Applying the Multidimensional Langevin Process in Stochastic Physics Parameterization for the GEFS

Jian-Wen Bao[1] and Sara Michelson[1,2]

In collaboration with
Philip Pegion, Jeffrey Whitaker, Lisa Bengtsson, Cecile Penland

[1]NOAA/Physical Sciences Laboratory
[2]CIRES, University of Colorado Boulder

The 9th NOAA Ensemble Users Workshop
22–24 August 2023

Outline

1. The use of the multidimensional Langevin process (MLP) to represent the aleatory model uncertainty in subgrid physics parameterization

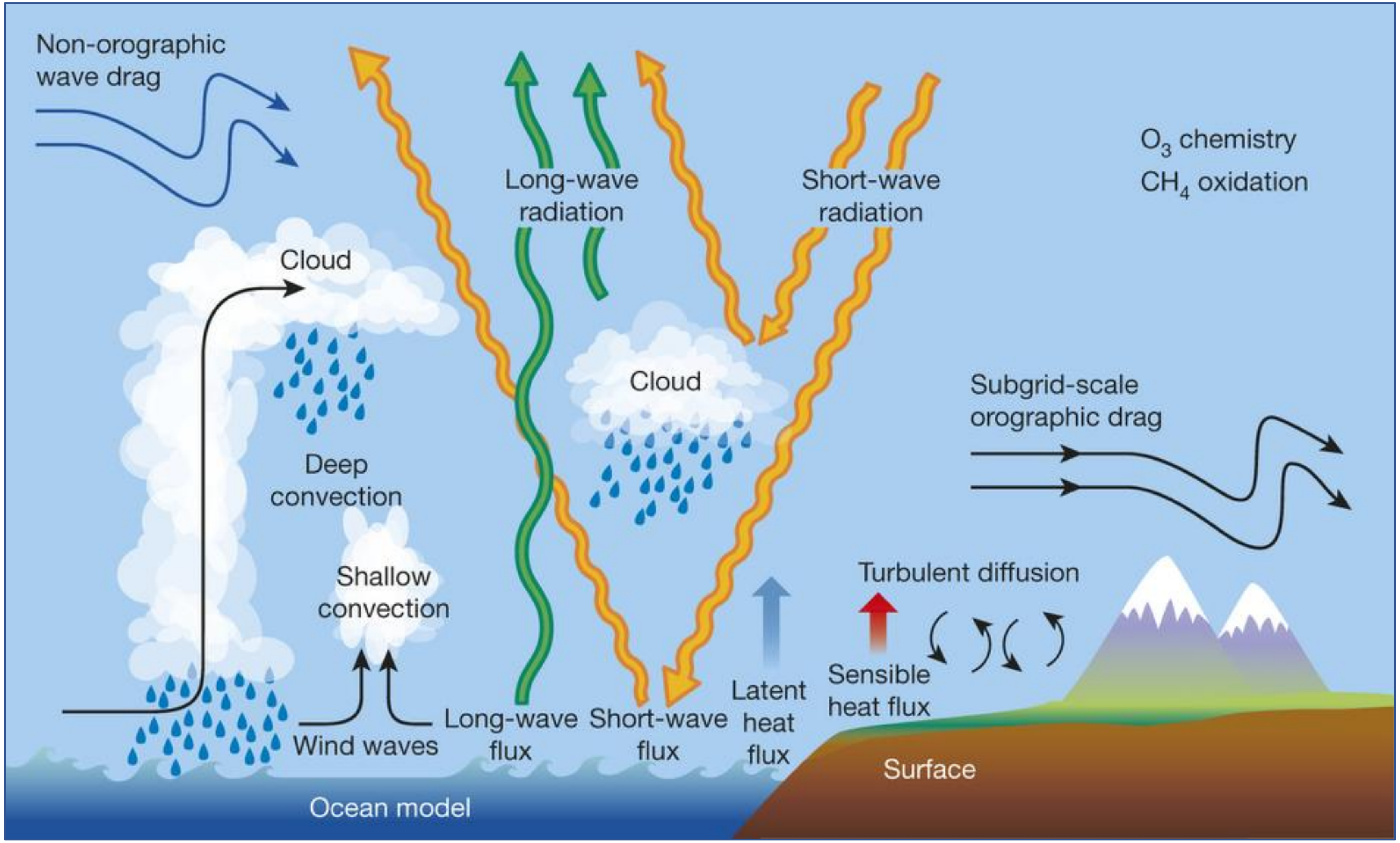
2. A GEFS example to demonstrate the potential of MLP in improving the quality of ensemble prediction

3. Interpretation of the positive results of the GEFS example in terms of stochastic calculus

4. Summary and future work

Major uncertainty in NOAA's UFS

Subgrid physical processes represented via parameterizations describing their contributions to the resolved scales in terms of mass, momentum and heat transfers



Peter Bauer et al. (2015)

The multidimensional Langevin process: A general representation of subgrid uncertainty

- The model system can generally be expressed in the phase space as the following:

$$\dot{\boldsymbol{x}} = \boldsymbol{M}(\boldsymbol{x}(t)),\ \boldsymbol{x}(0) = \boldsymbol{x}_0,\ \boldsymbol{x} = \overline{\boldsymbol{x}} + \widetilde{\boldsymbol{x}},$$

where $\overline{\boldsymbol{x}}$ and $\widetilde{\boldsymbol{x}}$ denote resolved and unresolved model state, respectively.

- Rewrite model as the so-called Liouville equation:

$$\partial \boldsymbol{z}/\partial t = \boldsymbol{L}\boldsymbol{z}(\boldsymbol{x}, t),\ \boldsymbol{z}(\boldsymbol{x}, 0) = \boldsymbol{a}(\boldsymbol{x}),$$

where the Liouville operator, $\boldsymbol{L}$, is defined as

$$\boldsymbol{L} = \boldsymbol{M} \cdot \nabla$$

(see, e.g., Chorin *et al*., PANS, 2000)

The multidimensional Langevin process (cont'd)

- The following generalized Langevin equation can be obtained by using the Mori-Zwanzig projection operators to map the Liouville equation on to the resolved and sub-grid variables

$$\dot{\overline{\boldsymbol{x}}} = \boldsymbol{e}^{t\boldsymbol{L}}\boldsymbol{P}\boldsymbol{L}\overline{\boldsymbol{x}}_0 + \int_0^t \boldsymbol{e}^{(t-s)\boldsymbol{L}}\boldsymbol{P}\boldsymbol{L}\boldsymbol{e}^{s\boldsymbol{Q}\boldsymbol{L}}\boldsymbol{Q}\boldsymbol{L}\overline{\boldsymbol{x}}_0 ds + \boldsymbol{e}^{t\boldsymbol{Q}\boldsymbol{L}}\boldsymbol{Q}\boldsymbol{L}\,\overline{\boldsymbol{x}}_0$$

the resolved dynamics

the "memory" term because it is an integration of quantities that are dependent on the model state prior to the current time

the "noise" term, representing the unresolved dynamics

where $\boldsymbol{P}$ is the projection to map $\boldsymbol{z}(\boldsymbol{x}, t)$ onto the resolved variables and $\boldsymbol{Q} = \boldsymbol{1} - \boldsymbol{P}$ is the projection to map $\boldsymbol{z}(\boldsymbol{x}, t)$ onto the subgrid variables

(Chorin *et al*., PANS, 2000)

The multidimensional Langevin process (cont'd)

In the physics literature, stochastic processes described by the generalized Langevin equation are called multi-dimensional Langevin Processes (MLP). Two approaches have been pursued to reduce the stochastic simulation of model uncertainty from the generalized Langevin equation to either (1) autoregressive models, AR(q) or (2) autoregressive moving average models, ARMA(q, p). Thus, the minimal form of the MLP for model uncertainty simulation is the following AR(1) process

$$\delta \boldsymbol{x}(t + \Delta t) = \phi \delta \boldsymbol{x}(t) + \rho \eta(t) \Delta t [d \delta \boldsymbol{x}(t)/dt]_{physics} ,$$

which is a general form of the currently widely-used stochastic random pattern generators,

$$\hat{e}(t + \Delta t) = \phi \, \hat{e}(t) + \rho \eta(t) ,$$

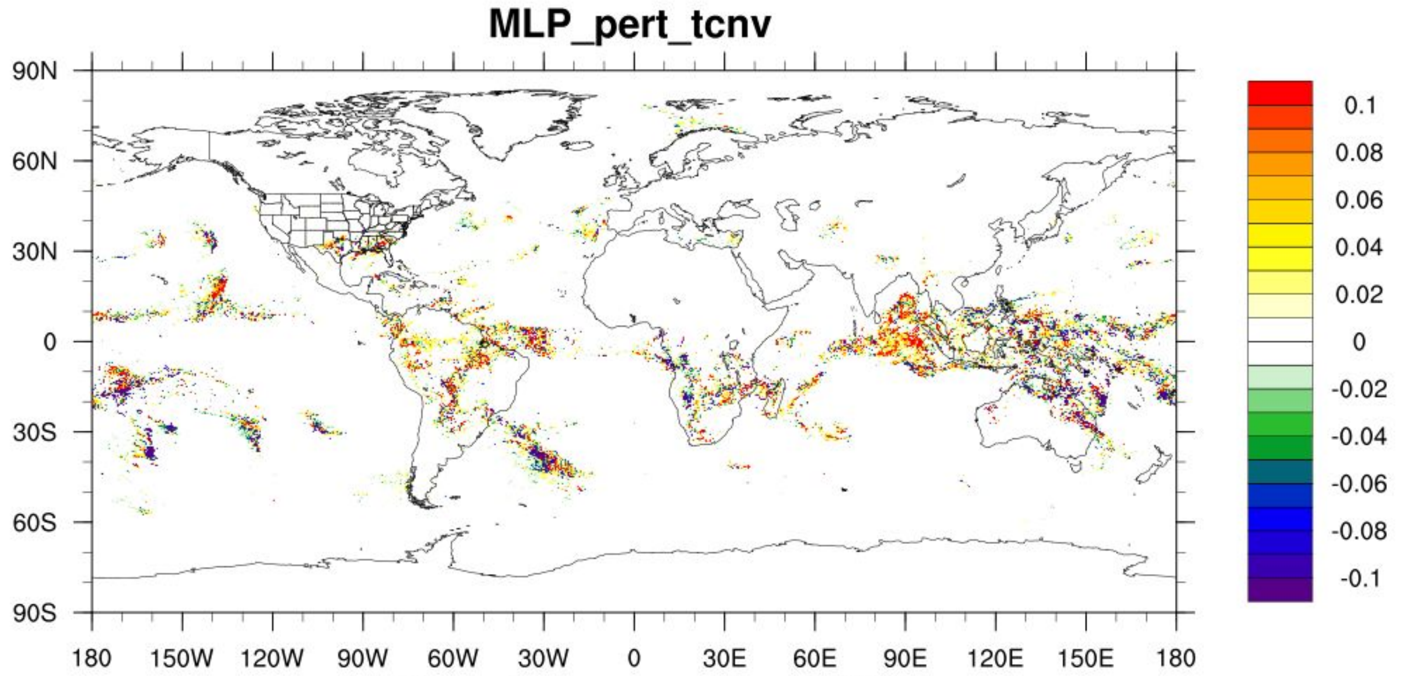
such as that used in the SPPT and SKEB schemes.

The GEFS experiment setup

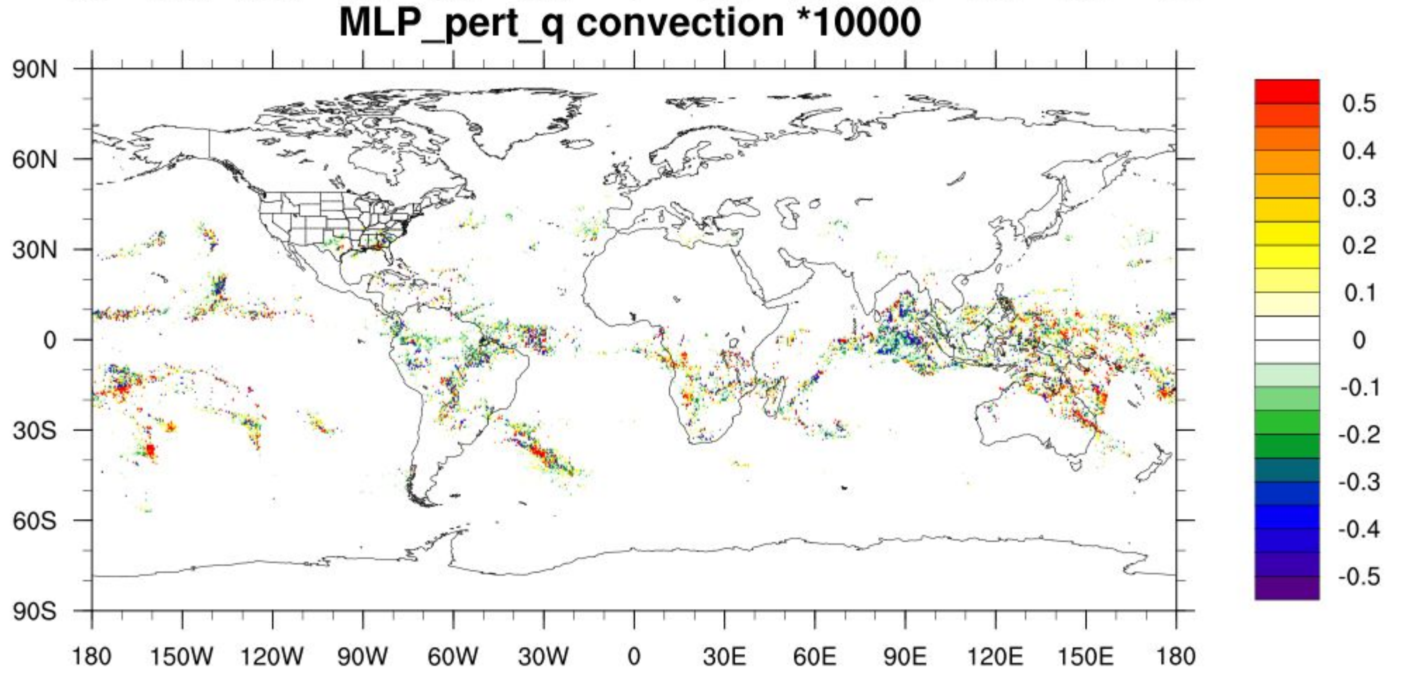
- The GEFS setup based on the GFSv16: the same physics configuration and 120 hours forecasts at the C384 resolution with 126 layers
- The same horizontal and temporal correlation scales of L_r = 500 km and τ_r = 6 hours used for individual physics schemes in both the MLP and the SPPT experiments
- 36 cases initialized at 0000 UTC on different days (Jan 1, Jan 15, Apr 1, Apr 15, Jul 1, Jul 15, Oct 1, Oct 15, and Nov 1) for the years 2014-2017
- 10 members in each ensemble forecast initialized with the perturbed initial conditions used in the operational GEFS
- Output of the ensemble runs interpolated to 0.5-degrees longitude-latitude resolution in order to compare the results to the GDAS analysis of the same resolution

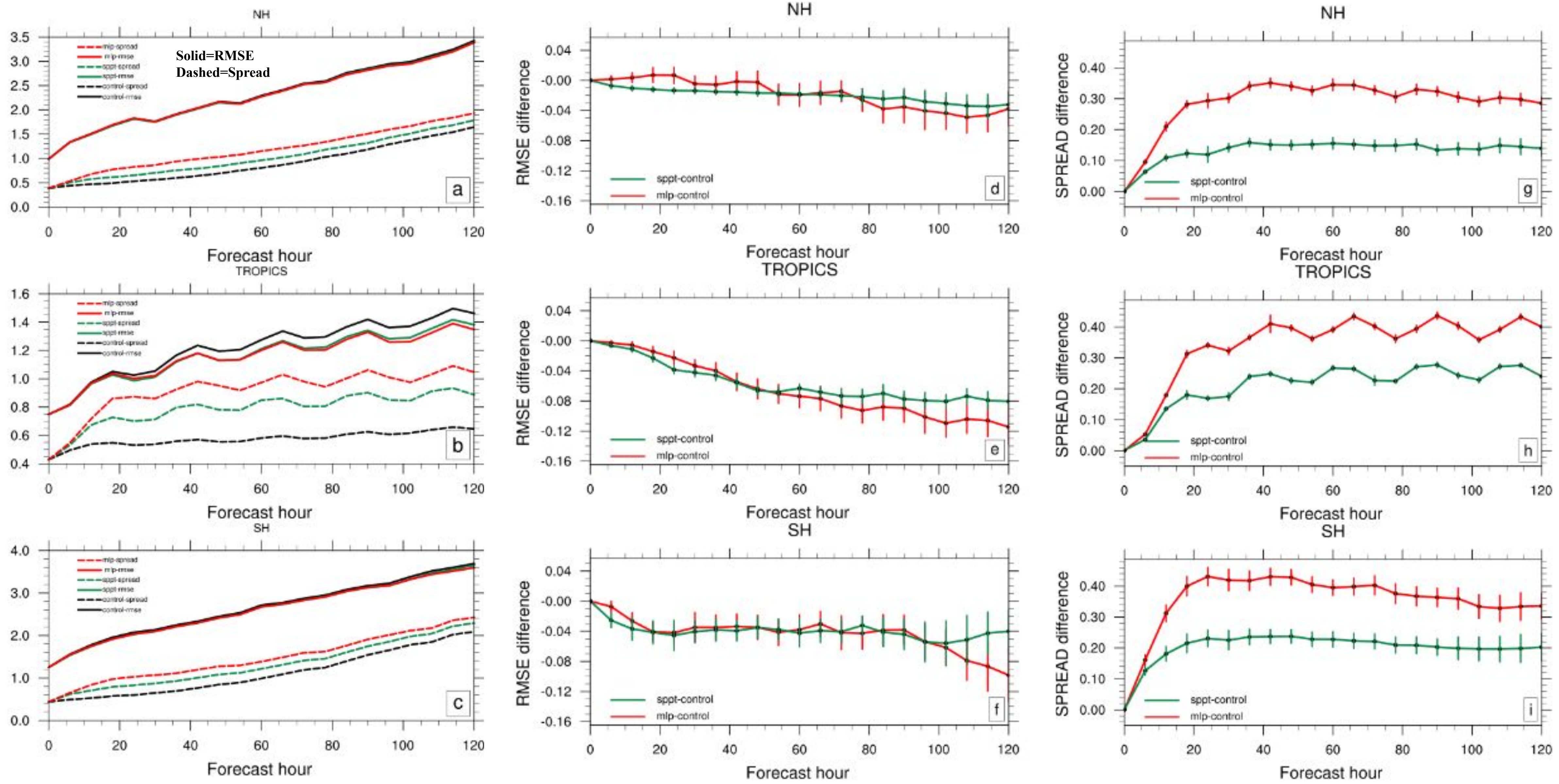
Perturbations of temperature and moisture from the convection scheme at 500 mb as an example

δT_{conv} →



δq_{conv} →





Ensemble spread (dashed lines) and RMSE (solid lines) for the 850 mb temperature (K) averaged over the (a) Northern Hemisphere (20°N-90°N), (b) Tropics (20°S-20°N) and (c) Southern Hemisphere (90°S-20°S). The black lines are for the control experiment, the green lines are for the SPPT experiment and the red lines are for the MLP experiment. Panels d-f show differences in the RSME and panels g-i show differences in the ensemble spread between the control and experiment ensemble runs using the MLP (red lines) and SPPT (green lines) schemes. The vertical bars in panels d-i indicate the 95% confidence range.

Interpretation of the positive results in terms of stochastic calculus

(Sardeshmukh *et al.*, 2023, Journal of Climate)

Take the following one dimensional system for simplicity:

$$dx/dt = a(x) + B(x)\eta \ ,$$

where η is Gaussian red noise of $N(0, 1)$ with a correlation time scale τ. The change of x over finite but still short intervals Δt is

$$\Delta x = a(x)\Delta t + \overline{\beta(x)}\sqrt{\Delta t}\phi \ ,$$

where $\beta(x) = B(x)\sqrt{\Delta\tau}$, $\phi = N(0, 1)$ is a standard Gaussian noise, $\overline{\beta(x)} \approx [\beta(x) + \beta(x + \Delta x)]/2$ (consistent with the "Stratonovich Interpretation") and $\beta(x + \Delta x) \approx \beta(x) + (d\beta/dx)\Delta x$. With these definitions, rearranging the above equation gives

$$\Delta x = [a(x) + \phi^2 d\beta(x)/dx]\Delta t + \beta(x)\sqrt{\Delta t}\phi \ .$$

The second term in [ ] is the noise-induced drift which directly affects the deterministic dynamics and hence the ensemble mean, whereas the $\sqrt{\Delta t}$ term represents a Wiener process which directly affects the ensemble spread.

Summary and future work

- A new stochastic physics parameterization scheme has been developed based on the generalized multidimensional Langevin equation, which governs the exact partition of resolved and unresolved dynamic processes.

- The previously implemented stochastic uncertainty quantification schemes in NOAA's UFS are particular cases of this new scheme.

- The positive impact of the MLP can be interpreted as due to the effect arising from the modification of the model's dynamics by a stochastic noise-induced drift.

- In the future, we will explore optimizing the performance of the new scheme using the linear inverse modeling technique.