Nonlinear ensemble filtering with diffusion models: Application to the SQG model

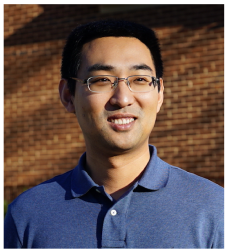
AMS Annual Meeting/IOAS-AOLS, 14 Jan 2025

Hristo G. Chipilski
Assistant Professor
hchipilski@fsu.edu

FSU | DEPARTMENT OF SCIENTIFIC COMPUTING
COLLEGE OF ARTS & SCIENCES

Research team

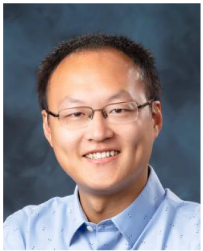
Feng Bao
(FSU)



Siming Liang
(FSU)



Guannan Zhang
(ORNL)



Jeffrey Whitaker
(NOAA PSL)



Ensemble DA: The two ends of the spectrum

Particle filters (PFs)

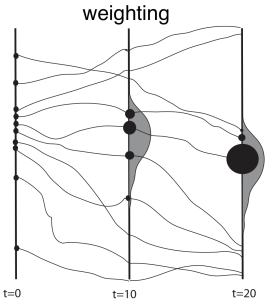
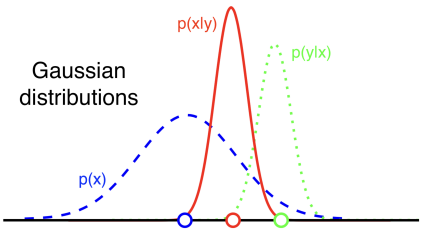


Figure: Van Leeuwen (2009)

- Solve Bayes' theorem exactly.
- Hard to implement with small ensembles.

Ensemble Kalman filters (EnKFs)



- Widely used in the field (low cost, stable).
- Gaussian assumptions can be problematic.

Generative AI

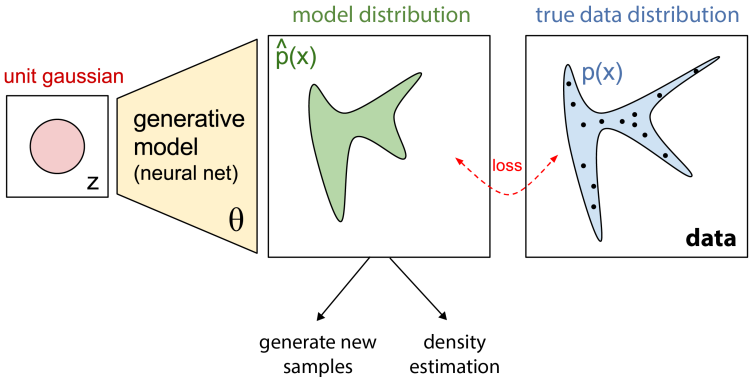


Figure: OpenAI (2016).

- Appealing framework for ensemble DA (e.g., better approximation of prior distributions).

2 new GenAI-inspired ensemble filtering frameworks

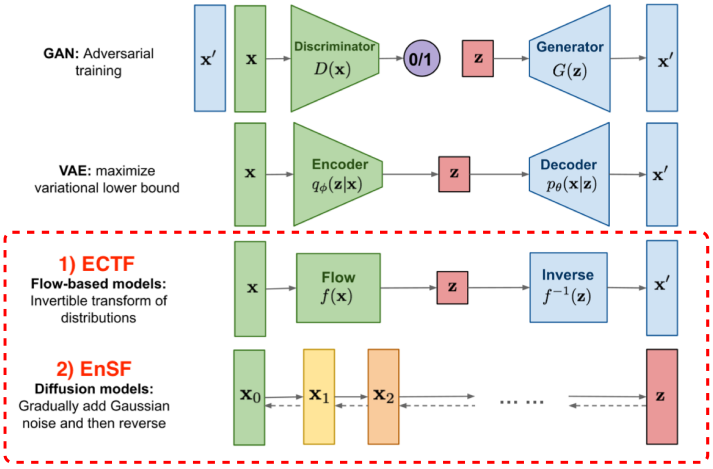
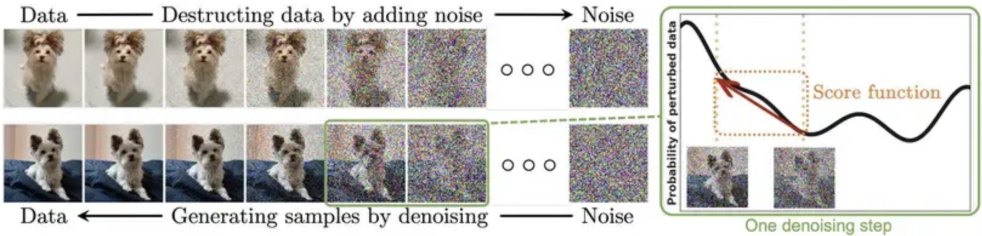


Figure: OpenAI (2016).

1. Ensemble Conjugate Transform Filter (**ECTF**)
   - EOR to appear in JAS soon.
2. Ensemble Score Filter (**EnSF**)
   - Focus of this talk.

Diffusion models



- **Idea**: Start with an image and progressively add noise to it until the blurred image becomes unrecognizable.
- The goal of diffusion models is to **learn how to reverse this process**; i.e., how to effectively denoise the corrupted image.
  - The **score function** $\mathbf{s}(\mathbf{x}) \coloneqq \nabla_{\mathbf{x}} \log p_{\text{data}}(\mathbf{x})$ plays a key role in restoring the image.

Training-free score estimation

- The standard technique to calculate $\mathbf{s}(\mathbf{z}_t, t)$ is known as *score matching* and uses a **neural network**.
  - Not suitable for ensemble DA where observations are assimilated frequently.

- Instead, we propose a **training-free approach** in which the score function is estimated cheaply (on the fly):

$$\hat{\mathbf{s}}(\mathbf{z}_t, t) = \sum_{i=1}^{N} -\frac{\mathbf{z}_t - \alpha_t \mathbf{x}^i}{\beta_t^2} \hat{\mathbf{w}}(\mathbf{z}_t, \mathbf{x}^i).$$

  - $\{\mathbf{x}^i\}_{i=1}^N$ is a set of ensemble members (the forecast ensemble in EnSF).
  - $\mathbf{z}_t$ is the dependent variable in the diffusion equations.
  - α_t and β_t are some fixed coefficient functions.

Analysis step

- Let $\hat{\mathbf{s}}_{k|k-1} := \nabla_{\mathbf{x}} \log p(\mathbf{x}_k|\mathbf{y}_{1:k-1})$ be the **prior score** estimated from the forecast ensemble $\{\mathbf{x}^i_{k|k-1}\}_{i=1}^N$.
- From Bayes' theorem, the posterior is $p(\mathbf{x}_k|\mathbf{y}_{1:k}) \propto p(\mathbf{x}_k|\mathbf{y}_{1:k-1})p(\mathbf{y}_k|\mathbf{x}_k) \Rightarrow$ the **posterior score** at intermediate times in the reverse diffusion process can be written as

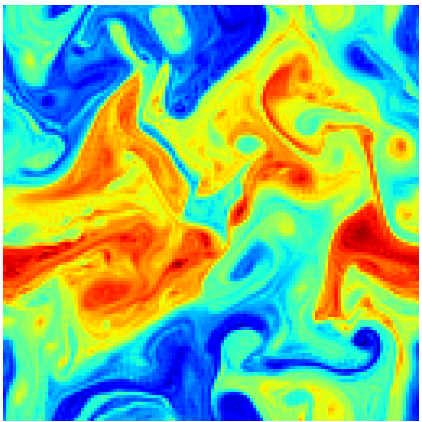
$$\hat{\mathbf{s}}_{k|k}(\mathbf{z}_t, t) = \hat{\mathbf{s}}_{k|k-1}(\mathbf{z}_t, t) + h(t)\nabla_{\mathbf{z}_t} \log p(\mathbf{y}_k|\mathbf{z}_t).$$

  - $h(t)$ is a damping function $\rightarrow$ gradually incorporate observational information in the reverse diffusion process.

- **EnSF update**: At the terminal pseudo-time $t = T$, (1) draw a standard Gaussian sample $\{\mathbf{z}(t = T)^i \sim \mathcal{N}(\mathbf{0}, \mathbf{I})\}_{i=1}^N$ and (2) solve the reverse diffusion equations using the estimated posterior score $\hat{\mathbf{s}}_{k|k}(\mathbf{z}_t, t)$. At $t = 0$, we get the desired **analysis ensemble**:

$$\{\mathbf{x}^i_{k|k}\}_{i=1}^N \leftarrow \{\mathbf{z}(t = 0)^i\}_{i=1}^N.$$

Surface-quasigeostrophic model

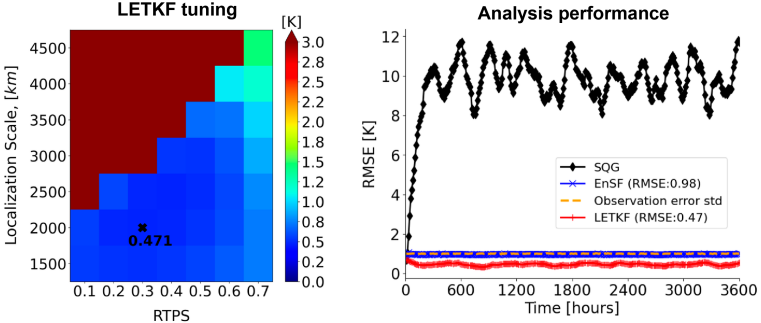


Animation of θ on upper model surface.

- Simulates the dynamics of potential temperature θ on top and bottom surfaces.
- **Why this model?** Produces turbulence spectra close to that of real geophysical flows.
- GitHub model repo from Jeff Whitaker:

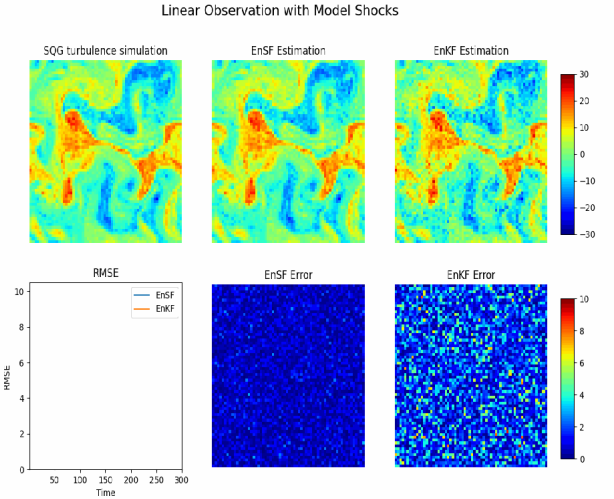
Full linear observations + perfect model

- Using LETKF as an EnKF reference (comes with Jeff's SQG package).
- First experiment is highly idealized → LETKF yields optimal performance.



- **Important finding**: However, EnSF remains close to LETKF and maintains filtering stability despite the lack of any localization and inflation tuning.

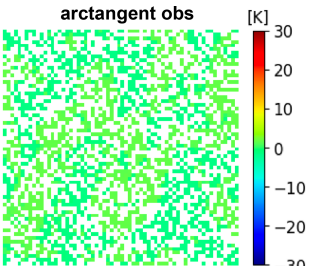
Full linear observations + imperfect model

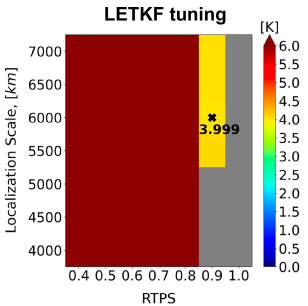


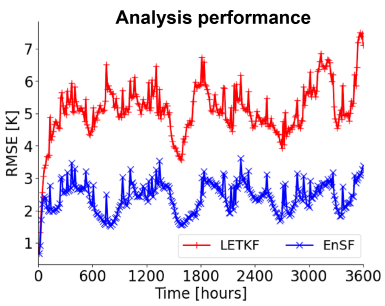
- Still having linear observations of the full state, but **unknown model errors** are now added to the nature run.
- Despite this nearly-Gaussian setting, **LETKF** is very sensitive to model errors $\rightarrow$ **progressively worse** analysis estimates.

Partial nonlinear observations + imperfect model

- The most challenging setting where we expect LETKF to be suboptimal.







- **Key points**: (i) EnSF is consistently better. Moreover, (ii) LETKF avoids filter divergence only over a limited set of localization and inflation values.

Summary and outlooks

- The SQG experiments highlight EnSF's potential for meteorological applications, especially for assimilating highly **nonlinear observations** (e.g., satellite radiances).
- In another recent study from our group (*see Guannan Zhang's talk J12.2 at 5pm tomorrow, La Nouvelle C*), we have also shown that EnSF can be successfully coupled with **data-driven numerical models** (specifically, a ViT surrogate of the SQG dynamics).
- There are many interesting future directions, including optimal **localization** and **inflation** strategies, treatment of **sparse observations**, and tests with even **more realistic models** (e.g., SPEEDY, OpenIFS).
- EnSF papers:

EnSF SQG

EnSF ViT

SDE-based diffusion: Theory from Song et al. (2021)

- Forward diffusion process (adding noise to an image) is modeled as the solution to an **Itô stochastic differential equation (SDE)**:

$$d\mathbf{z} = \mathbf{f}(\mathbf{z}, t)dt + g(t)d\mathbf{w}.$$

  - Here $\mathbf{w}$ is the standard Brownian motion, $\mathbf{f}$ is the vector-valued drift coefficient of $\mathbf{z}$, and g is a scalar function (could also be a matrix) known as the diffusion coefficient of $\mathbf{z}$.
  - Appropriate choices for $\mathbf{f}$ and g ensure that the terminal random variable $\mathbf{z}$ will be a sample from a standard Gaussian distribution (and that the SDE has a unique strong solution).

- Backward diffusion process described according to a **reverse-time SDE**:

$$d\mathbf{z} = \left[\mathbf{f}(\mathbf{z}, t) - g(t)^2 \nabla_{\mathbf{z}} \log p(\mathbf{z}, t)\right] dt + g(t)d\bar{\mathbf{w}}.$$

  - Here $\bar{\mathbf{w}}$ is another standard Brownian motion but with time going backwards, and $\mathbf{s}(\mathbf{z}, t) := \nabla_{\mathbf{z}} \log p(\mathbf{z}, t)$ is the score function.
  - We have the same drift and diffusion coefficients as the forward SDE!

SDE-based diffusion: Illustration from Song et al. (2021)

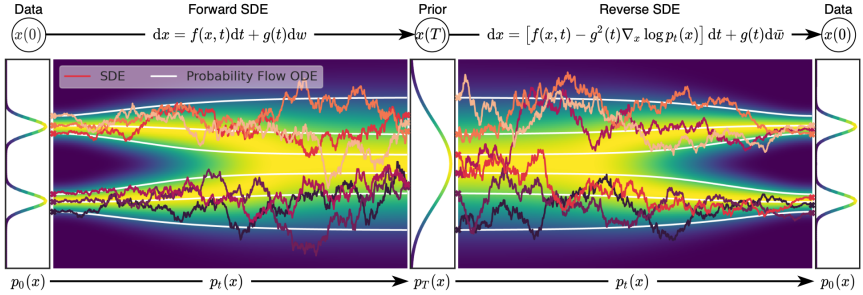


Figure 2: **Overview of score-based generative modeling through SDEs**. We can map data to a noise distribution (the prior) with an SDE (Section 3.1), and reverse this SDE for generative modeling (Section 3.2). We can also reverse the associated probability flow ODE (Section 4.3), which yields a deterministic process that samples from the same distribution as the SDE. Both the reverse-time SDE and probability flow ODE can be obtained by estimating the score $\nabla_{\mathbf{x}} \log p_t(\mathbf{x})$ (Section 3.3).

- We implement the reverse-time SDE numerically using an **Euler scheme**.

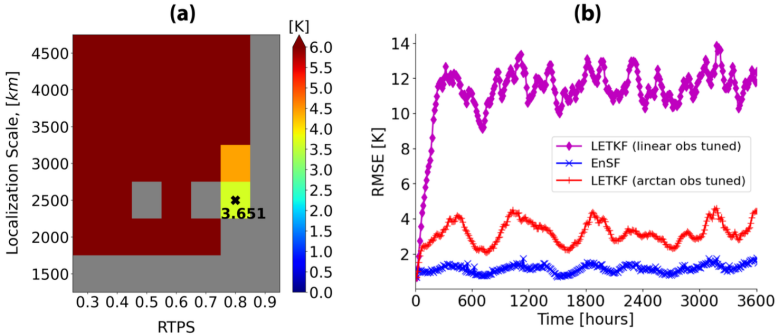
All experimental settings

Linear observations	**Nonlinear observations**
EXP_L1: fully observed state	EXP_NL1: fully observed state
EXP_L2: fully observed state subject to unknown model errors	EXP_NL2: partially observed state subject to unknown model errors

TABLE 1. A summary of the 4 main experimental settings used to compare EnSF and LETKF in Section 4. The first column lists experiments in which state variables are directly observed, whereas experiments in the second column use the nonlinear arctangent operator. Further note that EXP_L1 contains more than one numerical test (e.g., using the SQG model at different resolutions).

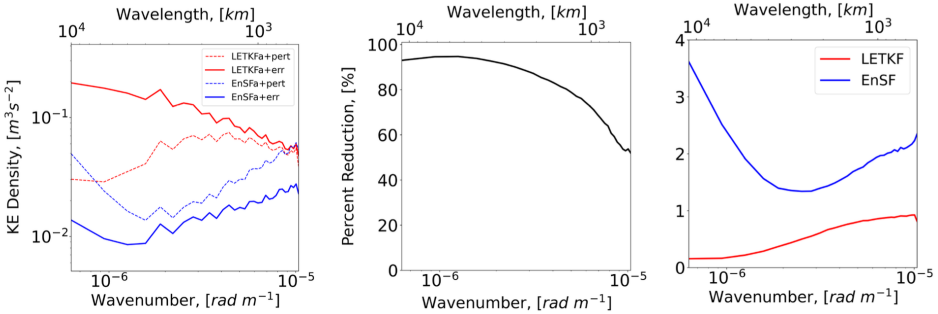
- Observation operator: $\mathbf{h}(\mathbf{x}) = \mathbf{x}$ [linear case] and $\mathbf{h}(\mathbf{x}) = \arctan(\mathbf{x})$ [nonlinear case].
- Imperfect model: $\mathbf{x} \mapsto \mathbf{f}(\mathbf{x}) + \boldsymbol{\xi}$, independent errors $\boldsymbol{\xi} \sim \mathcal{N}(\mathbf{0}, \mathbf{D})$, $\boldsymbol{\xi} = \sum_{i=1}^{4} \boldsymbol{\xi}_i$, sum of 4 error processes (from high frequency/low amplitude noise to low frequency/high amplitude noise).

EXP_NL1



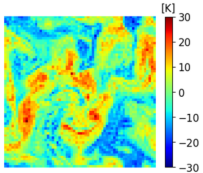
- Model state is **observed fully** but **nonlinearly**: $\mathbf{h}(\mathbf{x}) = \arctan(\mathbf{x})$.
- Perfect model assumption (**no model errors**): $\boldsymbol{\xi} = \mathbf{0}$ (a.s.).

KE spectra in EXP_NL2

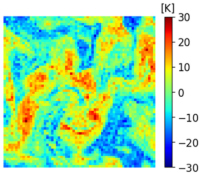


- EnSF has consistently lower analysis errors, biggest improvements occur at **large scales**.
- LETKF is **underdispersive**, especially at large scales, while EnSF – **overdispersive** (RTPS=1 in all experiments) $\rightarrow$ motivates future work into optimal EnSF configurations.

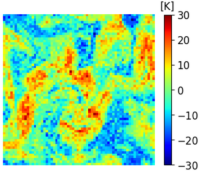
Model errors: Visualization



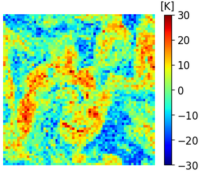
(a) Truth



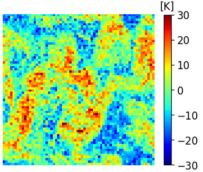
(b) 20% noise



(c) 30% noise



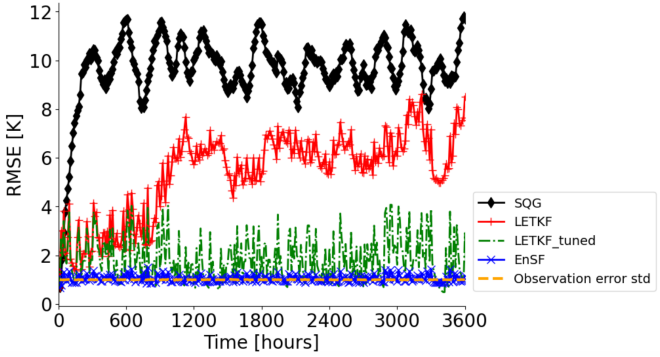
(d) 40% noise



(e) 50% noise

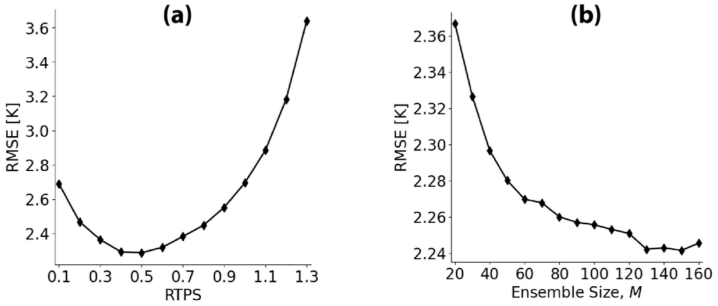
- Panels (b)-(e) show example model error-contaminated SQG fields related to the **4 error processes** $\{\boldsymbol{\xi}_i\}_{i=1}^4$ considered in our imperfect model experiments.

Model errors: Tuned LETKF



- LETKF_tuned assumes that the high frequency/low amplitude model errors $\boldsymbol{\xi}_1$ are in the training dataset.
- This leads to improved performance, but the presence of model errors still causes **spikes in the analysis performance**.

Preliminary EnSF sensitivities: Inflation and ensemble size (for EXP_NL2)



- (a) All EnSF experiments use a suboptimal RTPS $= 1$; better performance can be achieved through a **carefully tuned inflation** $\rightarrow$ more research needed in this direction.
- (b) Less EnSF sensitivity to the **ensemble size** – on one hand this suggests filtering stability, but we need to further improve the rates of error decrease.