

# Technical Talk - PhD Project

Kamal Sharma

Department of Mathematics, University of Hamburg

February 13, 2026



## Introduction

---

# What is it about?

Development of a **structure-preserving idealized stochastic climate model**

- ▶ **Climate model:** set of equations modelling coupled ocean-atmosphere dynamics
- ▶ **Idealized:** 2D model → solves for velocity, temperature, and pressure
- ▶ **Stochastic:** has stochastic/random terms (think of Brownian motion)
- ▶ **Structure-preserving:** preserves underlying geometric and physical structure

# What is it about?

Development of a **structure-preserving idealized stochastic climate model**

- ▶ **Climate model:** set of equations modelling coupled ocean-atmosphere dynamics
- ▶ **Idealized:** 2D model → solves for velocity, temperature, and pressure
- ▶ **Stochastic:** has stochastic/random terms (think of Brownian motion)
- ▶ **Structure-preserving:** preserves underlying geometric and physical structure

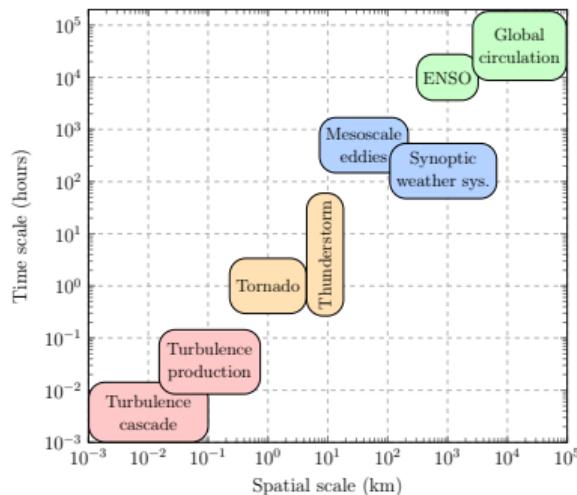
**We are solving stochastic PDEs which model ocean-atmosphere interactions!**

## Introduction

---

Why add stochastic terms?

# Multiscale phenomena

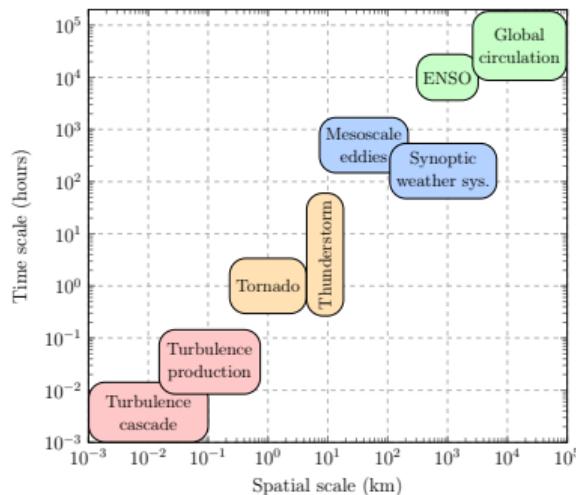


Spatial and temporal scales of various meteorological phenomena<sup>1</sup>

- ▶ Numerical models have fixed resolution
- ▶ Processes occurring below the model resolution are not captured → leads to prediction error!

<sup>1</sup>Roland B. Stull. *Practical Meteorology: An Algebra-Based Survey of Atmospheric Science*. 2017.

# Multiscale phenomena



Spatial and temporal scales of various meteorological phenomena<sup>1</sup>

- ▶ Numerical models have fixed resolution
- ▶ Processes occurring below the model resolution are not captured → leads to prediction error!
- ▶ **Solution:** Parameterization or subgrid-scale modeling

<sup>1</sup>Roland B. Stull. *Practical Meteorology: An Algebra-Based Survey of Atmospheric Science*. 2017.

# Parameterization

- ▶ Additional terms are added to account for the missing effect of unresolved/small scales
- ▶ Parameterization techniques can be divided into two categories:
  1. Deterministic
  2. Stochastic
- ▶ We explored **stochastic parameterization** : stochastic advection by Lie transport (SALT)

# Parameterization

- ▶ Additional terms are added to account for the missing effect of unresolved/small scales
- ▶ Parameterization techniques can be divided into two categories:
  1. Deterministic
  2. Stochastic
- ▶ We explored **stochastic parameterization** : stochastic advection by Lie transport (SALT)
- ▶ Why SALT?
  - ▶ preserves conservation properties → reliable long simulations
  - ▶ better forecast skills → uncertainty due to unresolved transport processes can be quantified
  - ▶ conducive to data-driven approaches → can be estimated from observations/satellite data

## Research objectives

---

# Main objectives

1. Explore the efficacy of stochastic parameterization (in terms of UQ skills) for a climate model
  - ▶ Previous studies focused on Euler, quasi-geostrophic, and shallow water equations
2. Numerically solve the idealized stochastic climate model

## The climate model

---

# Model equations

2D climate model equations:

$$\text{Atmosphere : } d\mathbf{u}^a + ((\mathbf{u}^a dt + \sum_i \boldsymbol{\xi}_i \circ dW^i) \cdot \nabla) \mathbf{u}^a + \frac{1}{Ro^a} \hat{\mathbf{z}} \times (\mathbf{u}^a dt + \sum_i \boldsymbol{\xi}_i \circ dW^i)$$

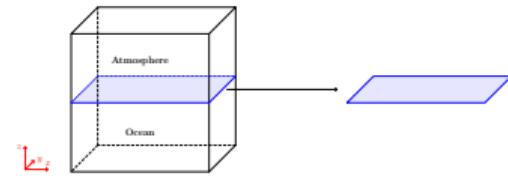
$$+ \sum_i (u_1^a \nabla \xi_{i,1} + u_2^a \nabla \xi_{i,2}) \circ dW^i = (-\frac{1}{C^a} \nabla \theta + \frac{1}{Re^a} \Delta \mathbf{u}^a) dt,$$

$$d\theta^a + \nabla \cdot (\theta^a (\mathbf{u}^a dt + \sum_i \boldsymbol{\xi}_i \circ dW^i)) = (\gamma(\theta^a - \theta^o) + \frac{1}{Pe^a} \Delta \theta^a) dt,$$

$$\text{Ocean : } \frac{\partial \mathbf{u}^o}{\partial t} + (\mathbf{u}^o \cdot \nabla) \mathbf{u}^o + \frac{1}{Ro^o} \hat{\mathbf{z}} \times \mathbf{u}^o + \frac{1}{Ro^o} \nabla p^a = \sigma(\mathbf{u}^o - \mathbb{E}\mathbf{u}_{sol}^a) + \frac{1}{Re^o} \Delta \mathbf{u}^o,$$

$$\nabla \cdot \mathbf{u}^o = 0,$$

$$\frac{\partial \theta^o}{\partial t} + \mathbf{u}^o \cdot \nabla \theta^o = \frac{1}{Pe^o} \Delta \theta^o.$$



Model domain

- $\mathbf{u}$  : velocity,
- $\theta$  : temperature,
- $p$  : pressure

Incompressible/compressible Navier-Stokes + Advection-diffusion equations

# How to get the stochastic terms?

- ▶ Difference between Lagrangian trajectories at different resolutions:

$$\sum_i \xi_i dW^i \approx \mathbf{u}_{true} dt - \mathbf{u} dt$$

$\mathbf{u}_{true}$  : true velocity

$\mathbf{u}$  : mean flow velocity i.e. the large-scale component of  $\mathbf{u}_{true}$

- ▶ Ways to estimate the true velocity  $\mathbf{u}_{true}$ :

1. Observation data (for example, from satellites)
2. Synthetic data (from high-resolution numerical simulations)

# How to get the stochastic terms?

- ▶ Difference between Lagrangian trajectories at different resolutions:

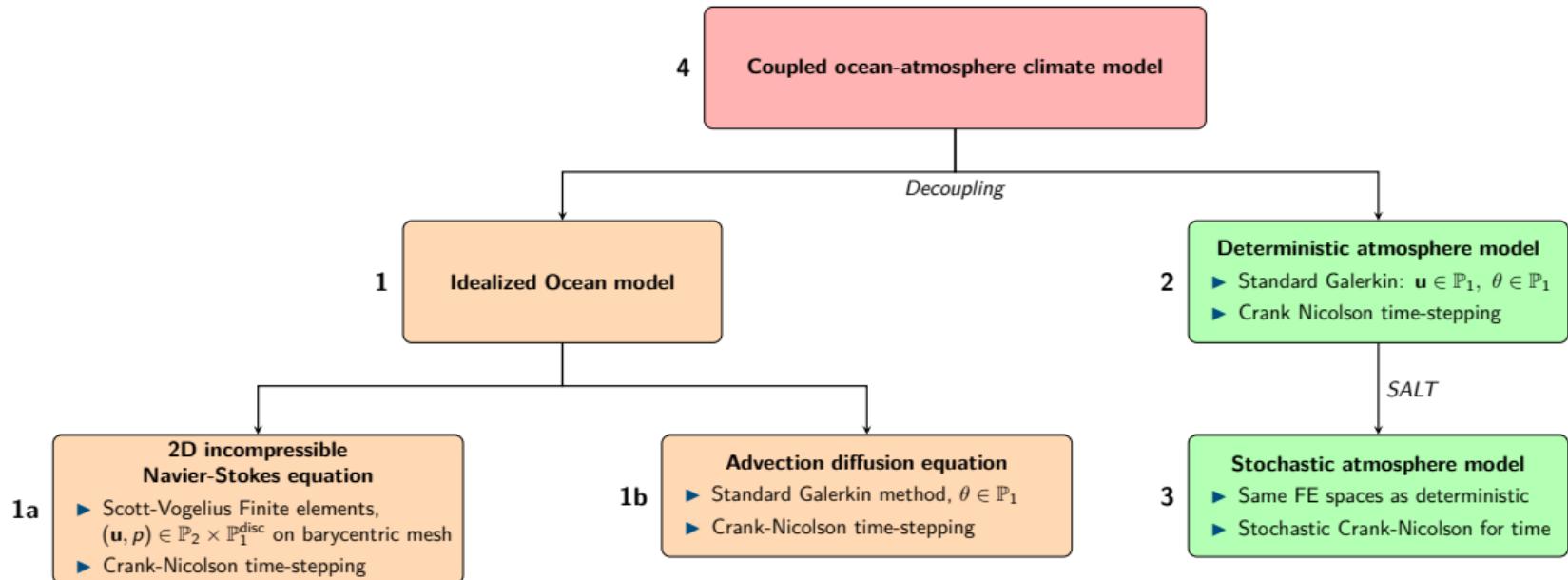
$$\sum_i \xi_i dW^i \approx \mathbf{u}_{true} dt - \mathbf{u} dt$$

$\mathbf{u}_{true}$  : true velocity

$\mathbf{u}$  : mean flow velocity i.e. the large-scale component of  $\mathbf{u}_{true}$

- ▶ Ways to estimate the true velocity  $\mathbf{u}_{true}$ :
  1. Observation data (for example, from satellites)
  2. Synthetic data (from high-resolution numerical simulations)
- ▶ **We use synthetic data (proof of concept)!**

# Our methodology



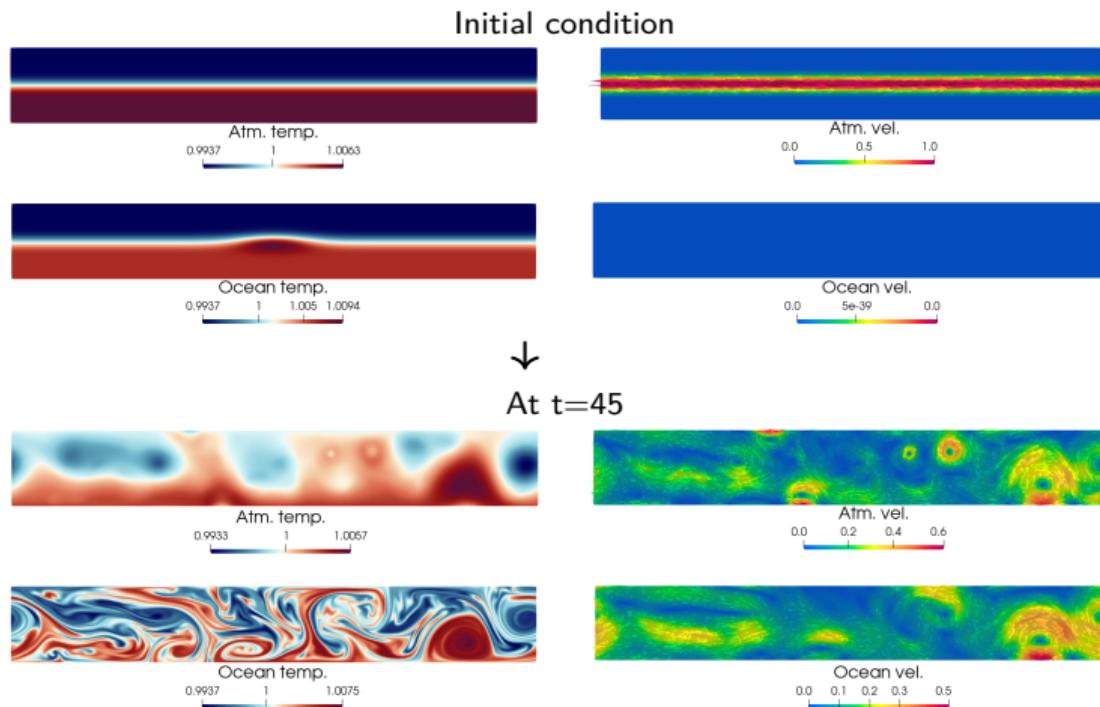
Incremental validation approach - solve simpler components first, verify at each step!

## Numerical simulation of the climate model

---

# Data collection and model calibration

- We run the deterministic climate model in high-res. ( $\Delta x = 1/128 \sim 30$  km)



- Analyze data → extract small-scale features (using statistical algorithms) → obtain  $\sum_i \xi_i dW^i$

## Numerical simulation of the climate model

---

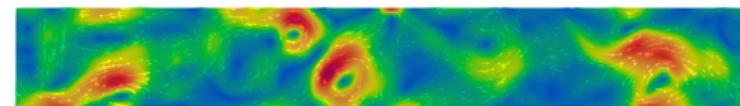
Stochastic model simulation

# Stochastic model simulation

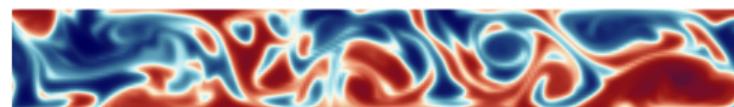
- We run the stochastic model on a coarse grid ( $\Delta x = 1/32 \sim 120$  km)
- Initial conditions correspond to coarse grained velocity and temperature fields at  $t = 25$



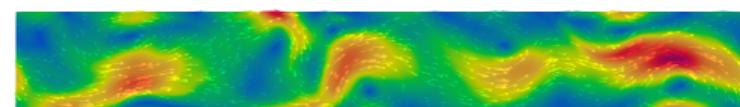
Atm. temp.  
0.9934 1 1.0054



Atm. vel.  
0.0 0.2 0.4 0.7

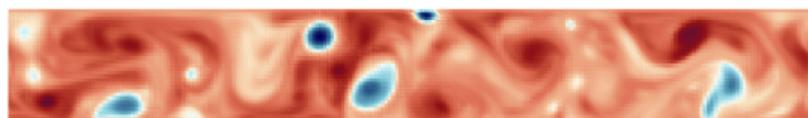


Ocean temp.  
0.9938 1 1.0072



Ocean vel.  
0.0 0.1 0.2 0.3 0.4

Atmospheric vorticity at  $t = 25$

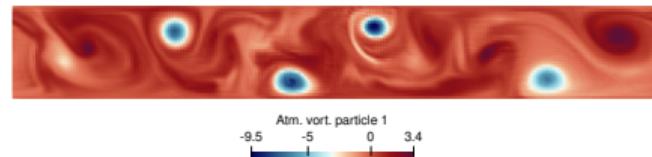


Atm. vort.  
-12.1 -5 0 6.4

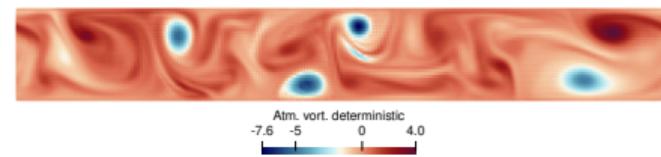
# Simulation results

- ▶ Stochastic model is simulated for 20 time units ( $t = 25$  to  $t = 45$ )
- ▶ Stochastic vs. deterministic (without parameterization) model results at  $t = 35$

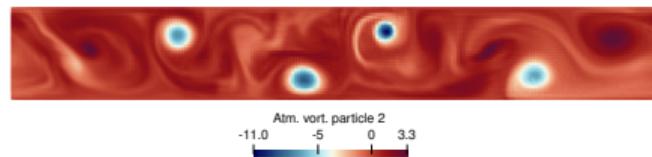
Particle 1



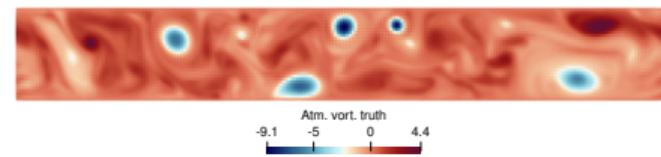
Deterministic solution



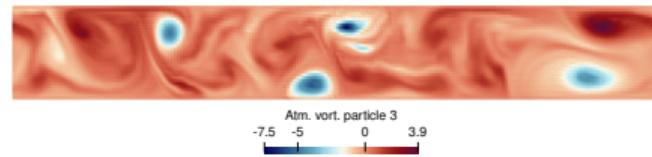
Particle 2



True solution



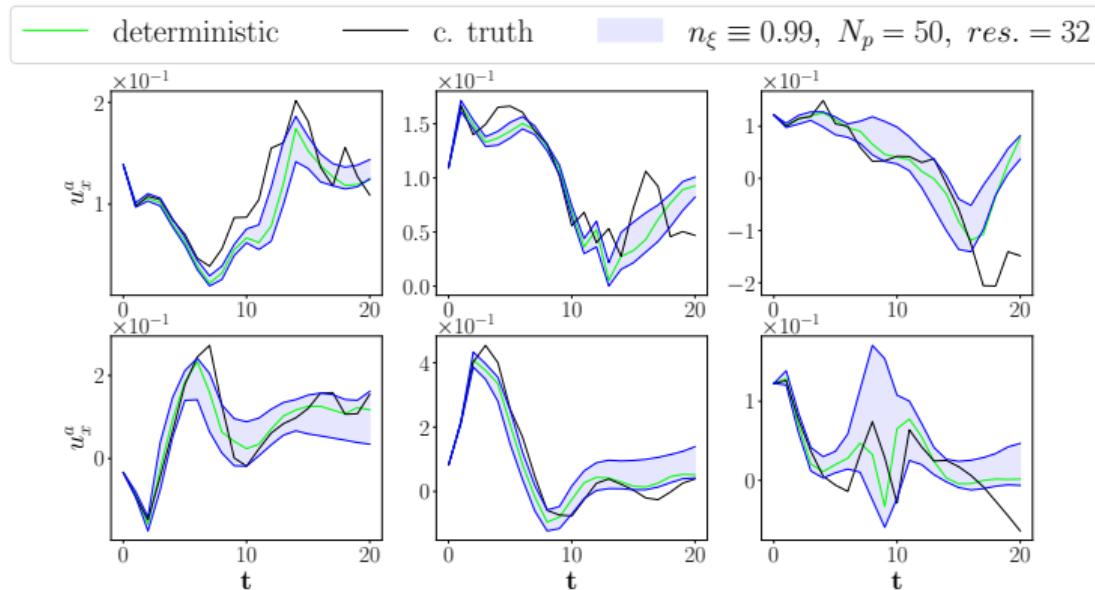
Particle 3



# Simulation results

- ▶ Comparison between the stochastic ( $\mu \pm 1\sigma$ ) and deterministic sol. at 6 grid points:

**Atmospheric velocity (x component)**

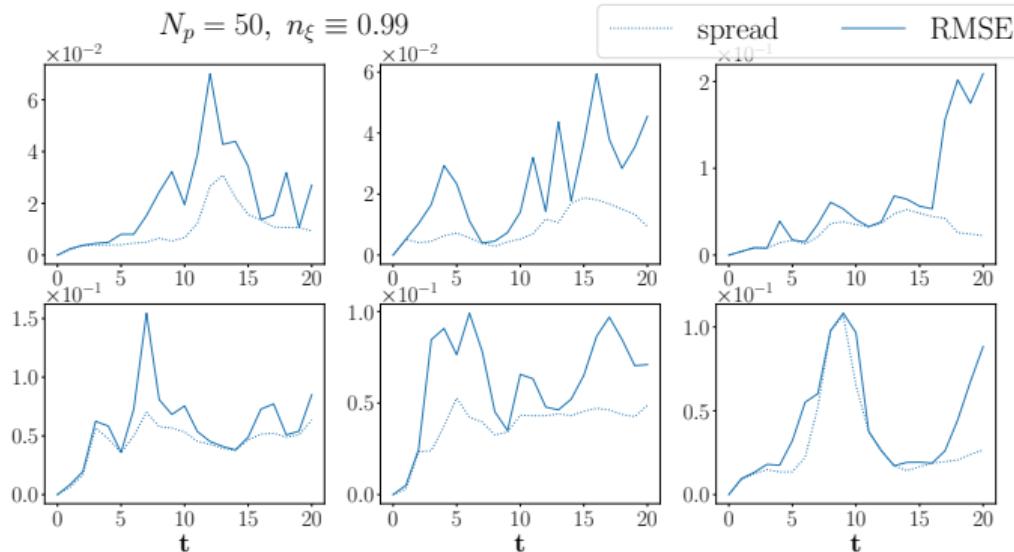


1. Ensemble spread increases over time
2. Stochastic model captures the truth for 5 to 10 time units

# Uncertainty quantification

- ▶ Plots for the evolution of ensemble spread and RMSE at six different locations on the grid

**RMSE vs. spread for  $u_x^a$**



- ▶ Ensemble spread  $\propto$  RMSE (for at least 15 time units)
- ▶ Error in the stochastic solution can be estimated by its own spread

## Summary

---

# Summary & outcomes

- ▶ **Aim:** solve a stochastic coupled ocean-atmosphere climate model
- ▶ **Novelty:** data-driven model
  - ▶ Stochastic terms (modelling unresolved processes) are obtained from observation data
  - ▶ Exhibit desired conservation properties
  - ▶ High-res. model sim. data is used to estimate  $\xi_i \rightarrow$  data analysis & statistical modelling
- ▶ **Methodology** for solving the model equations:
  - ▶ Solve simpler models first
  - ▶ build the climate model step-by-step
  - ▶ verify numerical schemes at each step
- ▶ **Results:**
  - ▶ Stochastic model solution captures the truth for some time units
  - ▶ Shows promising UQ test results: ensemble spread size  $\propto$  RMSE
- ▶ **Outcome:** first step towards using SALT in a highly complex coupled system!

## Additional slides

---

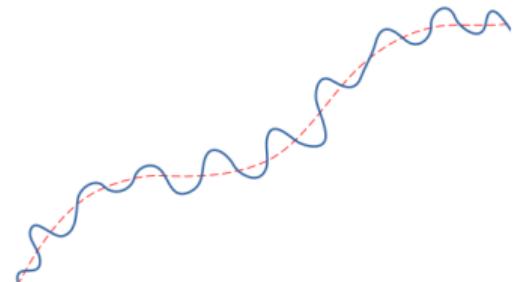
# What is Stochastic Advection by Lie Transport (SALT)?

- ▶ Assumption (in the deterministic approach): fluid particles follow

$$\frac{d\mathbf{X}(a, t)}{dt} = \mathbf{u}(\mathbf{X}(a, t), t), \quad \mathbf{X}(a, 0) = a \in \mathbb{R}^2 \text{ or } \mathbb{R}^3,$$

$\mathbf{X}(a, t)$  : particle trajectory (with Lagrangian label  $a$ )

$\mathbf{u}$  : velocity field



- ▶ In the SALT approach, Lagrangian particle paths follow a Stratonovich stochastic process

$$d\mathbf{X}(a, t) = \bar{\mathbf{u}}(\mathbf{X}(a, t))dt + \sum_i \xi_i(\mathbf{X}(a, t)) \circ dW_t^i,$$

$\bar{\mathbf{u}}$ : mean flow velocity       $\sum_i \xi_i(\mathbf{X}(a, t)) \circ dW_t^i$ : stochastic perturbation about the mean flow  
 $W_t^i$ : Ind. Wiener processes       $\xi_i$ : spatial correlation of the small-scale velocity fluctuations

- ▶ Fluid dynamics equations are derived based on this assumption

# Estimation of $\xi_i$ from high resolution data

$$(\mathbf{u} - \bar{\mathbf{u}})\Delta t \approx \sum_{i=1}^N \xi_i \Delta W_m^i$$

- $W^i$ : Brownian motion  $\rightarrow \Delta W_m^i$ : normal distribution with  $\mu = 0$  and  $\sigma^2 = \Delta t$

$$\begin{bmatrix} \Delta X_{11} & \Delta X_{12} & \cdots & \Delta X_{1n} \\ \Delta X_{21} & \Delta X_{22} & \cdots & \Delta X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \Delta X_{m1} & \Delta X_{m2} & \cdots & \Delta X_{mn} \end{bmatrix} \xrightarrow{\text{SVD}} \begin{bmatrix} | & | & & | \\ a_1 & a_2 & \dots & a_r \\ | & | & & | \\ \Delta X_{m1} & \Delta X_{m2} & \cdots & \Delta X_{mn} \end{bmatrix} \begin{bmatrix} \dots & v_1^T & \dots \\ \dots & v_2^T & \dots \\ \vdots & \vdots & \vdots \\ \dots & v_r^T & \dots \end{bmatrix}$$

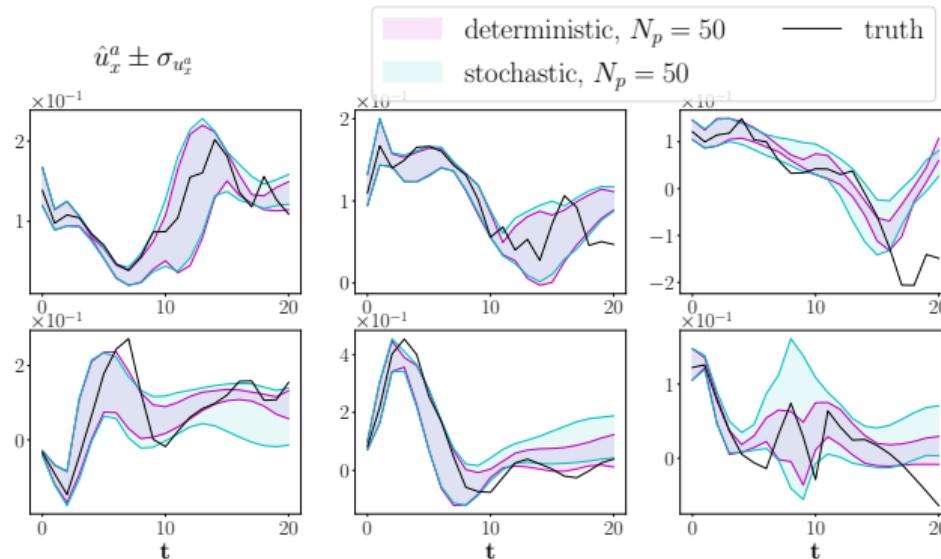
- Elements of time-series  $a_i$  are generally modeled using a normal distribution ( $\mu = 0$ ,  $\sigma^2 = \Delta t$ )
- $v_i^T$  represent the spatial correlation of the unresolved small-scale dynamics and hence are the  $\xi_i$
- Only a few  $\xi$ 's are needed to simulate most of the variance in  $(\mathbf{u} - \bar{\mathbf{u}})\Delta t$  data

# Stochastic versus deterministic ensemble

- ▶ How to quantify uncertainty due to unresolved scales with deterministic models?
- ▶ Perturb the atmosphere velocity field at  $t = 25$  to create an ensemble of initial conditions

$$\mathbf{u}_{\text{pert}} = \mathbf{u}_0 + 0.2 \times r \times \mathbf{u}_0, \quad \mathbf{u}_0 : \text{Initial velocity field}, \quad r \in \mathcal{N}(0, 1)$$

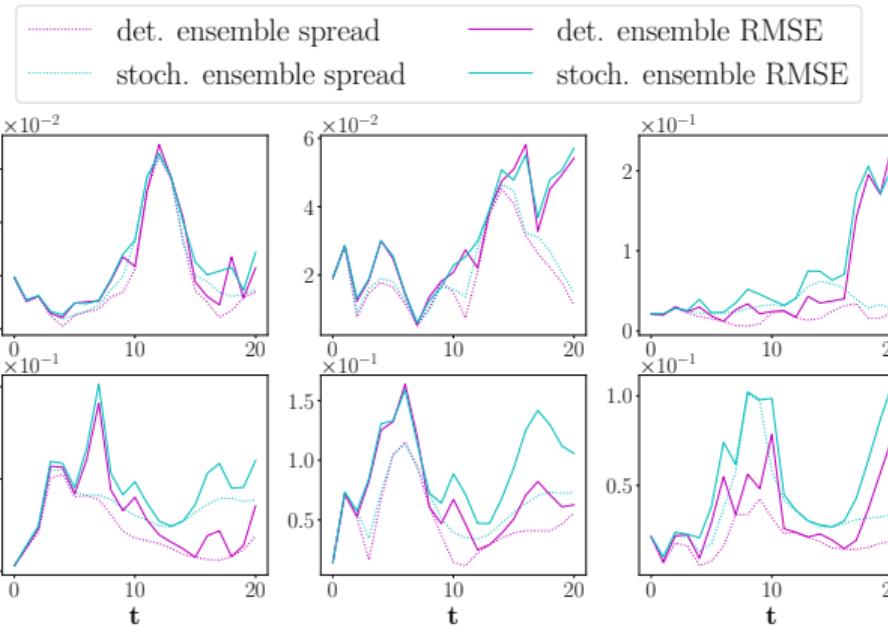
- ▶ Comparison between 50 independent realizations of the deterministic and stochastic model



- ▶ Stochastic model exhibits bigger spreads → true solution stays inside the spread for longer

# Uncertainty Quantification (UQ) skills

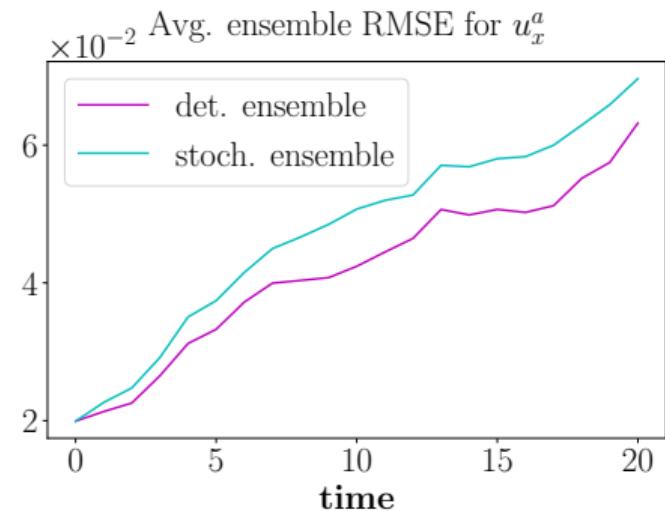
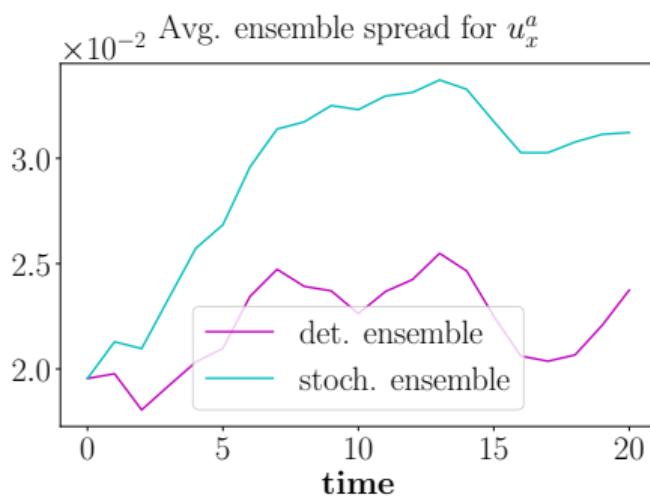
- ▶ Comparison between the ensemble spread and RMSE for atmospheric velocity ( $x$  component)



- ▶ Both models have good UQ skills (for at least 10 to 15 time units)
- ▶ Stochastic model indeed exhibits larger spreads
- ▶ RMSE of stochastic > RMSE of deterministic

# UQ skill comparison

- ▶ Average spread size and average RMSE comparison:



- ▶ Stochastic model exhibits bigger spread size but also higher RMSE!
- ▶ **Which model has better prediction skills?**

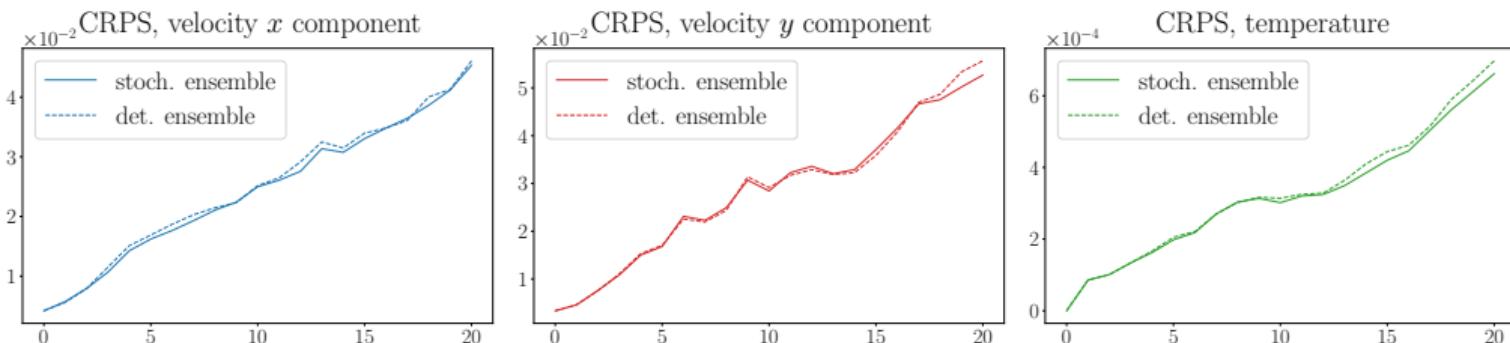
# Continuous Ranked Probability Score

- ▶ Continuous Ranked Probability Score (CRPS) evaluates the accuracy of the entire predicted probability distribution against the observed value

$$\text{crps}(F, y) = \int_{-\infty}^{\infty} (F(x) - \mathbb{1}(x \geq y))^2 dx,$$

$F(x)$  : Forecast CDF,  $y$  : Observation,  $\mathbb{1}$  : Heavyside step function

- ▶ CRPS plots comparing the performance of the stochastic and the deterministic ensemble:



- ▶ **Stochastic parameterization produces better forecasts than the deterministic approach**
  - ▶ Physics-based model
  - ▶ Produces bigger spreads, captures the truth for longer → more suitable for data assimilation