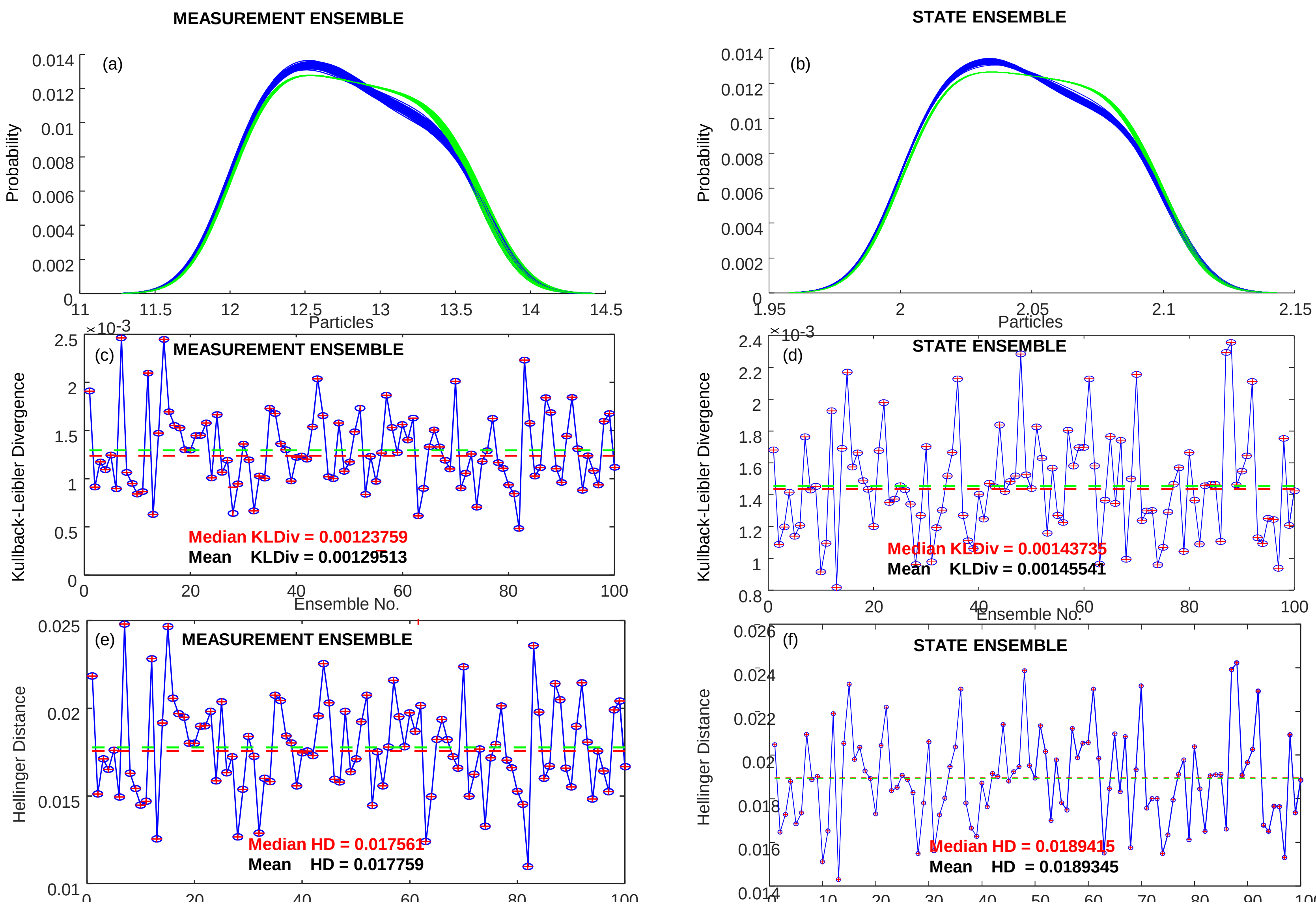
The Kullback-Leibler divergence (KLD) provides a reasonable metric for both the state and measurement particle filters (single realization)

$$\boxed{J_{KD}\left(\Pr\left(X_{\text{TRUE}}(t)\,|\,Y_t\right);\hat{\Pr}\left(\hat{X}_{\text{MAP}}(t)\,|\,Y_t\right)\right)}$$





BOTH Kullbach-Leibler & Hellinger metrics indicate a good PDF match



AN ADAPTIVE PARTICLE FILTERING APPROACH

TO

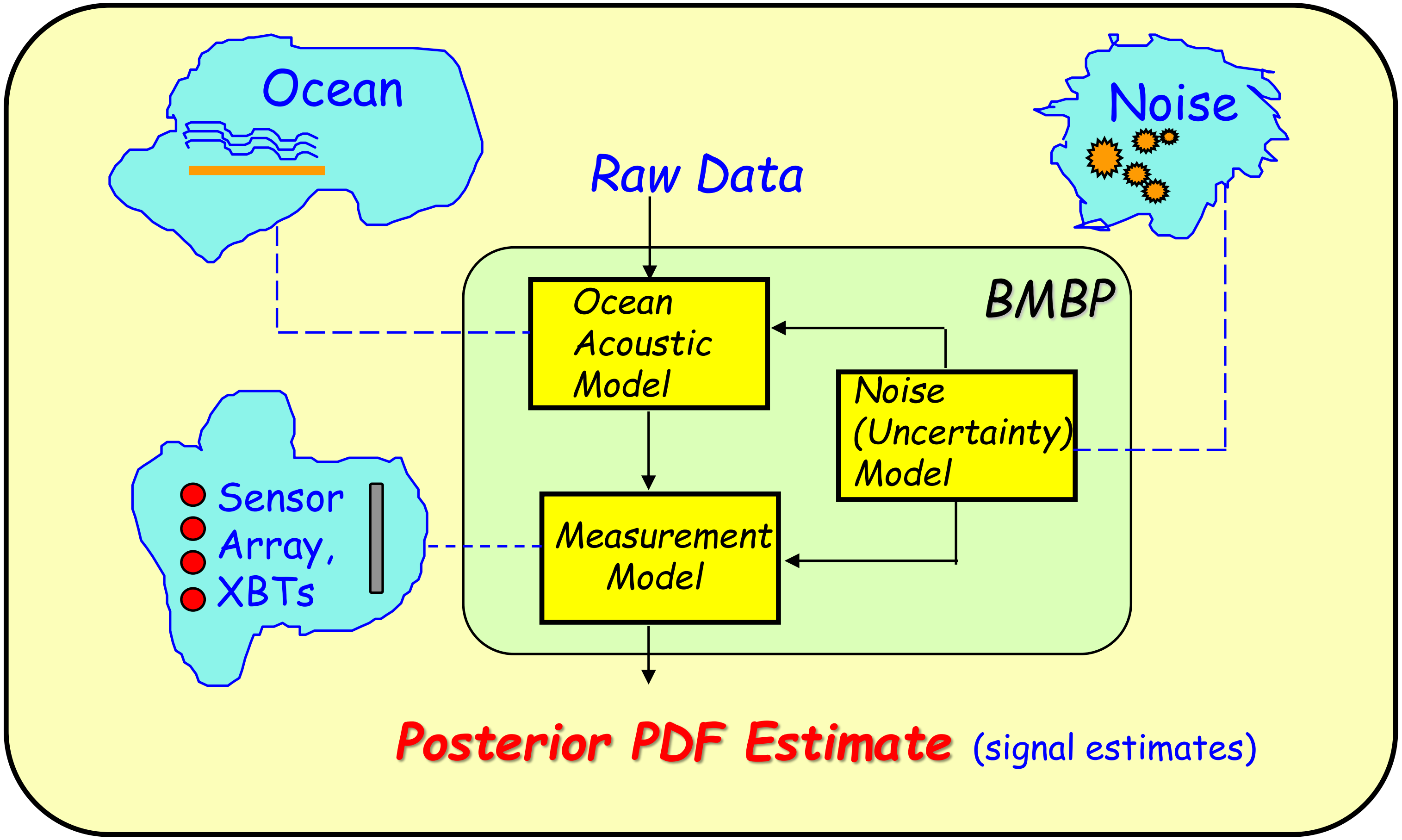
TRACKING MODES

IN A

VARYING SHALLOW OCEAN ENVIRONMENT

Bayesian model-based techniques incorporate "a priori" knowledge of the ocean acoustic phenomenology into a processing scheme to estimate the **posterior distribution** and therefore extract the desired **information**

The use of well-founded environmental propagation models coupled to both measurement and noise (ambient, shipping, etc.) models can be used to *enhance* critical signals



When statistics are "nonstationary," then *sequential MC methods* must be employed to solve the problem

Shallow Ocean Model: Normal-Modes

For a shallow water ocean environment, a normal-mode propagation is used to characterize sound propagation. Starting with the Helmholtz PDE, performing separation of variables and approximating range with a Hankel function, a set of ODEs (in depth) for each mode results:

The standard separation of variables technique and removing the time dependence leads to a set of ordinary differential equations, that is, we obtain a "depth only" representation of the wave equation which is an eigenvalue equation in z with

ODE

$$\boxed{\frac{d^2}{dz^2}\phi_m(z) + \kappa_z^2(m)\phi_m(z) = 0, \;\; m = 1, \cdots, M} \qquad (1)$$

whose eigensolutions $\{\phi_m(z)\}$ are the so called *modal functions* and κ_z is the wave number in the z-direction. These solutions depend on the sound speed profile, $c(z)$, and the boundary conditions at the surface and bottom as well as the corresponding *dispersion* relation given by

$$\kappa^2 = \frac{\omega^2}{c^2(z)} = \kappa_r^2(m) + \kappa_z^2(m), \qquad m = 1, \ldots, M \qquad (2)$$

where $\kappa_r(m)$ is the horizontal wave number associated with the m-th mode in the r direction and ω is the harmonic source frequency.

By assuming a known horizontal source range *a priori*, we obtain a range solution given by the Hankel function,

The corresponding pressure-field measurement using the Hankel solution is given by:

By assuming a known horizontal source range *a priori*, we obtain a range solution given by the Hankel function, $H_0(\kappa_r r_s)$ enabling the pressure-field to be represented by

$$\boxed{p(r_s, z) = \sum_{m=1}^{M} \beta_m(r_s, z_s)\phi_m(z)} \tag{3}$$

where p is the acoustic pressure; ϕ_m is the m^{th} modal function with the modal coefficient defined by

$$\boxed{\beta_m(r_s, z_s) := q \; H_0(\kappa_r r_s) \; \phi_m(z_s)} \tag{4}$$

for q is the source amplitude.

The ODEs are discretized using central differences, transformed to "state-space" form and are augmented with a parameter (θ_m) representing the *m-th* modal coefficient for each mode

$$\Phi_m(z_\ell;\theta_m) := \Phi_m(z_\ell) = [\phi_{m1}(z_\ell)\ \phi_{m2}(z_\ell)\ |\ \boxed{\theta_m(z_\ell)}]^T$$

With this choice of parameters (modal coefficients) the augmented state equations for the m-th mode become

Noise

$$\begin{aligned}
\phi_{m1}(z_\ell) &= \phi_{m2}(z_{\ell-1}) + w_{m1}(z_{\ell-1}) \\
\phi_{m2}(z_\ell) &= -\phi_{m1}(z_{\ell-1}) + (2 - \triangle z_\ell^2 \kappa_z^2(m))\phi_{m2}(z_{\ell-1}) \\
&+ w_{m2}(z_{\ell-1}) \\
\theta_m(z_\ell) &= \theta_m(z_{\ell-1}) + w_{\theta_m}(z_{\ell-1})
\end{aligned} \tag{16}$$

Modal Function

Noise

Modal Coefficient

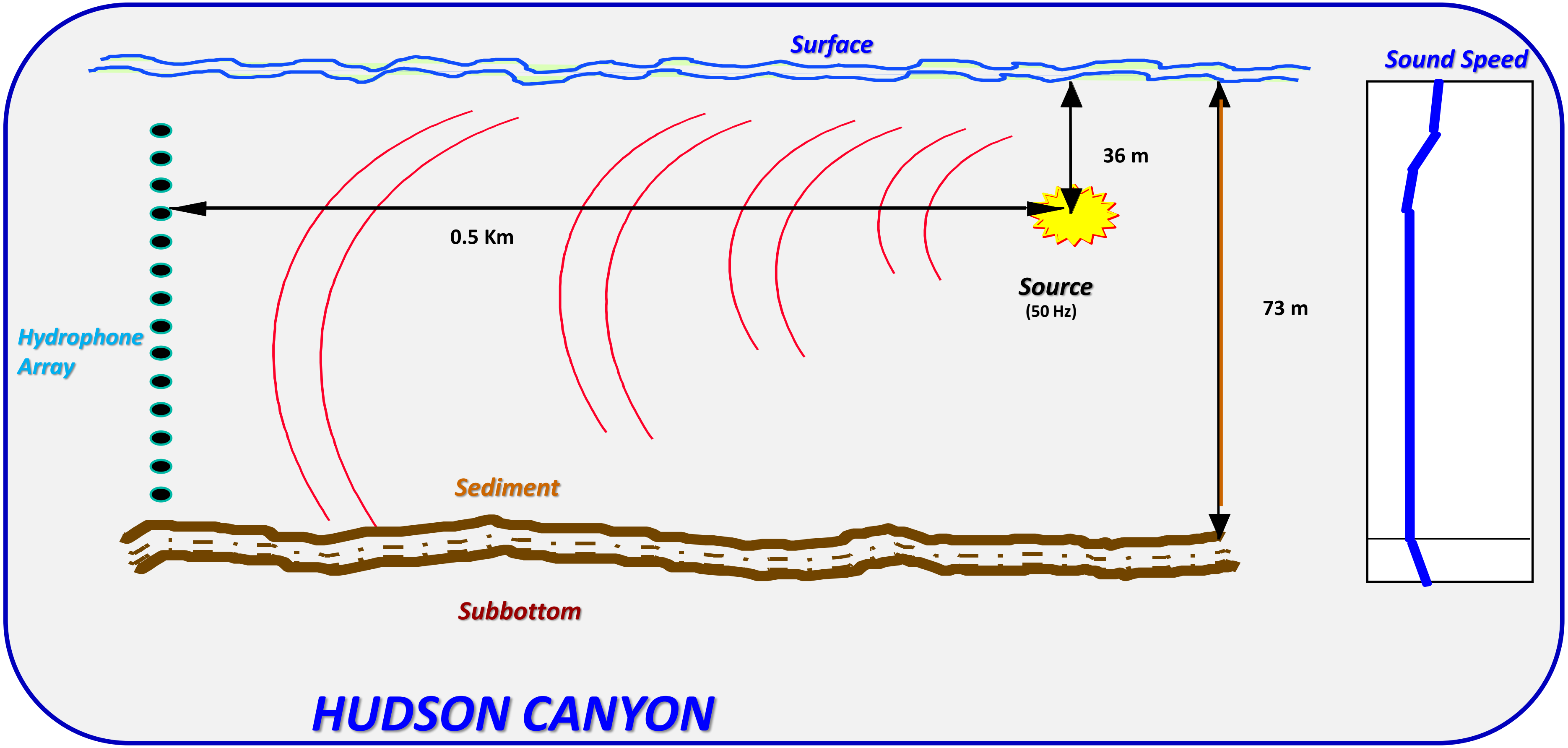
Noise

where we have selected a *random walk* model ($\dot{\theta}_m(z) = w_{\theta_m}(z)$) to capture the variations of the modal coefficients with additive, zero-mean, Gaussian noise of covariance $R_{w_{\theta_m} w_{\theta_m}}$.

The PROBLEM:

The adaptive problem is that of tracking modal functions in a shallow noisy ocean environment for the Hudson Canyon experiment

HC OCEAN: flat bottom, 3 layers, 73m water column, 2.5m sediment, 23 (46) element vertical array source at 0.5 Km range, 36m depth and 50Hz

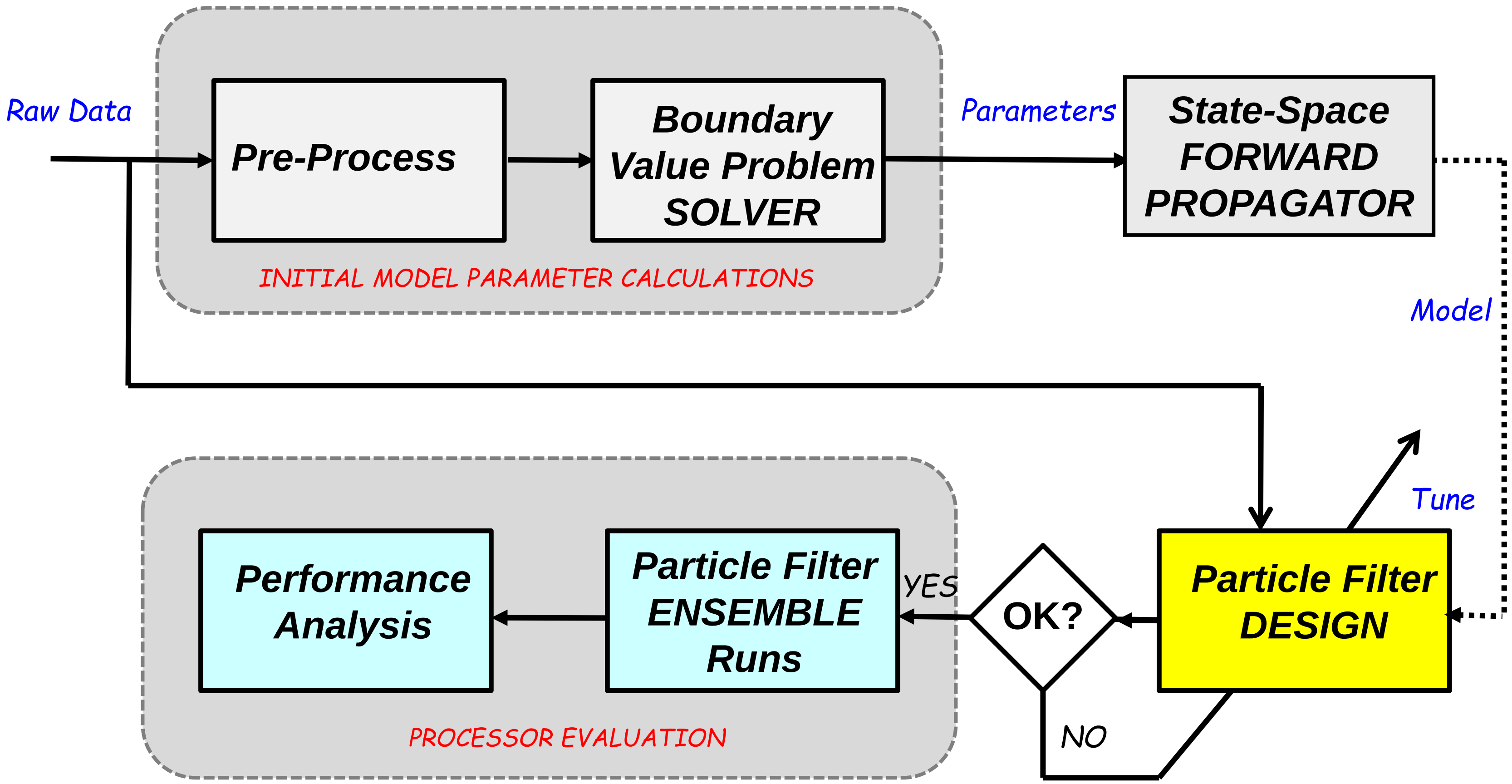


The problem is simply: *given a varying shallow ocean environment, "track" the evolving modal functions while adapting to the changes* or more formally …

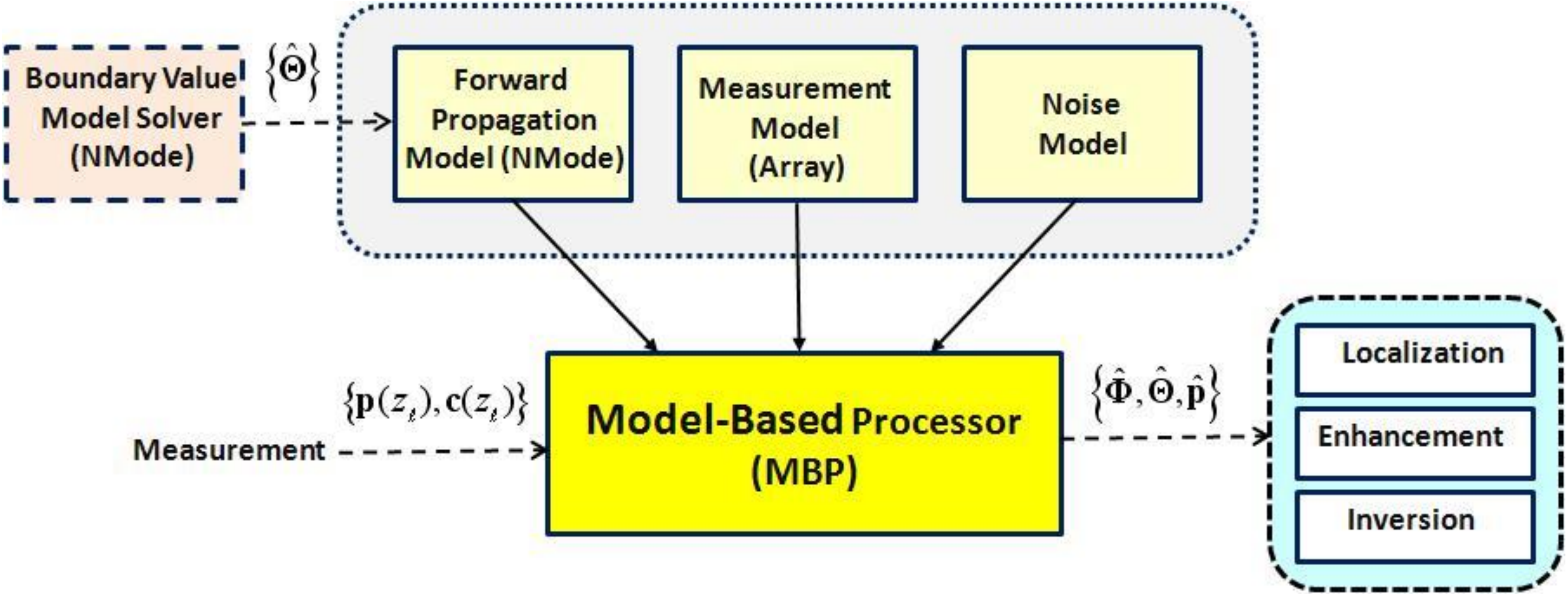
GIVEN a set of noisy pressure-field and sound speed measurements varying in depth, $[\{p(r_s, z_\ell)\}, \{c(z_\ell)\}]$ along with the underlying state-space model of Eqs. 18, 19 and 20 with unknown modal coefficients, FIND the "best" (minimum error variance) estimate of the modal functions, that is, $\{\hat{\phi}_m(z_\ell|z_\ell)\}, \{\hat{\theta}_m(z_\ell|z_\ell)\}; m = 1, \cdots, M$ and measurements (enhanced) $\{\hat{p}(r_s, z_\ell)\}$.

Particle Filter Design

Particle filter design consists of a set of initial parameter runs using *simulated* then actual *experimental* data



The outputs of the particle filter can be used for: localization, enhancement, inversion and detection



Particle filtering is used to estimate the posterior distribution and therefore the MAP modal function estimates while the CM (conditional mean) is found by MC integration

$$\hat{p}[\phi(z_\ell)|P_z] = \sum_{i=1}^{N_p} \hat{W}_i(z_\ell)\delta\big(\phi(z_\ell) - \phi_i(z_\ell)\big) \quad \forall z$$

$\hat{W}_i(z_\ell) \quad \propto \hat{\Pr}[\phi_i(z_\ell)]$ is the estimated weights at depth z_ℓ;

$\phi_i(z_\ell)$ is the i-th particle at depth z_ℓ;

$\hat{p}[\cdot]$ is the estimated empirical distribution;

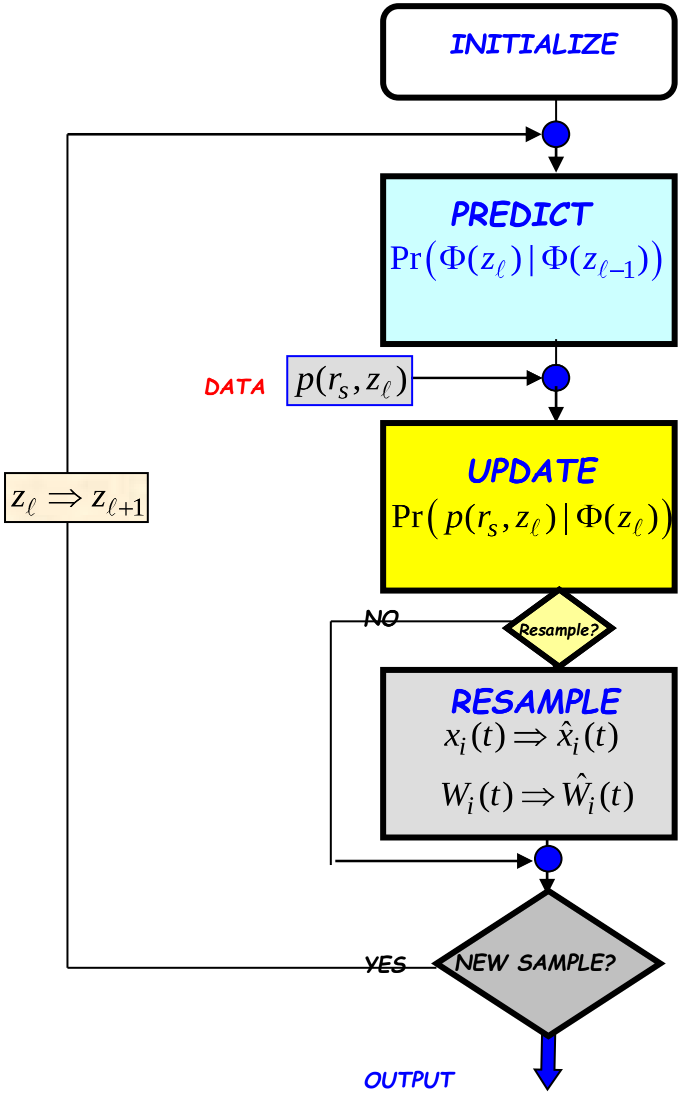
P_z is the set of batch pressure-field measurements,

$$P_z = \{p(z_1) \cdots p(z_L)\}.$$

the maximum a posteriori (MAP) estimate is simply found by locating the location of the particular particle $\hat{x}_i(z_\ell)$ corresponding to the maximum of the *PMF*, that is

$$\hat{\Phi}_i^{MAP}(z) = \max_i \quad \hat{p}[\phi_i(z)|P_z]$$

The simple BOOTSTRAP PF is used to estimate the posterior distribution



Thus, we estimate the posterior distribution using a sequential Monte Carlo approach and construct a *bootstrap particle filter* [19]-[24] using the following steps:

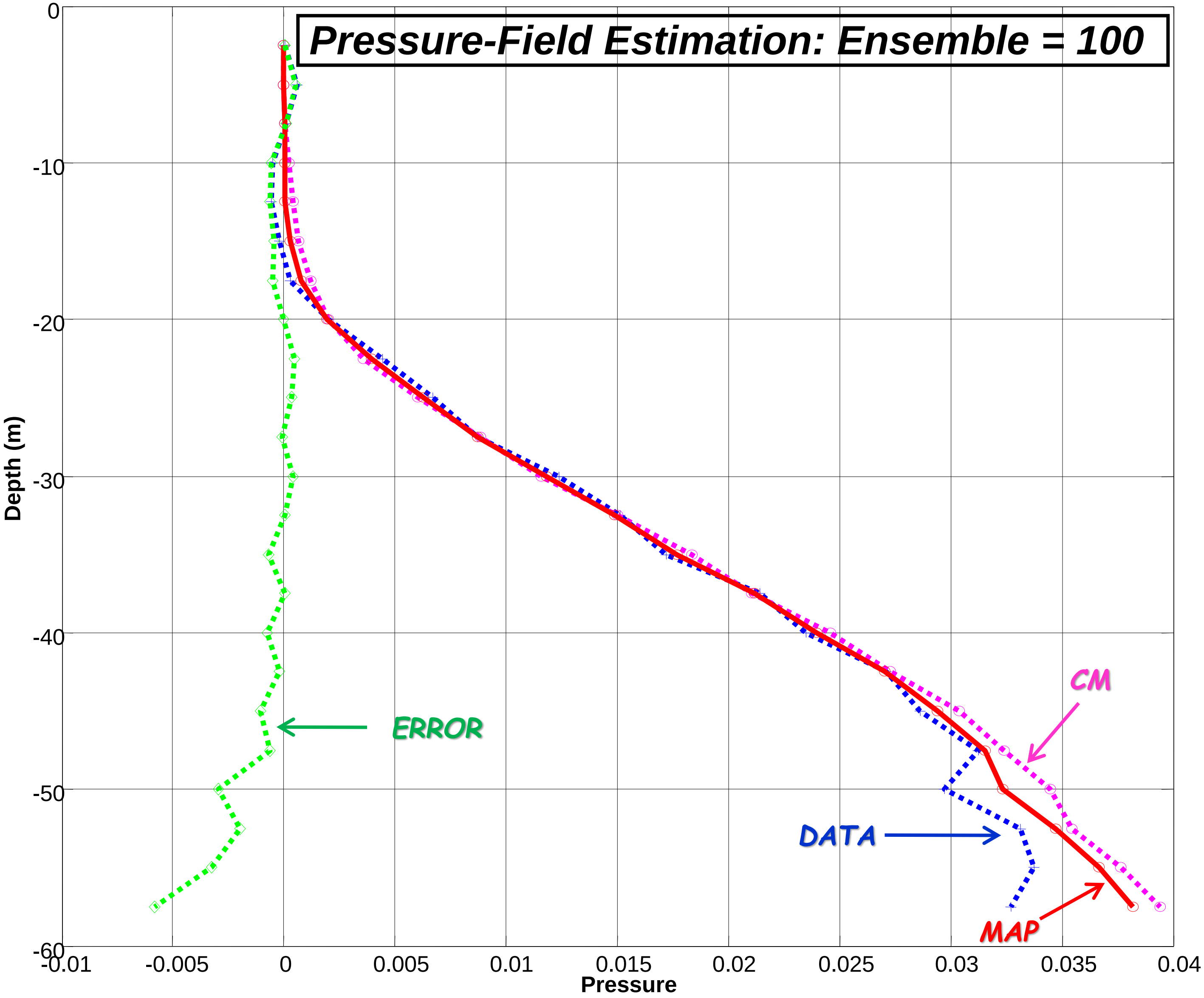
- Initialize: $\Phi_m(0), w_{z_\ell} \sim \mathcal{N}(0, R_{ww}), W_i(0) = 1/N_p; i = 1, \cdots, N_p$;
- State Transition: $\Phi_m(z_\ell) = \mathbf{A}_m(z_{\ell-1})\Phi_m(z_{\ell-1}) + \mathbf{w}_m(z_{\ell-1})$;
- Likelihood Probability: $\Pr[p(r_s, z_\ell)|\Phi(z_\ell)]$;
- Weights: $W_i(z_\ell) = W_i(z_{\ell-1}) \times \Pr[\Phi_m(z_\ell)|\Phi_m(z_{\ell-1})]$;
- Normalize: $\mathcal{W}_i(z_\ell) = \frac{W_i(z_\ell)}{\sum_{i=1}^{N_p} W_i(z_\ell)}$;
- Resample: $\tilde{\Phi}_i(z_\ell) \Rightarrow \Phi_i(z_\ell)$;
- Posterior: $\hat{\Pr}[\Phi_m(z_\ell)|P_z] = \sum_{i=1}^{N_p} \mathcal{W}_i(z_\ell)\delta(\phi(z_\ell) - \phi_i(z_\ell))$; and
- MAP Estimate: $\hat{\Phi}_i^{MAP}(z) = \max_i \hat{\Pr}[\phi_i(z_\ell)|P_z]$;
- MMSE Estimate: $\hat{\Phi}_i^{MMSE}(z) = \frac{1}{N_p}\sum_{i=1}^{N_p} \mathcal{W}_i(z_\ell)\phi_i(z_\ell)$

EXPERIMENTAL DATA RESULTS

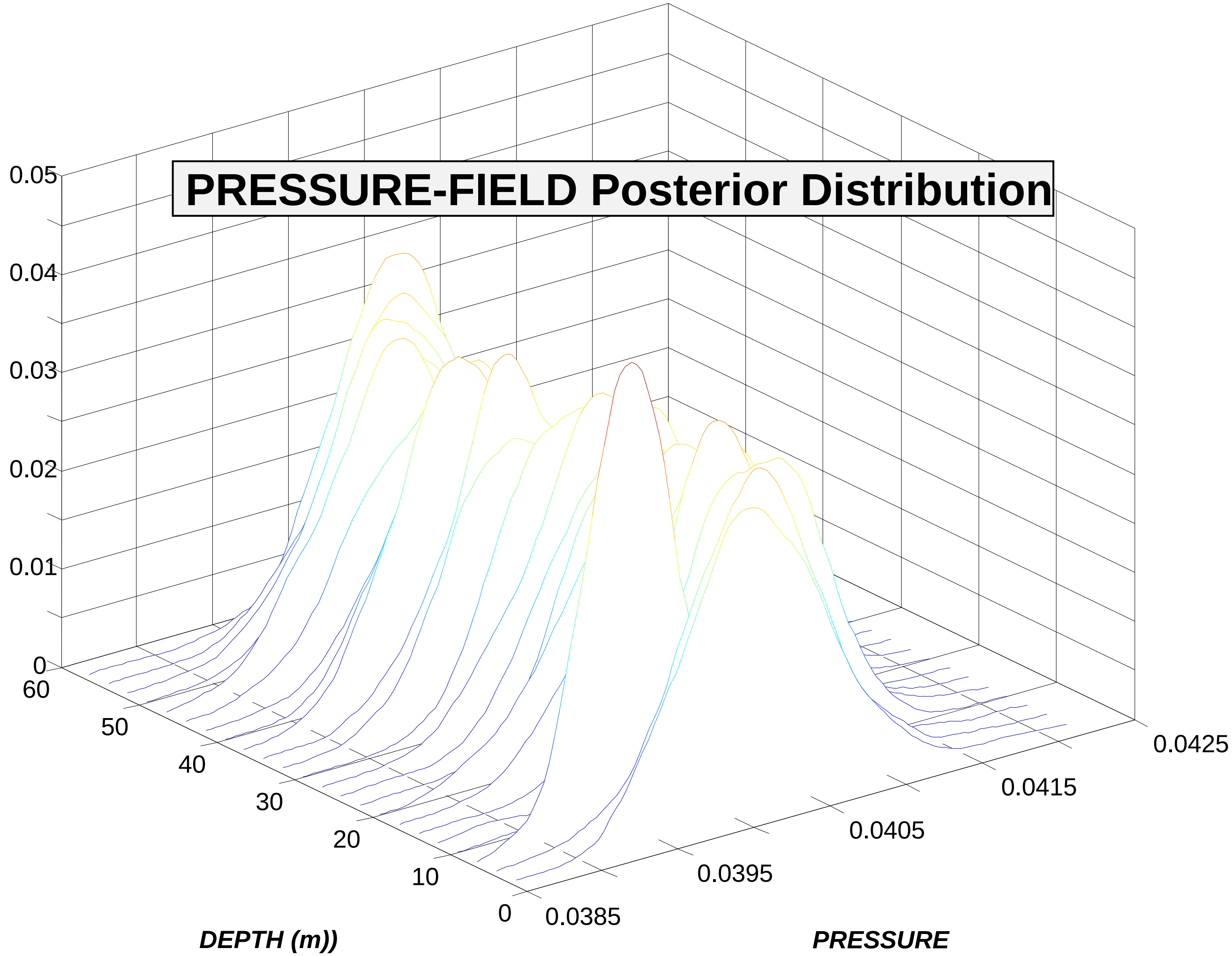
The Hudson Canyon experiment is:

- 23-element vertical array with 2.5m pitch
- Source at 0.5Km temporal frequency is 50Hz, Depth=36m
- 5 modes supporting the water column
- Noise is assumed AGWN for this test
- Length of particle filter, Nparticles=1000

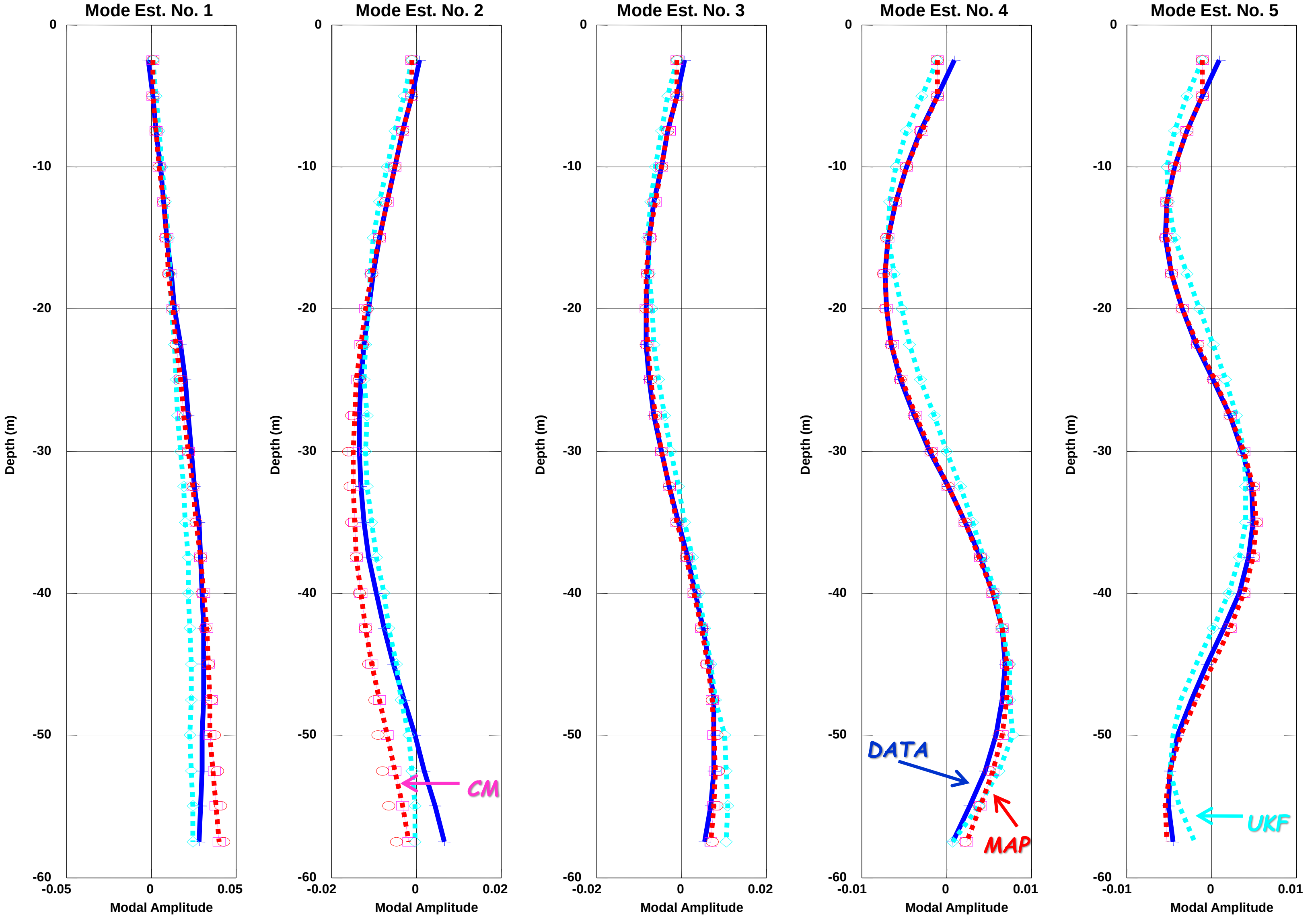
Predicting the pressure-field from noisy array measurements is quite reasonable



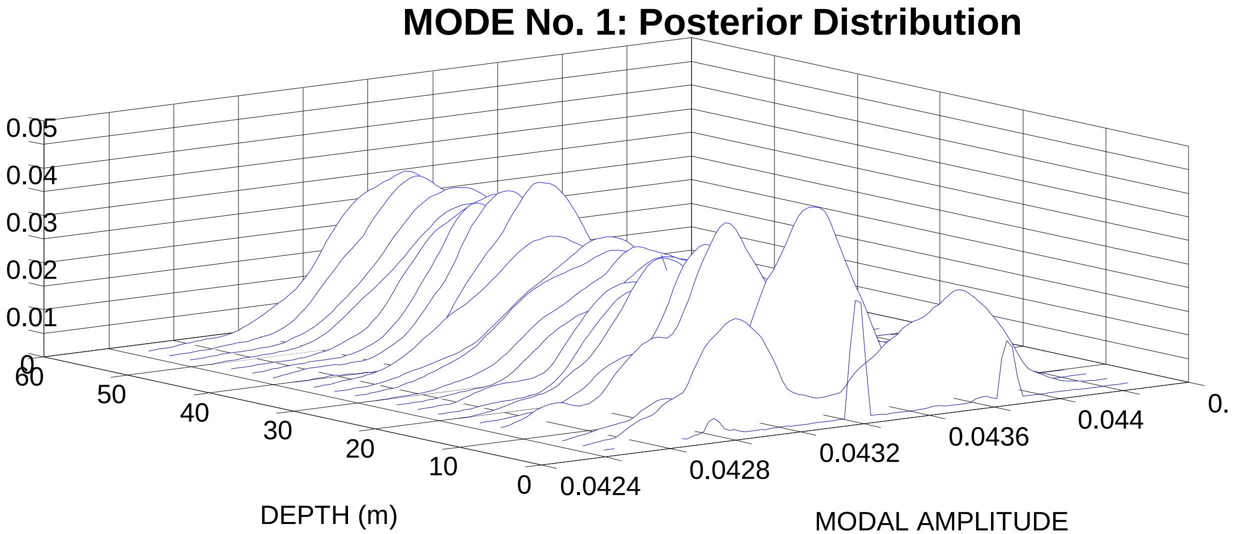
Pressure-field posterior PDFs can be approximated reasonably by a unimodal distribution explaining the good performance of the UKF processor for enhancement, but the modal estimates tell a different story

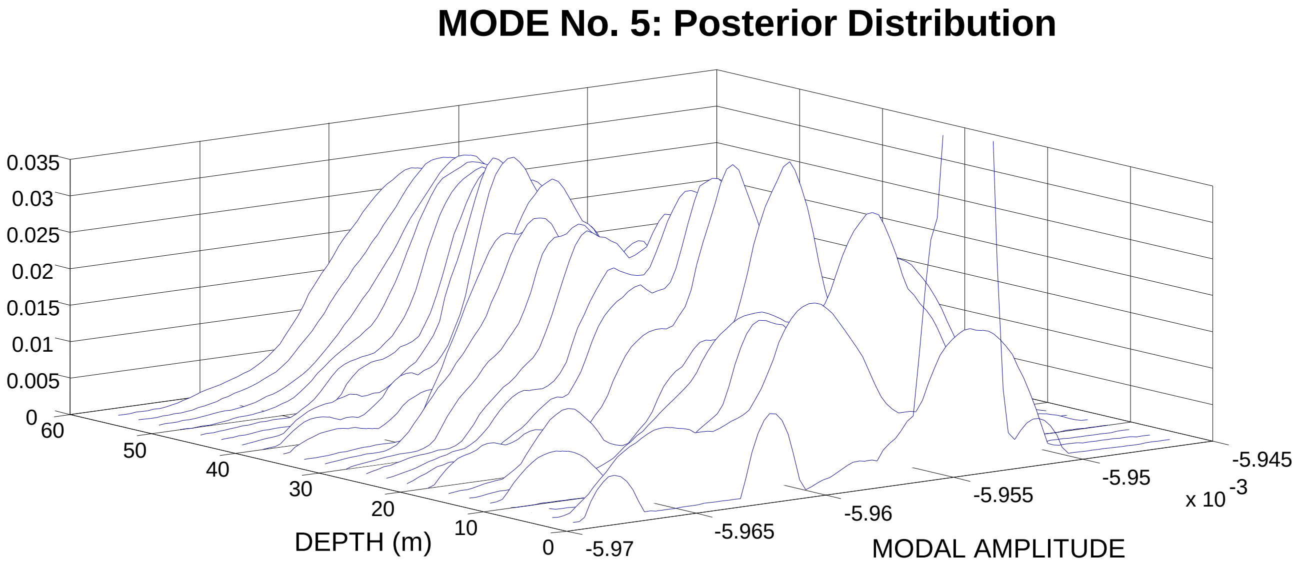


Modal function tracking is reasonable AND the PF results are GOOD (relative to UKF)

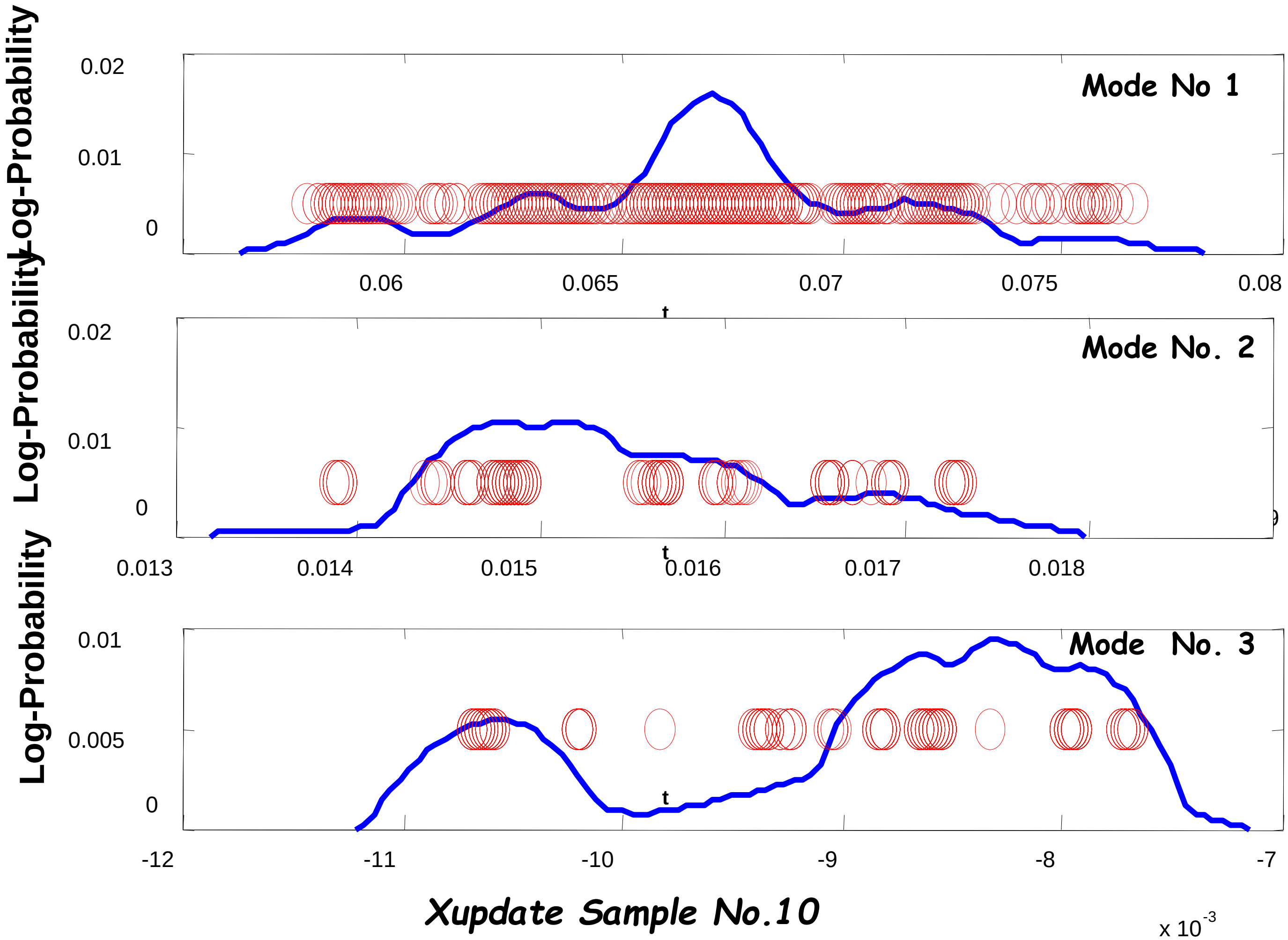


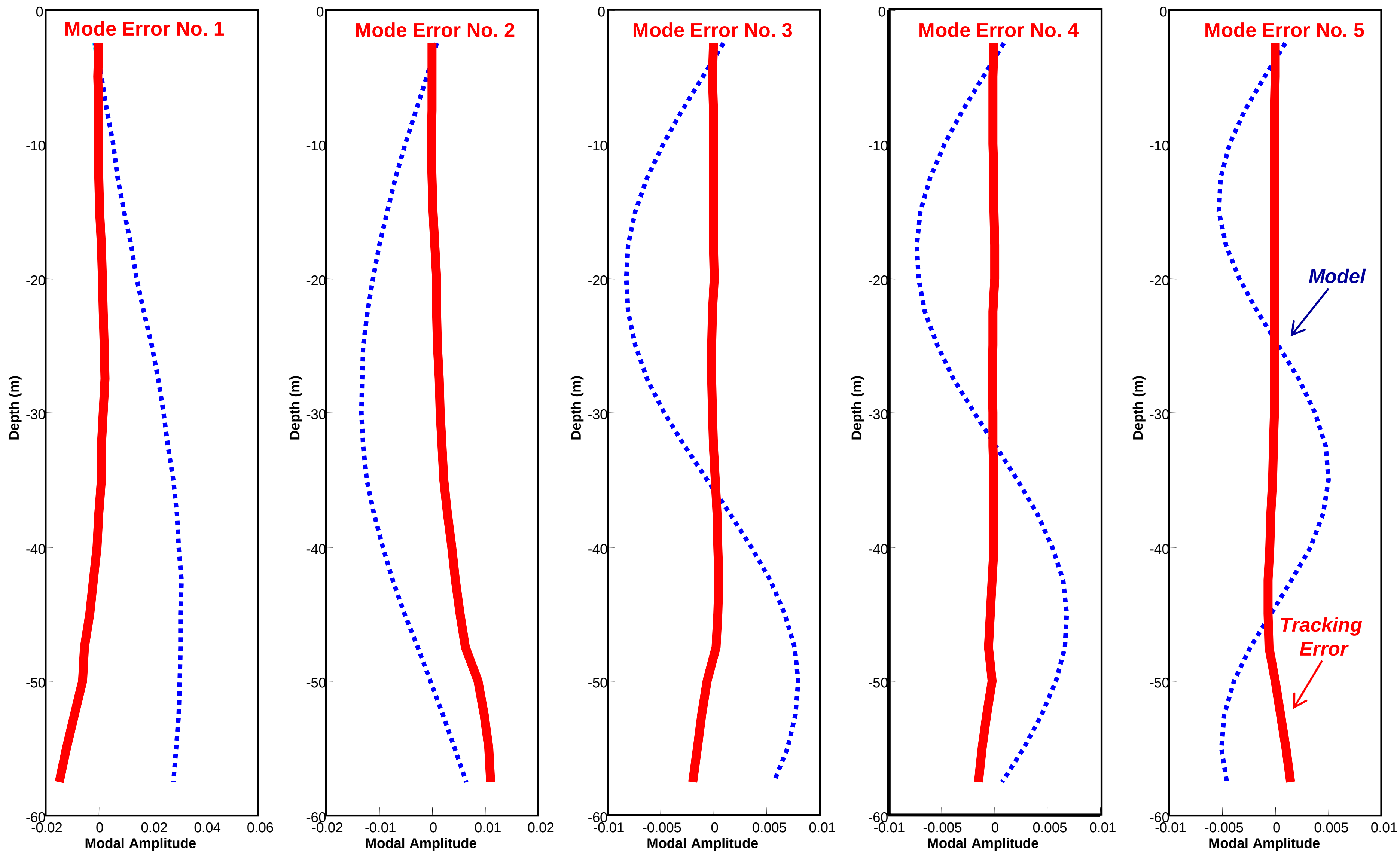
Modal posterior PDFs are not unimodal indicating the potential failure of unimodal estimators





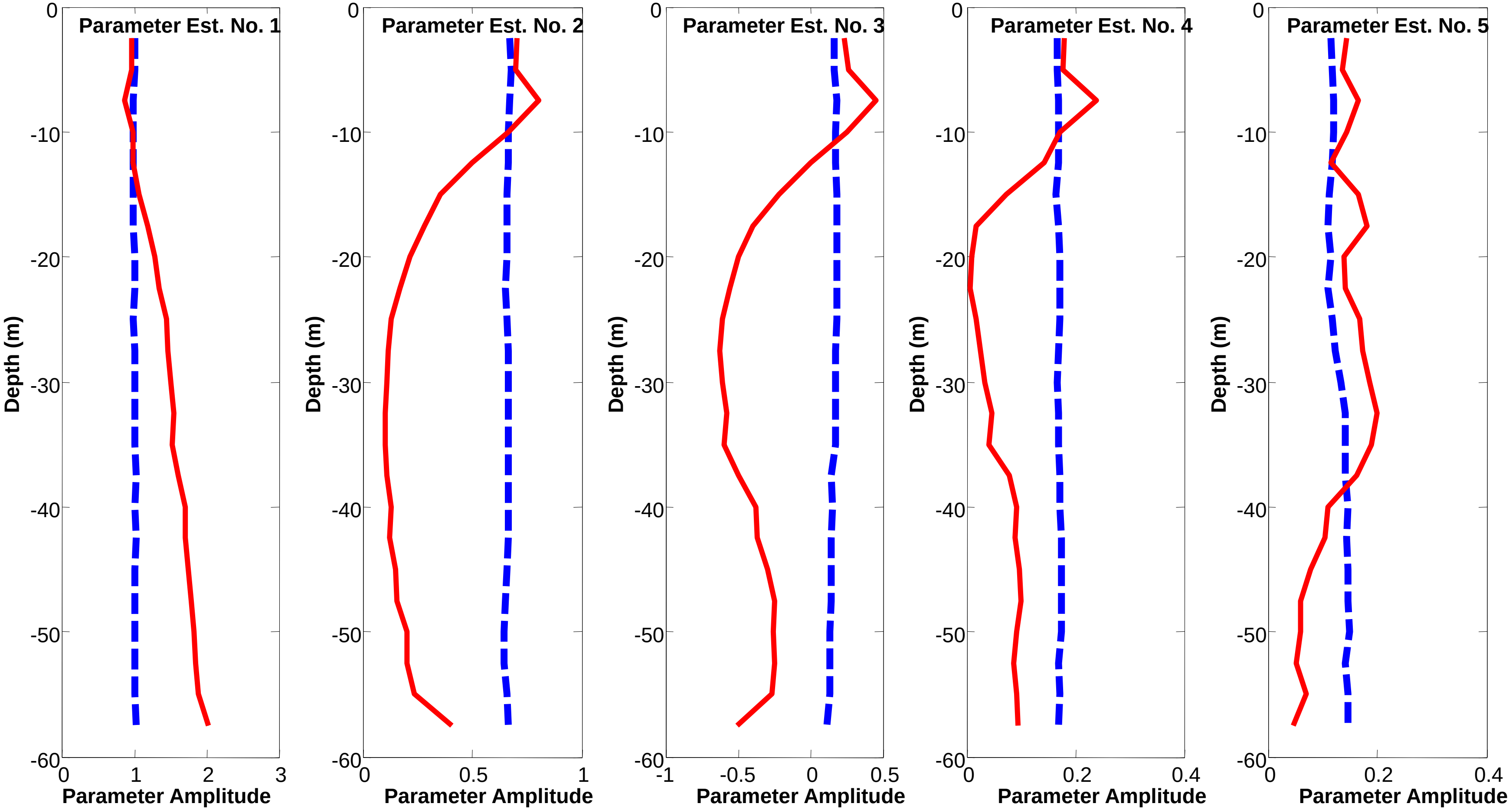
A cross-sectional slice through the modal PDFs confirms this representation (multi-modal)



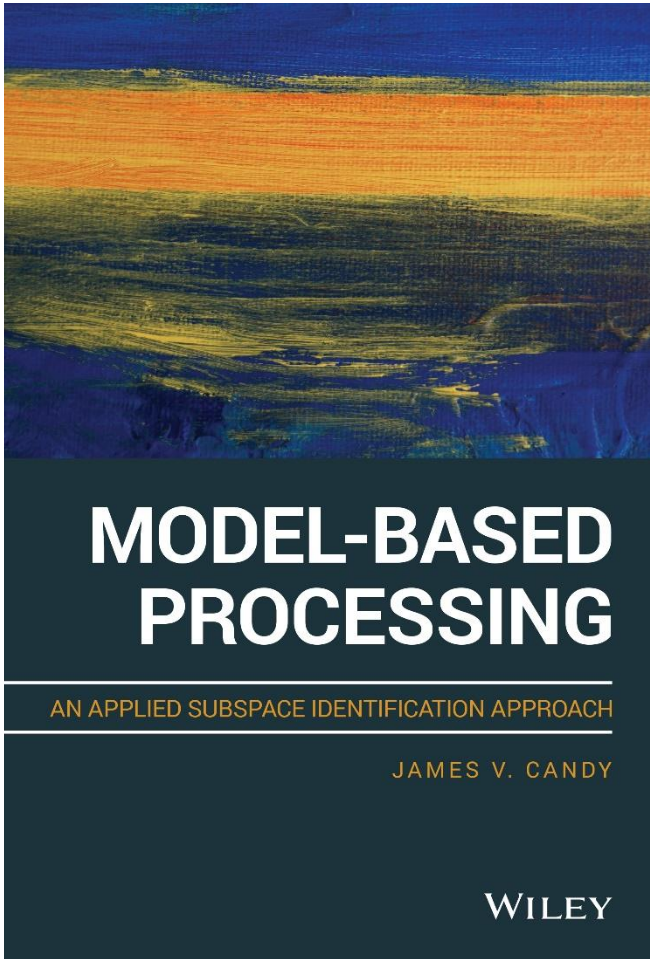


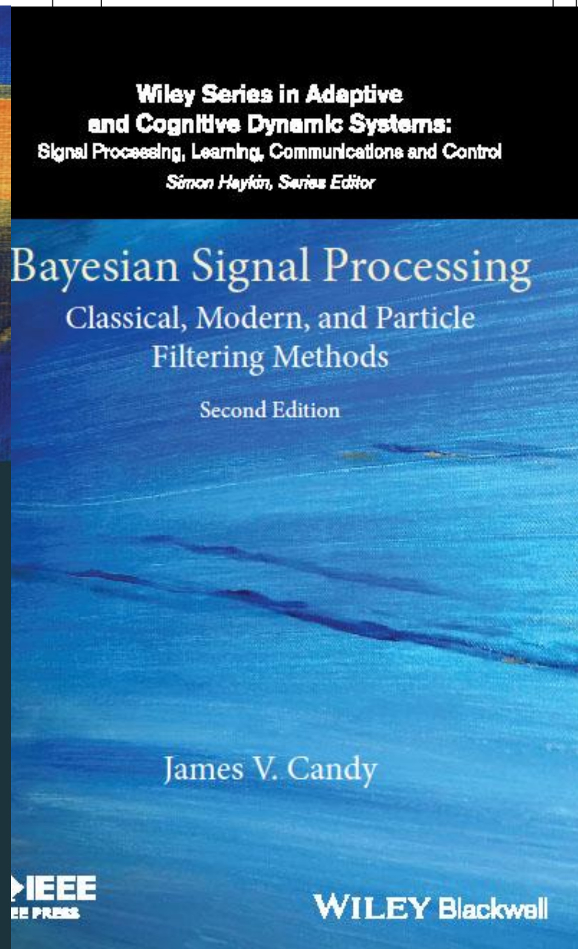
Mode Error No. 1
Mode Error No. 2
Mode Error No. 3
Mode Error No. 4
Mode Error No. 5
Model
Tracking Error
Depth (m)
Modal Amplitude
0
-10
-20
-30
-40
-50
-60
-0.02
0
0.02
0.04
0.06
-0.01
0.01
-0.005
0.005

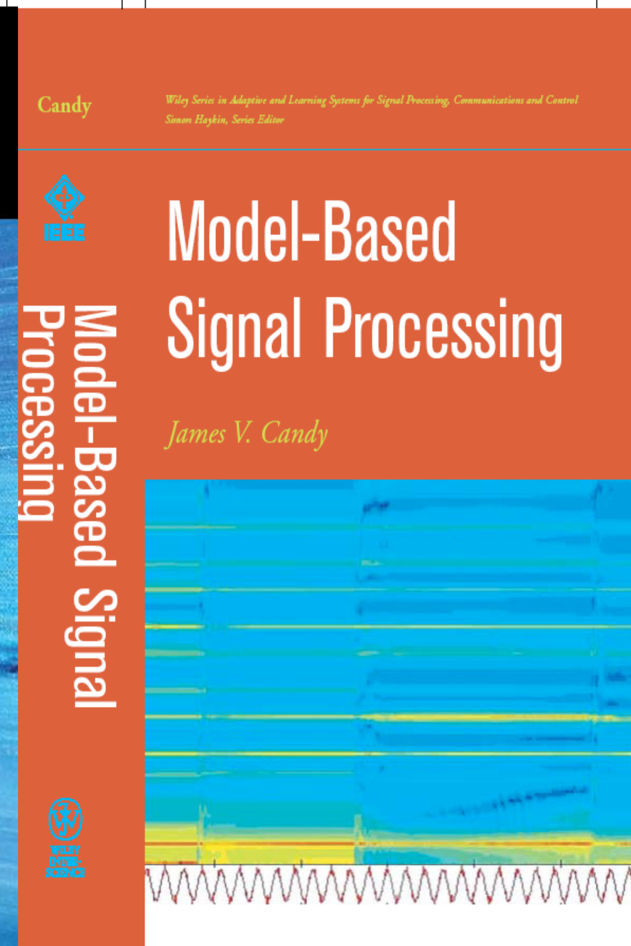
ADAPTIVITY: Parametric adjustments for each modal coefficient is shown



Texts on Subspace, Bayesian & Model-Based Processing







SUMMARY:

- ***We have presented an overview of Bayesian signal processing evolving into the design of a particle filter***
- ***We have developed a solution to the mode tracking problem using a particle filter and compared it to the unimodal UKF***
- ***We have demonstrated the BOOTSTRAP PF performance on pressure-field data from the Hudson Canyon and a 23-element hydrophone array***
- ***We have shown that the PF performance for this case is quite reasonable***

BACK-UPS

Another important metric for *particle filters* is based on the Kullback-Leibler information (KL) defined by:

Suppose $\Pr(x_i)$ is the *true* discrete posterior probability distribution and $\hat{\Pr}(\hat{x}_i)$ is the estimated distribution. Then the *Kullback-Leibler Information* (*KL*) quantity of the true distribution relative to the estimated is defined by using

$$\boxed{\begin{aligned}\mathcal{I}_{KL}\left(\Pr(x_i);\hat{\Pr}(\hat{x}_i)\right) \quad &:= \quad E_X\left\{\ln\frac{\Pr(x_i)}{\hat{\Pr}(\hat{x}_i)}\right\} = \sum_{i=1}^{N}\Pr(x_i)\ln\frac{\Pr(x_i)}{\hat{\Pr}(\hat{x}_i)} \\ &= \quad \sum_{i=1}^{N}\Pr(x_i)\ln\Pr(x_i) - \sum_{i=1}^{N}\Pr(x_i)\ln\hat{\Pr}(\hat{x}_i)(\end{aligned}}$$

where we chose $\log_b = \ln$. The *KL* possesses some very interesting properties which we state without proof (see [60] for details) such as

1. $\mathcal{I}_{KL}\left(\Pr(x_i);\hat{\Pr}(\hat{x}_i)\right) \geq 0$
2. $\boxed{\mathcal{I}_{KL}\left(\Pr(x_i);\hat{\Pr}(\hat{x}_i)\right) = 0 \Leftrightarrow \Pr(x_i) = \hat{\Pr}(\hat{x}_i) \quad \forall i}$
3. The negative of the *KL* is the *entropy*, $\mathcal{H}_{KL}\left(\Pr(x_i);\hat{\Pr}(\hat{x}_i)\right)$

The second property implies that as the *estimated* posterior distribution approaches the *true* distribution, then the value of the *KL* approaches *zero* (minimum). Thus,

There are other practical tests that can be performed to particle filters; however, one of the primary tests is the Kullback-Leibler divergence (KLD) metric defined by:

However, our interest lies in comparing two probability distributions to determine "how close" they are to one another. Even though $\mathcal{I}_{KL}$ does quantify the difference between the true and estimated distributions, unfortunately it is not a distance measure due to its lack of symmetry. However, the *Kullback divergence* (*KD*) defined by a combination of $\mathcal{I}_{KL}$

$$\boxed{\mathcal{J}_{KD}\left(\Pr(x_i);\hat{\Pr}(\hat{x}_i)\right) = \mathcal{I}_{KL}\left(\Pr(x_i);\hat{\Pr}(\hat{x}_i)\right) + \mathcal{I}_{KL}\left(\hat{\Pr}(\hat{x}_i);\Pr(x_i)\right)}$$

is a distance measure between distributions indicating "how far" one is from the other. Consider the following example of this calculation.

Before we close this section, let us see how the *KD* can be applied to *PF* design. Typically, we have a simulation model of the dynamics of the system under investigation in some form or another, that is, the model can range from a very detailed "truth model" as discussed in [5] to a simple signal processing representation (e.g. sinusoids). In any case using the truth model we can generate a "true distribution" of the system, say $\Pr\left(x(t)|Y_t\right)$ and incorporate it into the divergence criterion. The nonparametric estimate of the posterior distribution $\hat{\Pr}\left(x(t)|Y_t\right)$ provided by the particle filter can be used in the criterion enabling an estimate that can be compared to the truth, that is,

$$\boxed{\begin{aligned}\mathcal{J}_{KD}\left(\Pr\left(x(t)|Y_t\right);\hat{\Pr}\left(x(t)|Y_t\right)\right) \quad &= \quad \mathcal{I}_{KL}\left(\Pr\left(x(t)|Y_t\right);\hat{\Pr}\left(x(t)|Y_t\right)\right)\\ &+ \quad \mathcal{I}_{KL}\left(\hat{\Pr}\left(x(t)|Y_t\right);\Pr\left(x(t)|Y_t\right)\right)\end{aligned}}$$

Consider the following example of resampling emphasizing the generation of more particles in the high probability regions and the removal of the small weighted particles

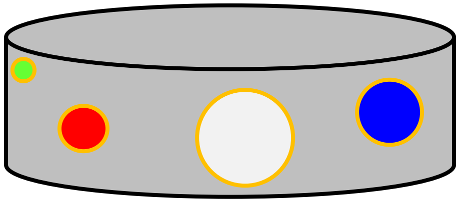
Assume a we have 100 balls (samples) with the following probabilities:

WHITE (W) = 40; Pr(W) = 0.40

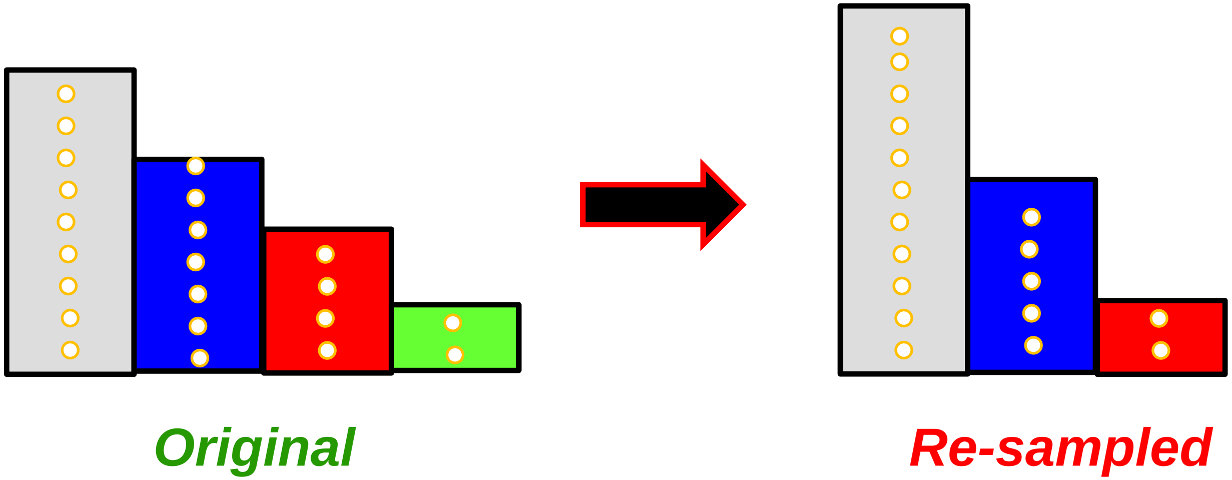
BLUE (B) = 25; Pr(B) = 0.25

RED (R) = 10; Pr(R) = 0.10

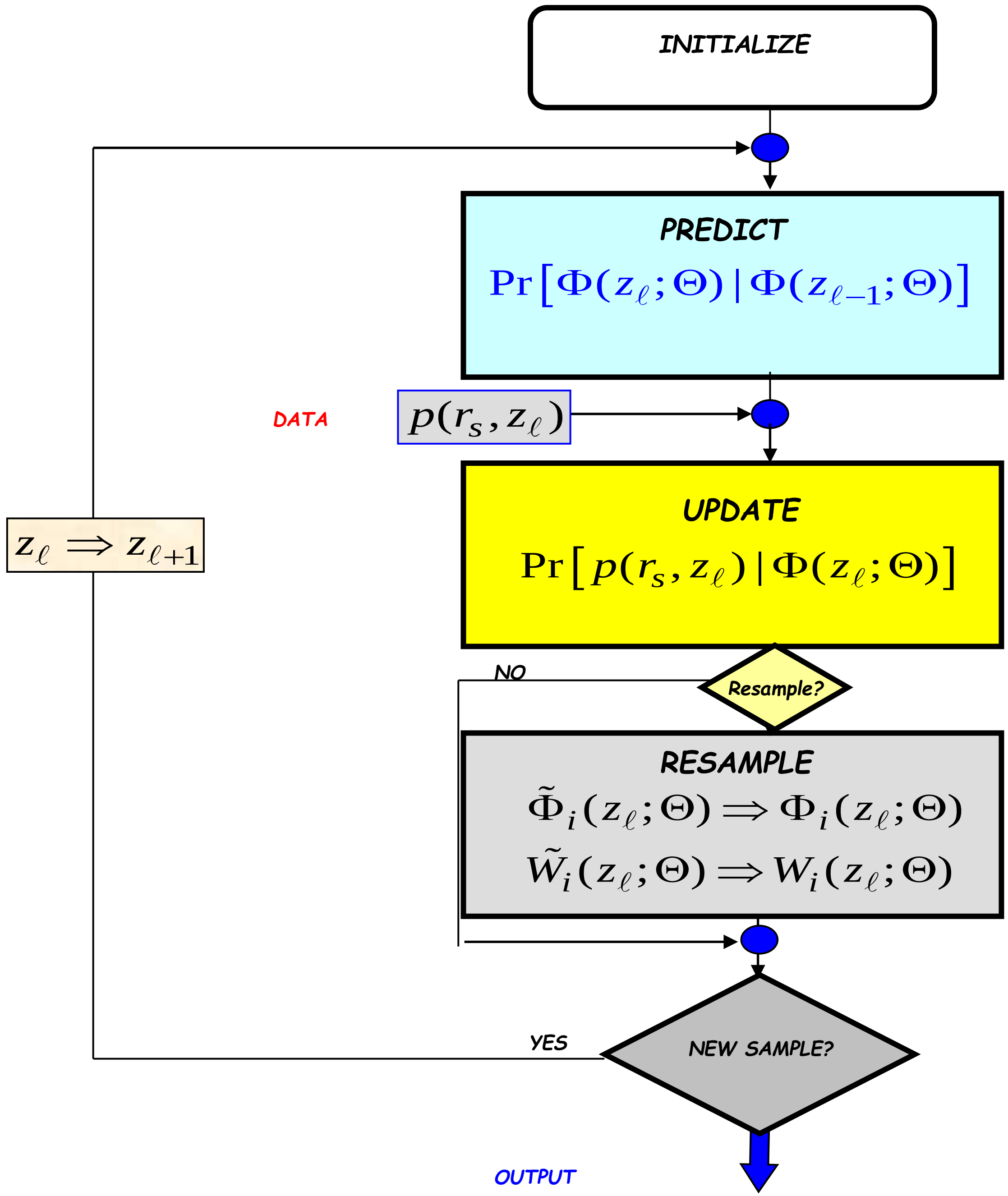
GREEN (G) = 2; Pr(G) = 0.02



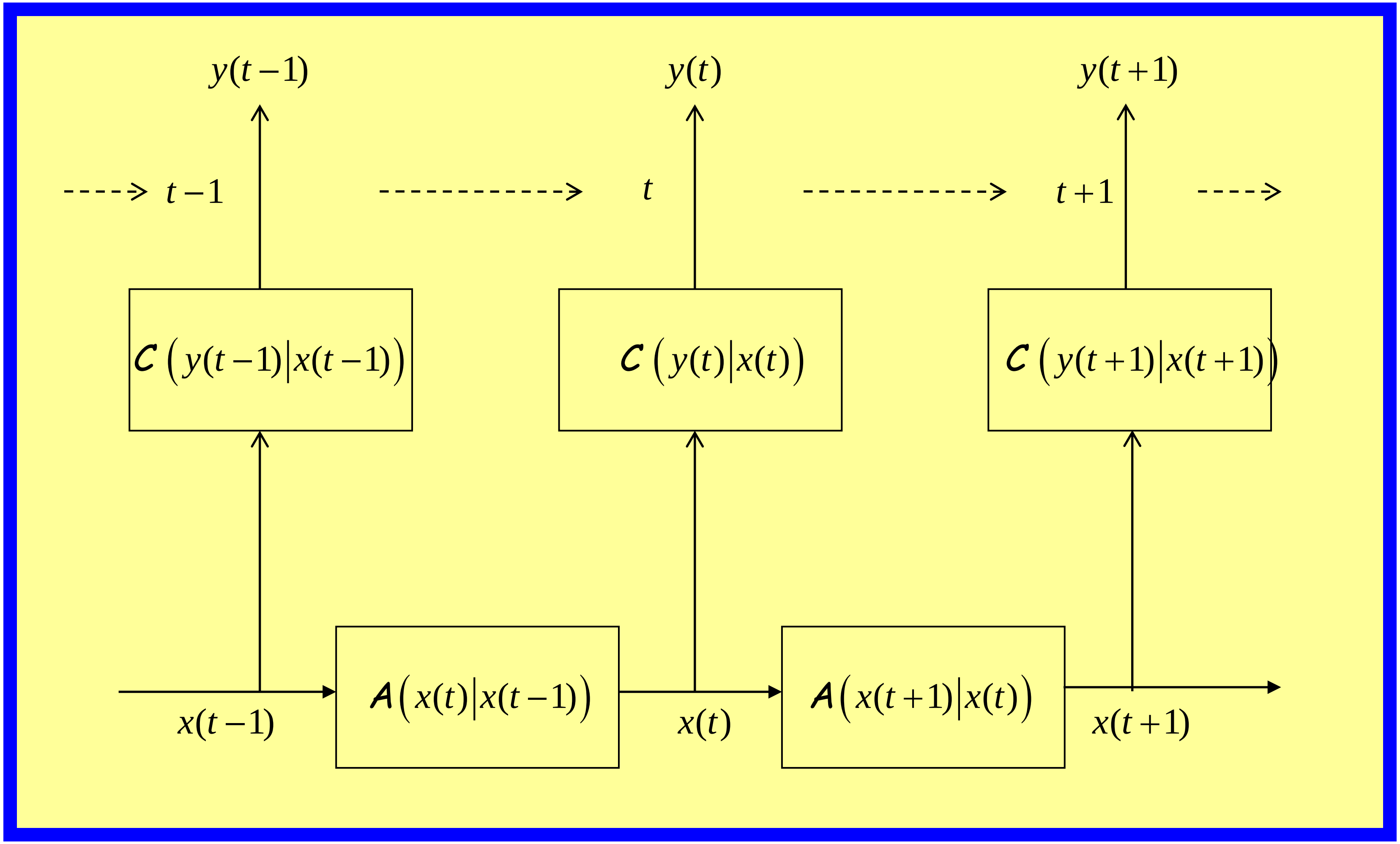
Placing the balls into an urn (individually --> uniform weighting 1/100) and then sampling WITH replacement we have a NEW or RE-SAMPLED histogram:



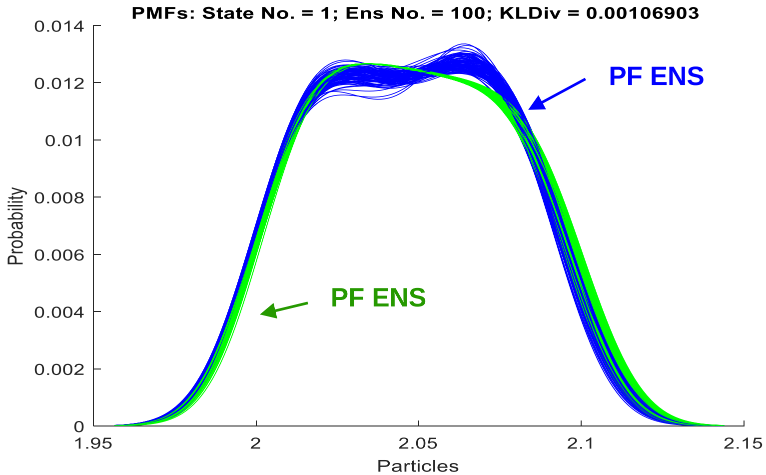
Making draws (resampling with replacement) implies that the new histogram (PDF) would have a different shape eliminating the smaller probabilities

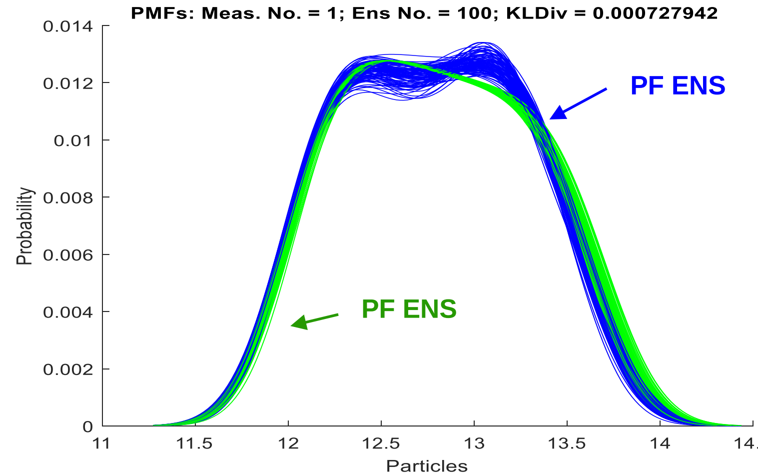

INITIALIZE
PREDICT
$\Pr\left[\Phi(z_\ell;\Theta)\,|\,\Phi(z_{\ell-1};\Theta)\right]$
DATA
$p(r_s, z_\ell)$
UPDATE
$\Pr\left[p(r_s, z_\ell)\,|\,\Phi(z_\ell;\Theta)\right]$
$z_\ell \Rightarrow z_{\ell+1}$
NO
Resample?
RESAMPLE
$\tilde{\Phi}_i(z_\ell;\Theta) \Rightarrow \Phi_i(z_\ell;\Theta)$
$\tilde{W}_i(z_\ell;\Theta) \Rightarrow W_i(z_\ell;\Theta)$
YES
NEW SAMPLE?
OUTPUT

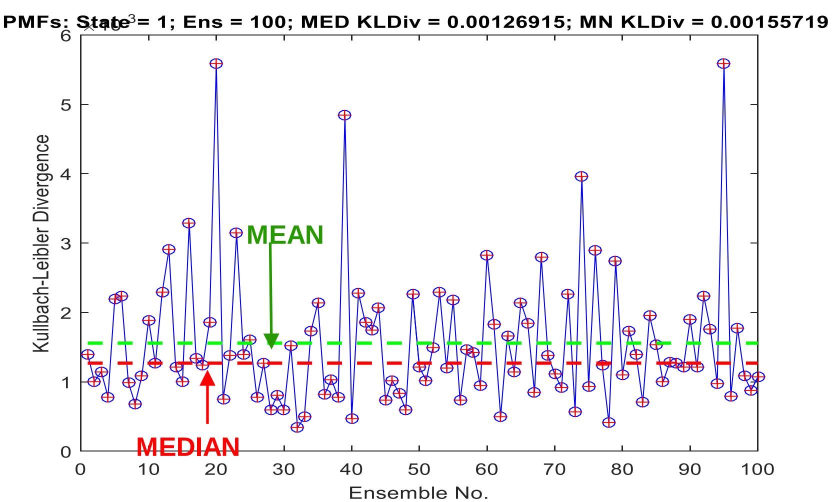
The BSP implemented in state-space is:

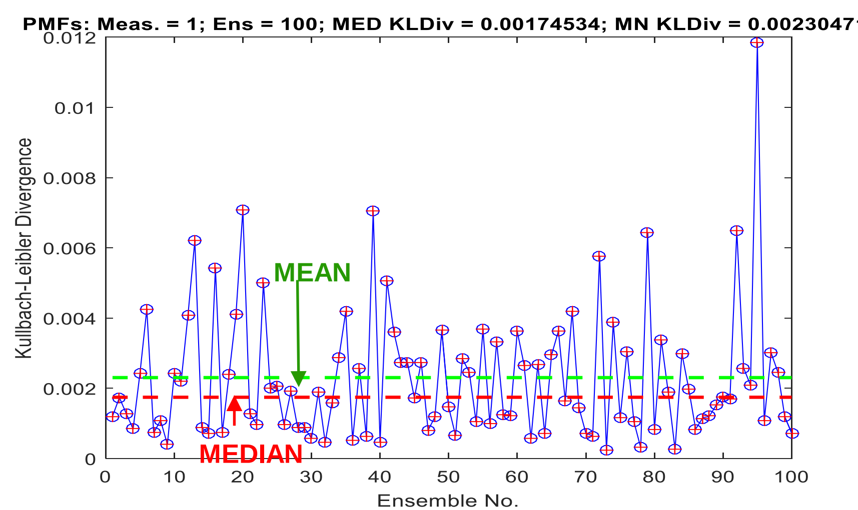


The KLD metric can also be applied to the ensemble and provide average results

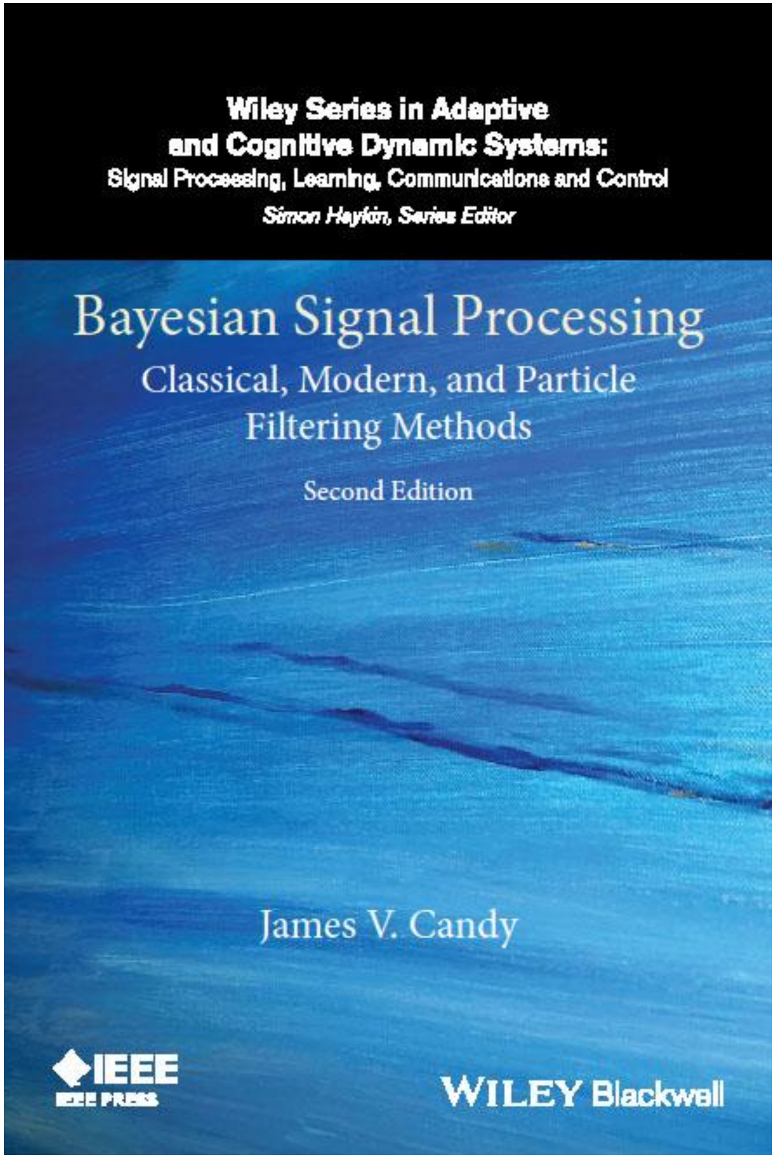




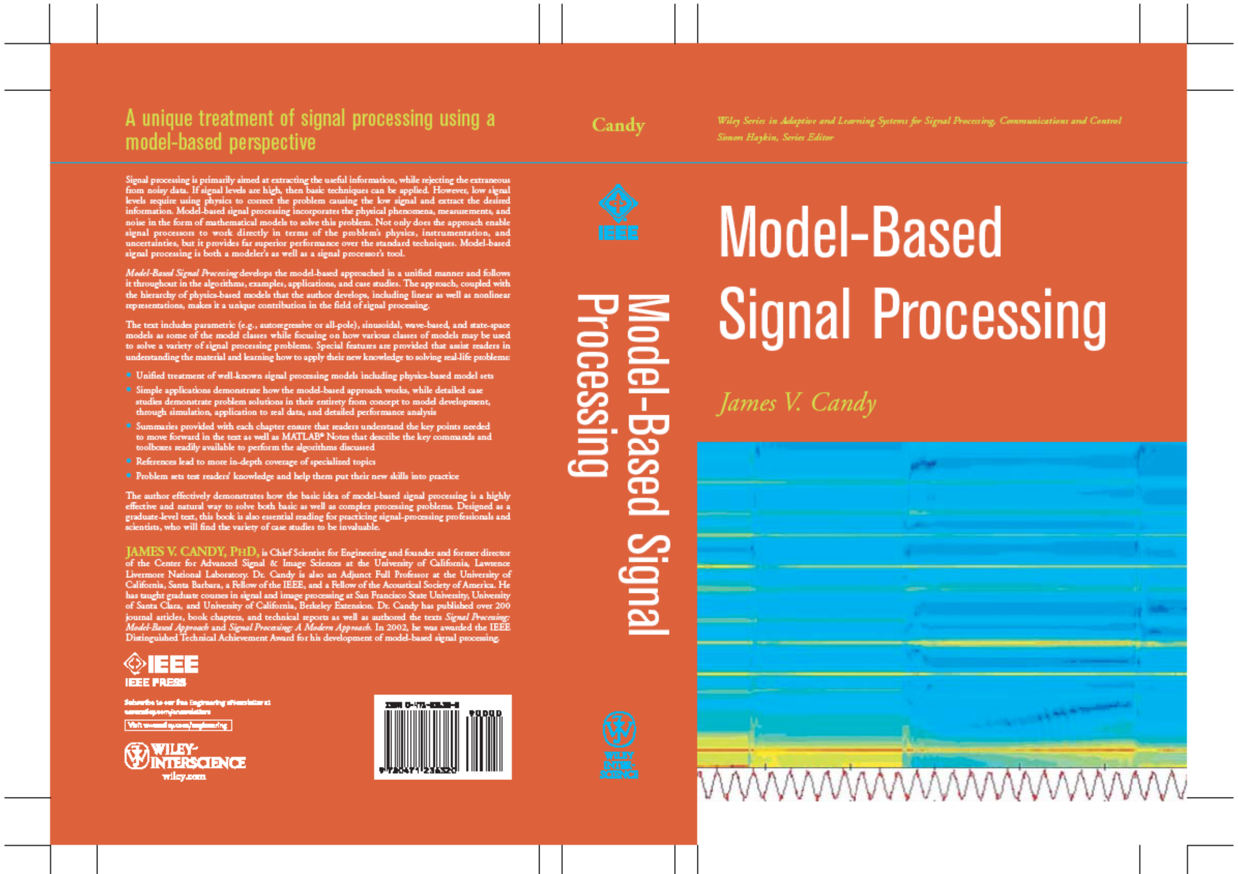




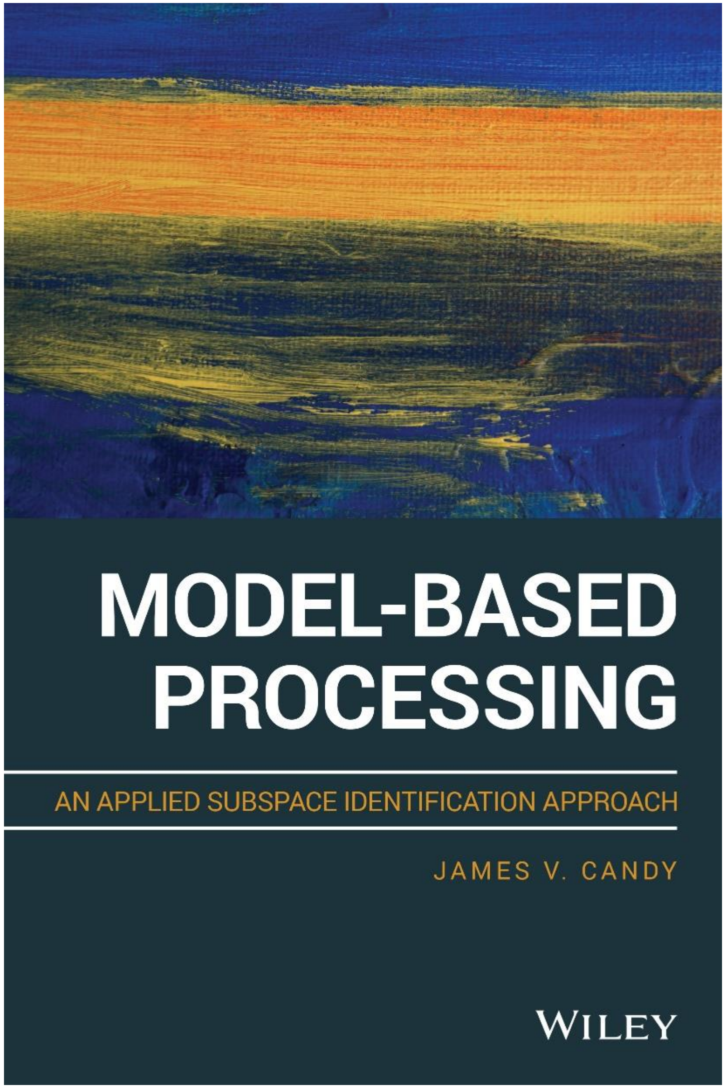
Text on Bayesian Processing



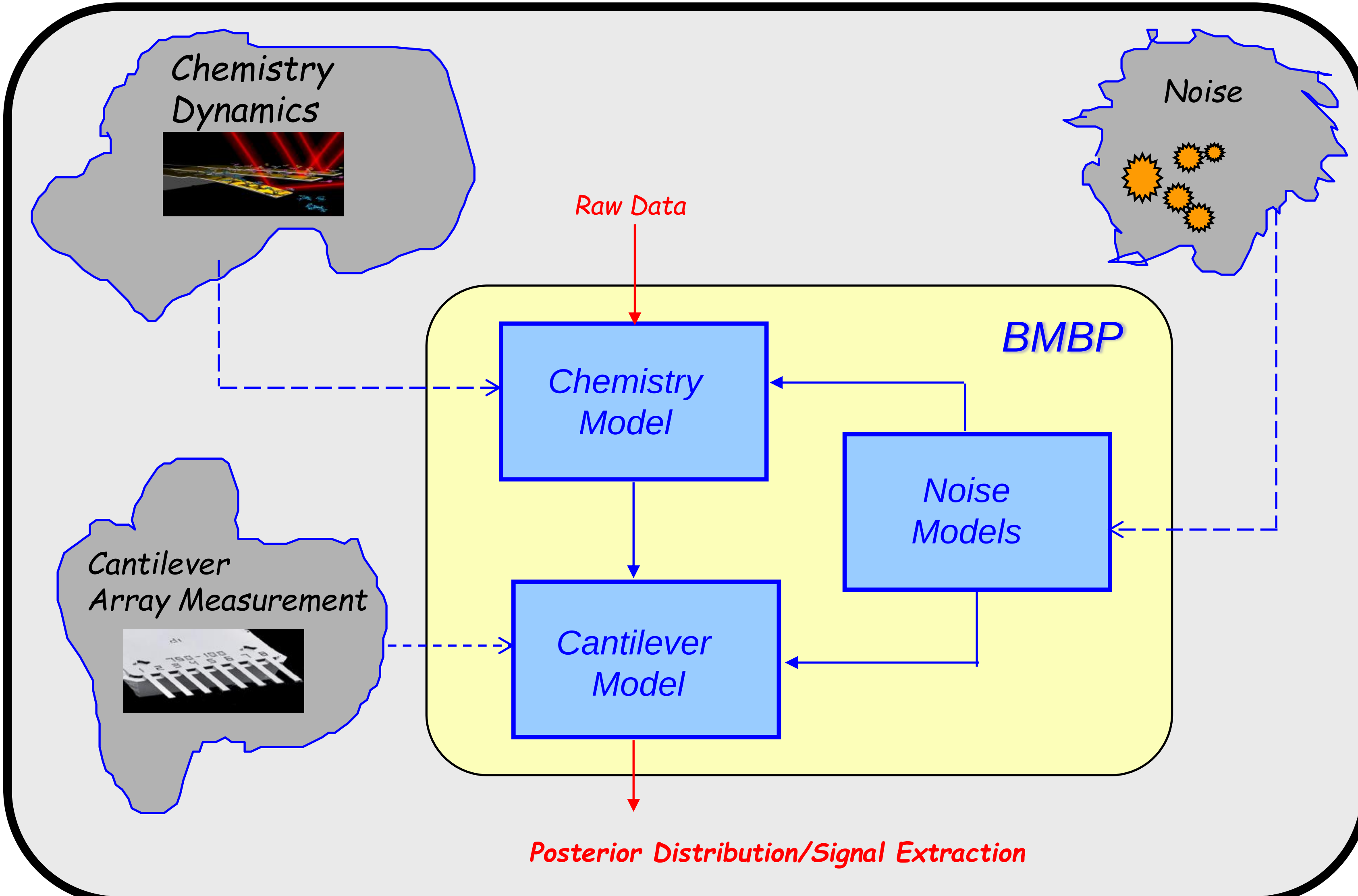
There is a textbook on MBSP (classical & modern nonlinear approaches & more …)



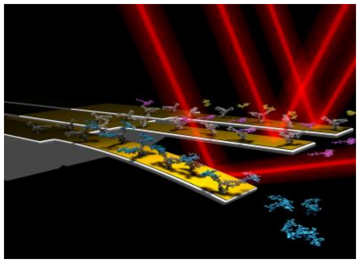
SUBSPACE techniques offer a viable and numerically robust way to obtain state-space models



In bio-threats, BMBP techniques can incorporate any "a priori" knowledge of the underlying physics into the processing scheme



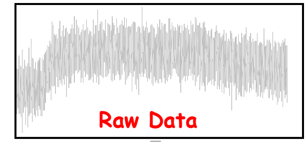
For this problem, *smart bio-sensors* incorporating a micro-cantilever array can be developed using the BMBP approach

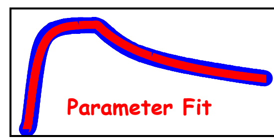


$$\Gamma(t)=\begin{cases} 0, & t<t_{\mathrm{ON}} \\ \left(\dfrac{c(t)}{c(t)+k_a/k_d}\right)\{1-\exp[-(k_a c(t)+k_d)(t-t_{ON})]\}, & t_{\mathrm{ON}}\le t\le t_{\mathrm{OFF}} \\ \sqrt{\dfrac{1}{2k_d(t-t_{OFF})}}, & t>t_{\mathrm{OFF}} \end{cases}$$

Concentration

PHYSICS





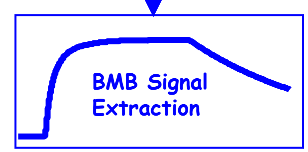
$\hat{\Theta}$

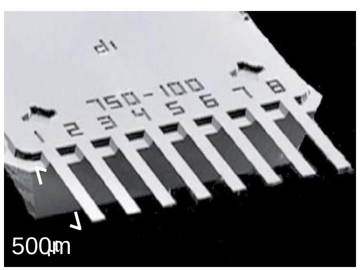
BAYESIAN MODEL-BASED PROCESSOR

$$\Delta\hat{G}(t\,|\,t-1)= \Delta\hat{G}(t-1\,|\,t-1) \quad \text{[Free Energy]}$$

$$\hat{\mathbf{y}}_\ell(t\,|\,t-1) = \beta_\ell\Gamma(t;\hat{\Theta})\Delta\hat{G}(t\,|\,t-1)+\Delta\hat{z}^T(t) \quad \text{[Deflection]}$$

$$\Pr[\Delta G\,|\,Y_N] \quad \text{[Posterior]}$$





$$y_\ell(t) = \beta_\ell\Gamma(t;\Theta)\Delta G(t)+\Delta z^T(t)+v_\ell(t)$$

for $\ell=1,\cdots,L$

Noise

MEASUREMENTS