NON-GAUSSIAN DATA ASSIMILATION METHODS FOR CLOUD MODELING

FSU | DEPARTMENT OF SCIENTIFIC COMPUTING
COLLEGE OF ARTS & SCIENCES

Hristo G. Chipilski
Assistant Professor, hchipilski@fsu.edu

1. INTRODUCTION AND MOTIVATION

- Many extreme weather events are connected to the physics taking place inside clouds (e.g., latent heat strengthens hurricanes).
- Simulation and forecasting of clouds and precipitation relies on numerical models, but accuracy is limited due to (i) incomplete knowledge and (ii) computational costs.
- Current practice is to **parameterize** the physics by modeling bulk properties such as cloud fraction, total mass, average droplet size, etc.

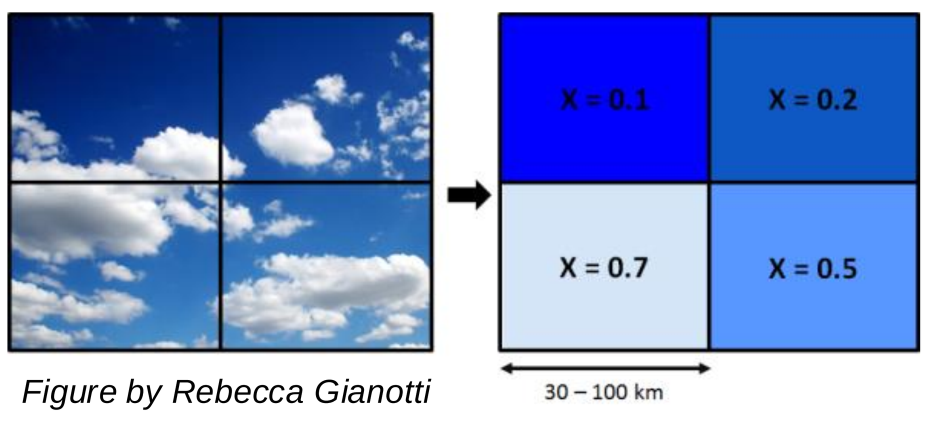


Figure by Rebecca Gianotti

- **Inverse problem**: Estimate the optimal model parameters from observations. The challenge is that existing methods are either not accurate enough or too expensive.

2. PARAMETER ESTIMATION METHODS

- Observations are modeled as $\mathbf{y} = \mathbf{h}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon}$. Here $\boldsymbol{\theta}$ are the set of true model parameters and $\boldsymbol{\varepsilon} \sim \mathrm{N}(\mathbf{0}, \mathbf{R})$ are the measurement errors.
- The parameter estimation task can be formulated as a Bayesian problem:

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\boldsymbol{\theta}) p(\mathbf{y}|\boldsymbol{\theta}).$$

- In this study, we will use several different **ensemble approaches** to solve Bayes' theorem: given the sample $\{\boldsymbol{\theta}^i_{\text{prior}}\}_{i=1}^{N_e}$ from $p(\boldsymbol{\theta})$, obtain $\{\boldsymbol{\theta}^i_{\text{post}}\}_{i=1}^{N_e}$ from $p(\boldsymbol{\theta}|\mathbf{y})$.

Ensemble Kalman filter (EnKF)

- Very popular method for both parameter and state estimation.
- Can handle high-dimensional problems but assumes the joint pdf $p(\boldsymbol{\theta}, \mathbf{y})$ is Gaussian. In view of this, the update is given by

$$\boldsymbol{\theta}^i_{\text{post}} = \boldsymbol{\theta}^i_{\text{prior}} + \mathbf{Cov}(\{\boldsymbol{\theta}^i_{\text{prior}}\}, \{\mathbf{y}^i\})\, \mathbf{Cov}(\{\mathbf{y}^i\}, \{\mathbf{y}^i\})^{-1} (\mathbf{y} - \mathbf{y}^i)$$

Ensemble Quadrature Filter (EnQF)

- Extends the linear update of EnKF by adding a quadratic term.
- Inclusion of higher-order moments improves the mean estimates in the case of skewed distributions.

Ensemble Conjugate Transform Filter (ECTF)

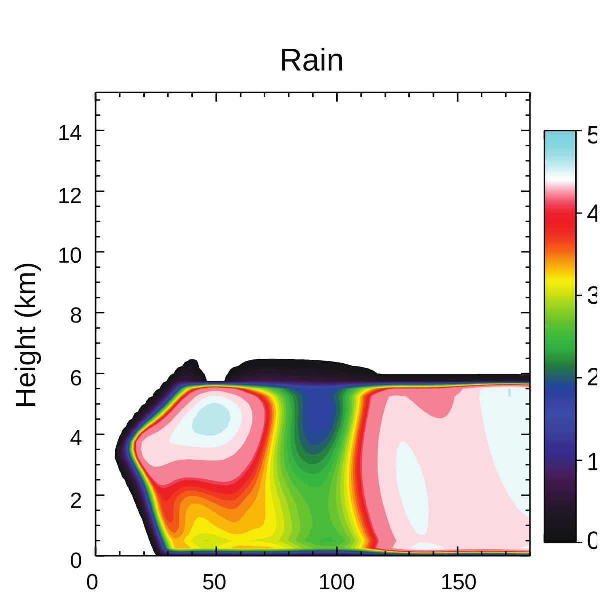
- Recently, the author developed a new framework which generalizes the Kalman filter to arbitrarily non-Gaussian distributions (scan QR code to learn more). Posterior is available in exact form and given by

$$p(\boldsymbol{\theta}|\mathbf{y}) = \mathbf{f}_{\boldsymbol{\Phi}_{\text{prior}}} \sharp \phi(\boldsymbol{\mu}_{\text{post}}, \boldsymbol{\Sigma}_{\text{post}})$$

- $\mathbf{f}_{\boldsymbol{\Phi}_{\text{prior}}}$ can be estimated via MLE from $\{\boldsymbol{\theta}^i_{\text{prior}}\}_{i=1}^{N_e}$, but initial tests use fixed choices (ECTF-ST) which reflect different parameter bounds: (i) $f(x) = \exp(x)$ if θ>0 and (ii) the standard logistic function $f(x) = \frac{1}{1+\exp(-x)}$ if $\theta \in (0,1)$.
- The key advantage is that $\{\boldsymbol{\mu}_{\text{post}}, \boldsymbol{\Sigma}_{\text{post}}\}$ can be obtained with a standard EnKF solver (since we need to approximate a Gaussian).

2. IDEALIZED CLOUD MODEL

- It is generally desirable to test new algorithms like ECTF in idealized and computationally-efficient settings first.
- The experiments in this study are based on a **simplified 1D (column) version of a large NASA model** used in real simulations of clouds and precipitation.
- The 1D model is **driven by** (i) prescribed profiles of temperature $T(z)$ and WV mixing ratio $q_{wv}(z)$ as well as (ii) time-varying profiles of vertical velocity $\mathrm{w}(z,t)$ and WV tendency $q'_{wv}(z)$.
- 3-hour model output to the right shows the rain mixing ratio $q_r(z,t)$, and represents a realistic depiction of a **mesoscale convective system** (MCS).



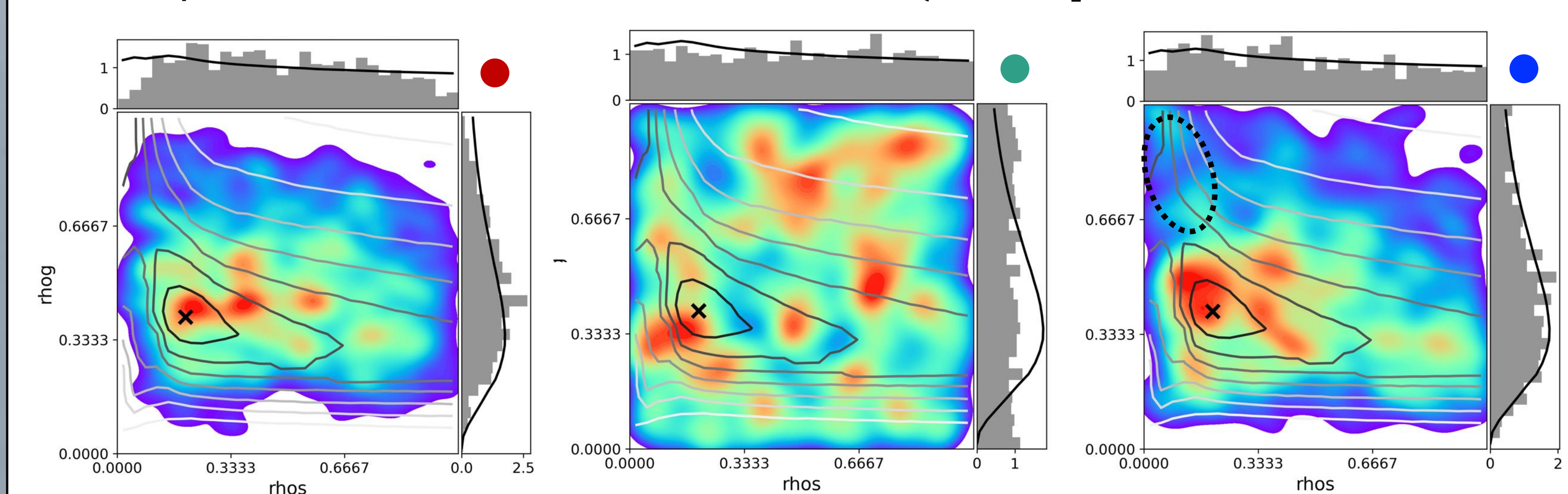
4. RESULTS

Data processing

- Posterior ensemble $\{\boldsymbol{\theta}^i_{\text{post}}\}_{i=1}^{N_e}$ from each filter is evaluated against the true Bayesian solution (obtained numerically).
- A rejection sampling procedure is used to post-process the raw filtering output and remove unrealistic members that violate physical and statistical bounds.

`EXP 1`: Unimodal posterior with a broad uncertainty range

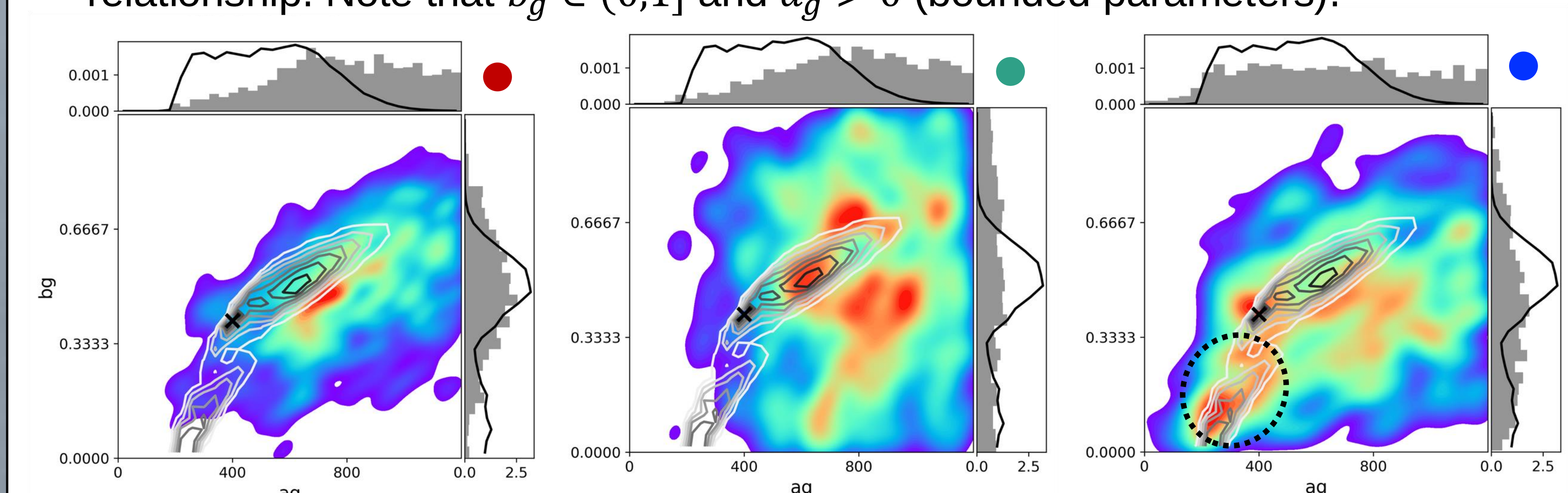
- Observe $\{\mathrm{OLR}, \mathrm{OSR}\}$ (outgoing LW and SW radiation) at t=120min with $\mathbf{R} = \{10{,}20\}$ Wm^{-2}.
- Inverse problem: Estimate the density of snow and graupel $\{\rho_s, \rho_g\}$. Note that both parameters are bounded in the interval $(0, 0.917]$.



- Reasonable performance by **EnKF**.
- Despite its nonlinear update, **EnQF** makes the results worse: large fraction of the original posterior members violate the parameter bounds.
- By contrast, **ECTF-ST** is very accurate and slightly better than EnKF.

`EXP 2`: Multimodal posterior with a narrow uncertainty range

- Observe $\{\mathrm{OLR}, \mathrm{OSR}\}$ at t=120min with $\mathbf{R}$=5 Wm^{-2}.
- Inverse problem: Estimate $\{a_g, b_g\}$, parameters in the graupel v_{fall}/D relationship. Note that $b_g \in (0,1]$ and $a_g > 0$ (bounded parameters).



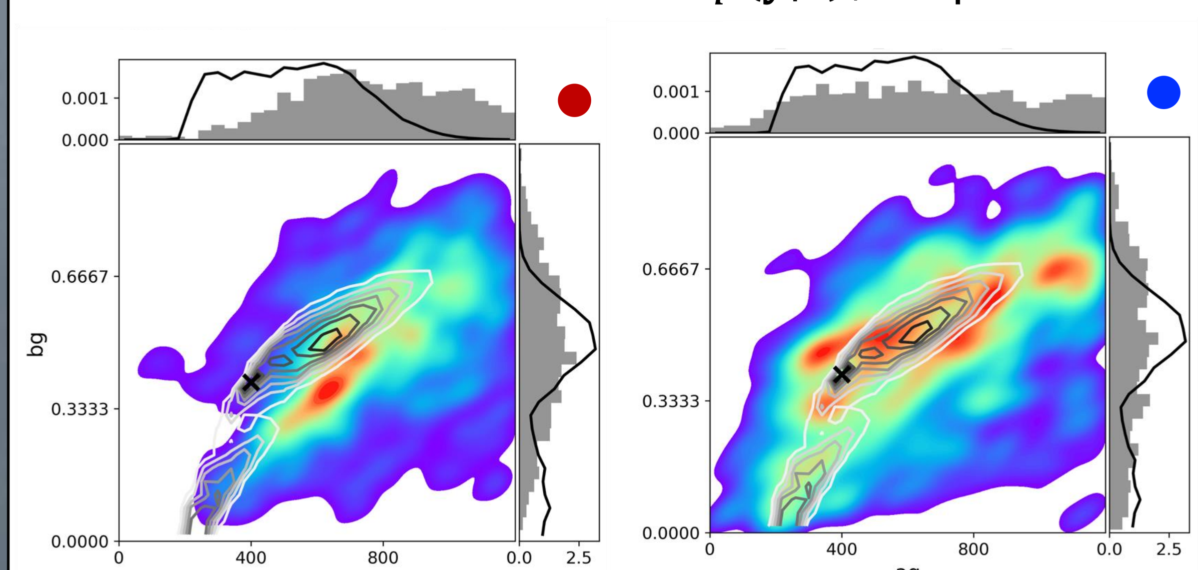
- Visible degradation in **EnKF**: single vs. double mode, large number of members outside the Bayesian posterior.
- **EnQF** has a better handle on the primary mode, but larger error in variance.
- While **ECTF-ST** solution is not ideal, analysis members cover both modes.

Likelihood tempering

- A more gradual incorporation of the observations $\mathbf{y}$ can further improve ECTF.
- Idea: Similar to the Ensemble Kalman Inversion (EKI), perform an *iterative update* by tempering the likelihood such that

$$p(\mathbf{y}|\mathbf{x}) = \prod_k p(\mathbf{y}|\mathbf{x})^{\alpha_k} \text{ with } \sum_k \alpha_k = 1.$$

- In our case with a Gaussian $p(\mathbf{y}|\mathbf{x})$, we perform k updates with $\mathbf{R}_k \leftarrow \mathbf{R}/\alpha_k$.



k=2 case

- **EnKF** (EKI) changes only slightly relative to k=1.
- **ECTF-ST** analysis visibly improved in terms of primary mode and narrower ensemble → conjugate approximation satisfied better when k=2.

5. SUMMARY AND NEXT STEPS

- **ECTF** shows promises for the efficient estimation of cloud parameters.
- Next, the algorithm will be tested with **learnable transformations**; e.g., the *rational splines* shown to the right, which can approximate complex distributions.
- Will also consider more **methods**, **experimental settings** and **objective skill metrics**.