

Ying Li\* and Samuel N. Stechmann

# Spatial and Temporal Averaging Windows and Their Impact on Forecasting: Exactly Solvable Examples

<https://doi.org/10.1515/mcwf-2018-0002>

Received June 29, 2018; accepted October 26, 2018

**Abstract:** In making weather and climate predictions, the goal is often not to predict the instantaneous, local value of temperature, wind speed, or rainfall; instead, the goal is often to predict these quantities after averaging in time and/or space—for example, over one day or one week. What is the impact of spatial and/or temporal averaging on forecasting skill? Here this question is investigated using simple stochastic models that can be solved exactly analytically. While the models are idealized, their exact solutions allow clear results that are not affected by errors from numerical simulations or from random sampling. As a model of time series of oscillatory weather fluctuations, the complex Ornstein-Uhlenbeck process is used. To furthermore investigate spatial averaging, the stochastic heat equation is used as an idealized spatiotemporal model for moisture and rainfall. Space averaging and time averaging are shown to have distinctly different impacts on prediction skill. Spatial averaging leads to improved forecast skill, in line with some forms of basic intuition. Time averaging, on the other hand, is more subtle: it may either increase or decrease forecast skill. The subtle effects of time averaging are seen to arise from the relative definitions of the time averaging window and the lead time. These results should help in understanding and comparing forecasts with different temporal and spatial averaging windows.

**Keywords:** prediction skill, lead time, temporal averaging, stochastic partial differential equations, Ornstein-Uhlenbeck process, stochastic heat equation

## 1 Introduction

Weather predictions are commonly made for *averaged* quantities. For example, the amount of rainfall might be predicted as an average in time over an hour, day, week, month, season, etc. [30, 36, 40]. For longer time averages, the forecasts are often associated with intraseasonal variability such as the Madden–Julian Oscillation (MJO) [1, 6, 16, 18–20, 22, 39] or El Niño–Southern Oscillation (ENSO) [3, 17, 25, 27, 33]. For shorter time averages, the forecasts are often associated with synoptic variability or individual convective storms. The question of spatial averaging becomes particularly relevant as further details of convective systems are resolved by newer numerical weather prediction models, and many new challenges arise in formulating forecasts of precipitation and in assessing errors and uncertainties [2, 4, 11, 12, 24, 28, 29].

One of the main questions here is: How does prediction skill vary as the space- and/or time-averaging window is varied? Here this question is investigated using exactly solvable models, which allows precise assessment of forecasts without being tarnished by errors from numerical approximations or pseudo-random Monte Carlo sampling. The first model used here is the complex Ornstein–Uhlenbeck (cOU) process, a stochastic

---

\*Corresponding Author: Ying Li: Department of Mathematics, University of Wisconsin–Madison, Madison, WI USA, E-mail: li526@wisc.edu

Samuel N. Stechmann: Department of Mathematics & Department of Atmospheric and Oceanic Sciences, University of Wisconsin–Madison, Madison, WI USA, E-mail: stechmann@wisc.edu

model for time series of weather fluctuations [21, e.g.]. The second model used here is the stochastic heat equation, which has been used as an idealized spatiotemporal model of tropical rainfall [13, 31, 32].

A priori, one would perhaps expect from basic intuition that spatially and/or temporally averaged variables should have longer predictability. Intuition would suggest that averaging should filter out the high-frequency, nearly unpredictable components of the system, and only the low-frequency, more-predictable components of the system would remain. For example, a time average over a week or a month should “average out” the small-scale fluctuations in rainfall associated with individual convective storms, and the remaining, averaged signal should be associated with low-frequency variability such as the MJO and/or ENSO.

In the results shown here, this basic intuition sometimes holds and sometimes does not. Different results will be seen for spatial versus temporal averaging. Spatial averaging leads to improved forecast skill, in line with the basic intuition described above. Time averaging, on the other hand, is more subtle: it may either increase or decrease forecast skill, depending on the relative definition of the lead time.

Our original motivation stemmed from analyzing some preliminary forecasts of precipitation associated with convectively coupled equatorial waves (CCEWs) and the MJO, using the model of [32]. In forecasting the rainfall associated with such multiscale systems, the skill could potentially be measured in a variety of ways, depending on the length scales and time scales and wave types of interest. Some very recent studies have now analyzed the forecast skill in this setting using operational numerical weather prediction models [10, 14, 37, 38]. In the future it would be interesting to conduct more detailed forecasts with the model of [32], which could offer theoretical perspectives owing to the model’s simplicity relative to numerical weather prediction systems.

The paper is organized as follows. The model equations and prediction setup are defined in section 2. Forecasts of time series of the cOU process are presented in section 3. Subtleties of time averaging are discussed in further detail in section 4, including comparisons of different definitions of lead time relative to the averaging window. Forecasts of the spatiotemporal rainfall model (the stochastic heat equation) are presented in section 5, allowing investigation of both time averaging and space averaging, and allowing investigation of time averaging for a more complicated system that has numerous degrees of freedom. Conclusions are described in section 6.

## 2 Models and Methods

In this section, the two exactly solvable models are introduced: the complex Ornstein-Uhlenbeck process as a model for a single oscillatory mode, and the stochastic heat equation as a model for spatiotemporal dynamics. In addition, several aspects of the forecasting setup are also described, including measures of forecast skill and definitions of time and space averaging.

### 2.1 Mathematical Models

#### 2.1.1 Complex Ornstein-Uhlenbeck(cOU) Process

A complex Ornstein-Uhlenbeck(cOU) process is applied for the first insight of the forecasting behaviors in predicting signals at single time points or signals averaged over a temporal window. Conducting forecasting analyses on a cOU process is a meaningful attempt for detecting the forecast skill for predicting signals with a single oscillation mode. We consider a complex linear stochastic differential equation for the Ornstein-Uhlenbeck process

$$du(t) = (-\gamma + i\omega)u(t) + \sigma dW(t), \quad (1)$$

where  $\gamma, \sigma > 0$  and  $\omega$  are real numbers and

$$dW(t) \equiv \frac{dW_1(t) + idW_2(t)}{\sqrt{2}} \quad (2)$$

is a complex Gaussian white noise where each component satisfies

$$dW_j(t) \equiv \dot{W}_j(t)dt, \quad j = 1, 2 \quad (3)$$

that is, white noise is a "derivative" of the Wiener process  $W_j(t)$  and it satisfies the following properties

$$\mathbb{E}[\dot{W}_j(t)] = 0 \quad (4)$$

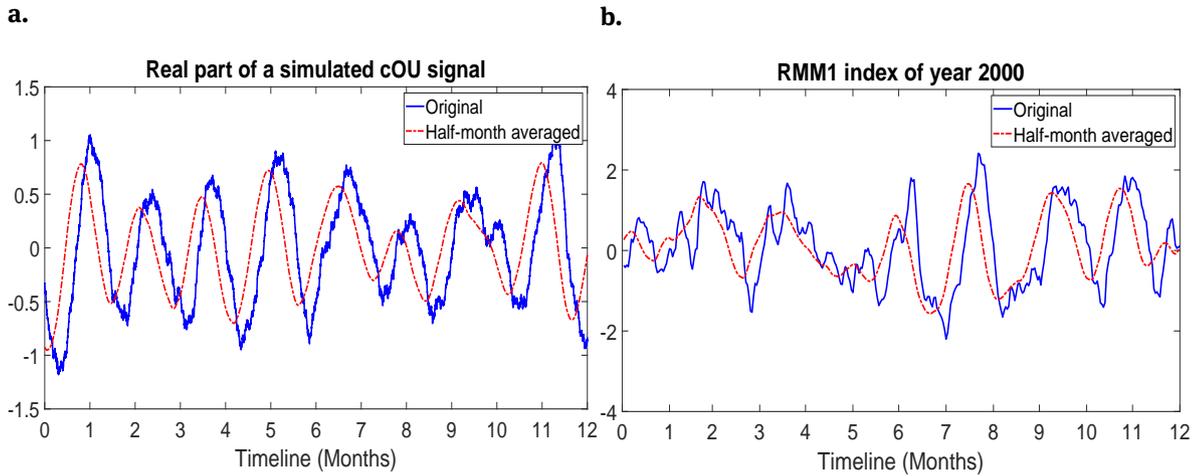
$$\mathbb{E}[\dot{W}_j(t)\dot{W}_j(s)] = \delta(t-s) \quad (5)$$

$$\mathbb{E}[\dot{W}_i(t)\dot{W}_j(s)] = 0 \quad \text{for } i \neq j. \quad (6)$$

The exact solution of (1) is

$$u(t) = e^{(-\gamma+i\omega)t}u(0) + \sigma \int_0^t e^{(-\gamma+i\omega)(t-s)} dW(s). \quad (7)$$

As  $t \rightarrow \infty$ ,  $u(t)$  will converge to a stationary Gaussian distribution with mean 0 and variance  $\frac{\sigma^2}{2\gamma}$  (see details in [21]).



**Figure 1:** Panel a: Time series of the real part of a simulated signal from a cOU process over 12 months; Panel b: Time series of RMM1 index in the year 2000. In both panels, the blue curve is the original signal while the dashed red curve is the half-month averaged signal (namely  $\hat{u}(t, T_w = 0.5 \text{ month})$  from  $t = 0$  to  $t = 12$  months in (12)).

In the simulation and plotting in this paper, we choose the parameters  $\gamma = 0.1 \text{ month}^{-1}$ ,  $\omega = \frac{4}{3}\pi \text{ month}^{-1}$ ,  $\sigma = 0.5$  to represent a wave with oscillation period  $2\pi/\omega = 1.5$  months like the MJO and a decorrelation time  $1/\gamma = 10$  months that is longer than the oscillation period, resulting in the occurrence of sequences of irregular wave oscillations. All the analytic formulas have been checked with the numerical simulations. A simulated time series over 12 months is shown in panel a of Fig. 1 along with the time series of Real-time Multivariate MJO (RMM) index in the year 2000. The similarities between the two time series show that the cOU signals are good representations for the dynamics to some extent. When a time averaging is applied, small oscillations will be removed from the original signal through the temporal averaging process so that the signal becomes smoother with an expectation for a lower variance.

For investigating more about influences of temporal and spatial averaging in forecasting, a one dimensional model based on the stochastic heat equation is analyzed for some insight. The simple model applied to the water vapor dynamics has been shown that it has behaviors very similar to the observational statistics in the article [13]. The model is a stochastic PDE

$$\frac{\partial q}{\partial t} = b_0 \nabla^2 q - \frac{1}{\tau}(q - q^*) + F + D_* \dot{W}, \quad (8)$$

where  $q(t, x)$  is the signal at time  $t$  and location  $x$  depending on a spatial interaction constant  $b_0$ , the relaxation time  $\tau$ , the relaxation target  $q^*$ , an external force  $F$  and the stochastic forcing variance  $D_*^2$ . Now,  $\dot{W}$  is a space-time white noise. It can be characterized as a centered Gaussian process such that  $\mathbb{E}[\dot{W}(s, x)\dot{W}(t, y)] = \delta(s - t)\delta(x - y)$ . We call (8) as SHE or SPDE for abbreviations in this paper. In one dimensional case, (8) can be solved analytically as

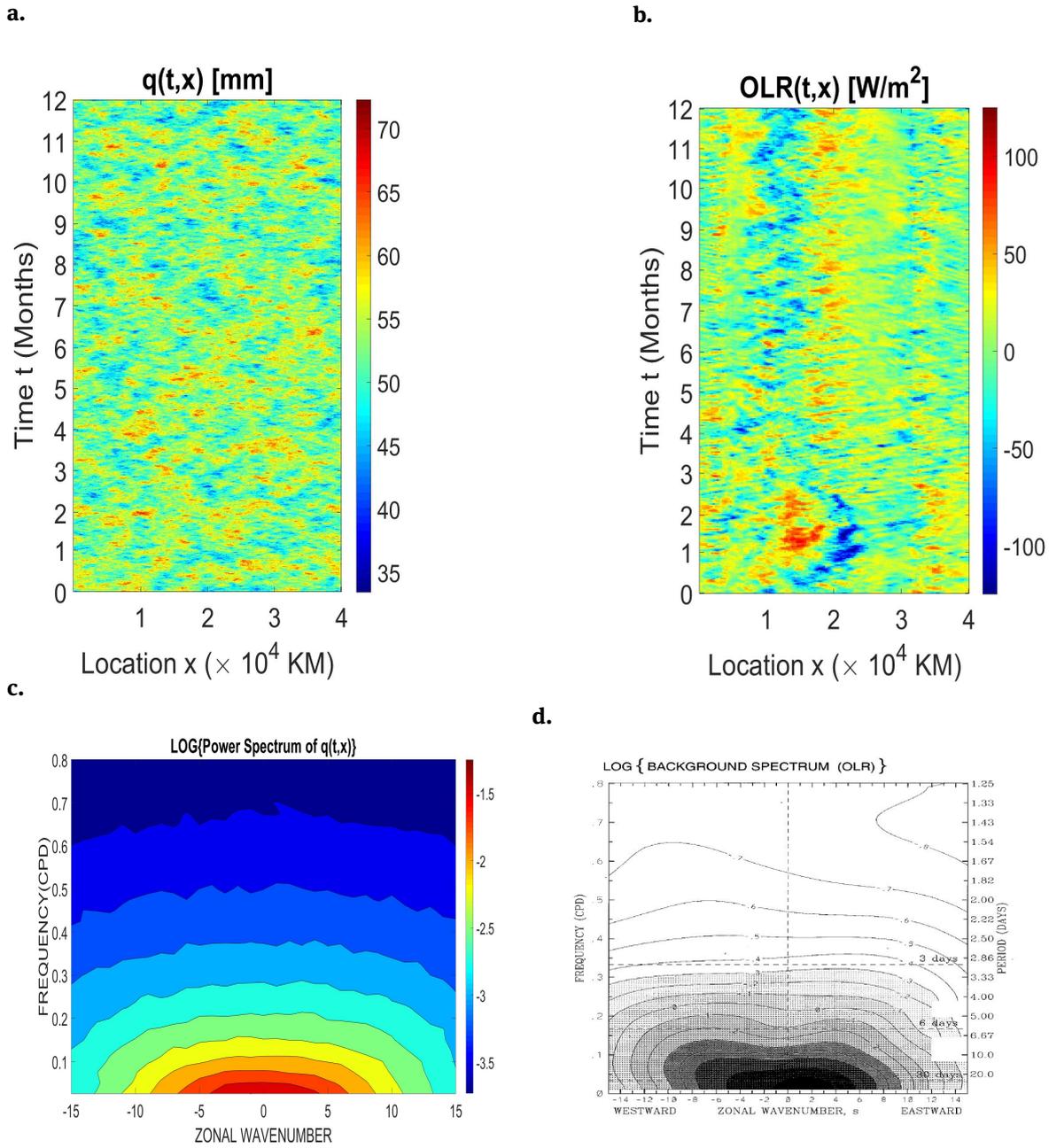
$$q(t, x) = q^* + \tau F + \frac{e^{-t/\tau}}{(4\pi b_0 t)^{1/2}} \int_{\mathbb{R}} e^{-\frac{|x-y|^2}{4b_0 t}} (q(0, y) - q^* - \tau F) dy \\ + e^{-t/\tau} \int_0^t \frac{1}{(4\pi b_0 |t-s|)^{1/2}} \int_{\mathbb{R}} e^{-\frac{|x-y|^2}{4b_0(t-s)}} e^{s/\tau} \dot{W}(s, y) dy ds. \quad (9)$$

Besides the analytical solution, there are many useful and important statistics that can be computed directly using the analytical solution, such as the temporal covariance  $cov[q(t, x)q(s, x)]$  and the spatial covariance  $cov[q(t, x)q(t, y)]$  which will help a lot in accessing the forecast skill (See in appendix section C). The signal arising from this type of stochastic partial differential equation like (8) will enter into a stationary Gaussian distribution as time goes to infinity. The stationary distribution has mean  $q^* + \tau F$  and variance  $\frac{D_*}{4} \sqrt{\frac{\tau}{b_0}}$ . In addition, seeing from spatial Fourier space,  $q(t, x)$  can be decomposed into the integral of a bunch of components satisfying the equation of a cOU process (1) independently. All the components have a corresponding decorrelation time  $1/\gamma$  in the cOU process that is not greater than  $\tau$  and  $\omega = 0$ . From this point of view,  $q(t, x)$  can be seen as the combination of numerous different waves with decorrelation time not greater than  $\tau$  and oscillation frequency  $\omega = 0$ . A detailed description and derivation for the discrete version of this Fourier transform can be seen in [13].

In our simulation and figures, we aim to simulate to recover the rainfall statistics according to the probability distribution in [13] and the power spectrum density plot in [35].  $\tau = 96$  hours,  $q^* = 65$  mm,  $F = -0.125$  mm · hour<sup>-1</sup>,  $b_0 = 10^4$  km<sup>2</sup> · hour<sup>-1</sup>,  $D_* = 25$  mm · km<sup>1/2</sup> · hour<sup>-1/2</sup> are chosen for this purpose. All the analytic formulas have been checked with the numerical simulation. Both the temporal averaging and the spatial averaging have an influence on eliminating the small oscillations and smoothing. A global dynamic view of the simulated  $q(t, x)$  over one year is shown in panel **a** of Fig. 2 along with the OLR observational data in the year of 2010 after averaging over the tropical belt, which provides a OLR signal  $OLR(t, x)$  for time  $t$  and longitude  $x$ . The power spectrum density plot of the simulated data in panel **c** of Fig. 2 performs a good match with the power density based on the observational OLR data in the paper [35] (shown in panel **d** of Fig. 2). The similarities between the observational data and the simulated data provide evidence for that stochastic heat equation is a reasonable and good model for representing the atmospheric dynamic system.

## 2.2 Measures of Forecasting Skill: Mean Square Error (MSE) and Pearson Correlation Coefficients ( $\rho$ )

Two types of frequently-used criterions, mean square error(MSE) and the Pearson correlation coefficients, are used in this paper for evaluating the forecasting skill. The mean square error measures the average of the squares of the errors that is the difference between the predictions and the true signal. When a temporal or spatial averaging is applied, the variance of the averaged signal will also be scaled from the variance of the original signal. Hence, instead of the MSE, the ratio of the MSE over the variance of the averaged signal seems to be a more reasonable and meaningful measure for the assessment. The Pearson correlation coefficient is a measure of the strength and direction of the linear relationship between two variables. Mathematically, if  $X(t)$  is the true signal from the cOU process (1) or the stochastic heat equation (8) and  $Y(t)$  is our prediction, both  $X(t)$  and  $Y(t)$  are stationary and ergodic processes and  $(X(t), Y(t))$  is also a jointly stationary and ergodic process. Stationarity ensures that the first moment  $\mathbb{E}[X(t)]$ ,  $\mathbb{E}[Y(t)]$ , the second moment  $\mathbb{E}[X(t)^2]$ ,  $\mathbb{E}[Y(t)^2]$  and the mixed moment  $\mathbb{E}[X(t)Y(t)]$  are all constants that are not depending on the time  $t$ . Then the MSE of



**Figure 2:** Panel **a.** A simulated global image of  $q(t, x)$ ; Panel **b.** A global image of of OLR observational data in the year 2010 after averaging over the tropical belt; Panel **c.** The power spectrum density of the simulated data; Panel **d.** Zonal wavenumber-frequency spectrum of the base-10 logarithm of the background power from the OLR observational data cited from Wheeler&Kiladis’s paper [35] ©American Meteorological Society. Used with permission.

the estimator  $Y(t)$  with respect to the unknown variable  $X(t)$  is defined as

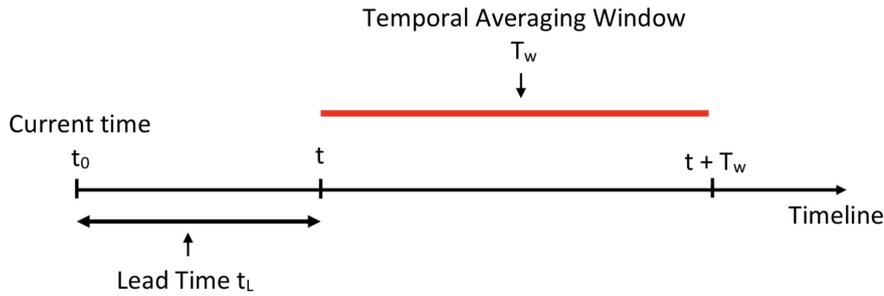
$$MSE(X(t), Y(t)) = \mathbb{E}[(Y(t) - X(t))^2], \tag{10}$$

which is a constant not depending on the time  $t$ . The formula for the Pearson correlation coefficients  $\rho$  can be expressed as

$$\rho(X(t), Y(t)) = \frac{cov(X(t), Y(t))}{\sigma_X \sigma_Y}, \tag{11}$$

where  $\text{cov}(X(t), Y(t)) = \mathbb{E}[(X(t) - \mathbb{E}[X(t)])(Y(t) - \mathbb{E}[Y(t)])]$  is the covariance between  $X(t)$ ,  $Y(t)$  and  $\sigma_X$ ,  $\sigma_Y$  are the standard deviations of  $X$ ,  $Y$  respectively.  $\rho(X(t), Y(t))$  is also a constant that doesn't change over time. In a situation of analyzing the real data, we have the true data from the true signal as  $X_1, X_2, \dots, X_N$  and our corresponding predictions  $Y_1, Y_2, \dots, Y_N$ . By ergodicity, as  $N \rightarrow \infty$ , the sample means  $\frac{1}{N} \sum_{i=1}^N X_i$ ,  $\frac{1}{N} \sum_{i=1}^N Y_i$  converge in squared mean to  $\mathbb{E}[X(t)]$ ,  $\mathbb{E}[Y(t)]$  and  $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N X_i^2 = \mathbb{E}[X(t)^2]$ ,  $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N Y_i^2 = \mathbb{E}[Y(t)^2]$ ,  $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N X_i Y_i = \mathbb{E}[X(t)Y(t)]$ . These connect the sample MSE and sample correlation coefficient of a real dataset with our idealized and theoretical definitions (10) and (11).

## 2.3 Temporal and Spatial Averaging: Definitions and Notations



**Figure 3:** Schematic of default lead time and temporal averaging window definitions in the analysis. The horizontal axis represents forecast time from the initial condition at  $t_0$ .

When we do the temporal averaging, there are different ways to place the temporal averaging window. Motivated by Fig. 1 in the paper [41], the default definitions of the lead time when there is a temporal averaging window in this paper are shown in Fig. 3. For the temporal averaged cOU signal  $\hat{u}(t, T_w)$  at time  $t$  with averaging window width  $T_w$ , it is defined as the averaged signal from time  $t$  to time  $t + T_w$ , namely

$$\hat{u}(t, T_w) = \frac{1}{T_w} \int_t^{t+T_w} u(s) ds. \quad (12)$$

Similarly, the temporal averaged stochastic heat equation signal at time  $t$  with averaging window width  $T_w$  at a single location  $x$  is defined as

$$\hat{q}(t, x, T_w) = \frac{1}{T_w} \int_t^{t+T_w} q(s, x) ds. \quad (13)$$

For the spatial averaging at location  $x$ , the signals are averaged over a spatial window  $[x - L_w/2, x + L_w/2]$  with  $L_w$  as the averaging window width and  $x$  as the center of the spatial averaging window, namely

$$\tilde{q}(t, x, L_w) = \frac{1}{L_w} \int_{x-L_w/2}^{x+L_w/2} q(t, y) dy. \quad (14)$$

The averaged signal with both temporal and spatial averaging is a combination of (13) and (14) as

$$\bar{q}(t, x, T_w, L_w) = \frac{1}{L_w T_w} \int_{x-L_w/2}^{x+L_w/2} \int_t^{t+T_w} q(s, y) ds dy. \quad (15)$$

The changes of variances in the cOU or SHE signals through the temporal and/or spatial averaging are also investigated, which is described in appendix section A in detail.

Note that both the complex Ornstein-Uhlenbeck process and the stochastic heat equation are Markovian processes. With the aids of Markovian properties of the two models, a lot of useful analytical formulas are derived and calculated when we assess the forecasting skills involving temporal and/or spatial averaging in the later sections. It would be interesting in the future to investigate whether similar results are still seen with non-markovian processes.

## 3 Forecasting the Complex Ornstein-Uhlenbeck Process

### 3.1 Forecasting at a Single Time Point

In an idealized situation, we have the perfect model of a cOU process for predicting, which means we have the exact true values of the parameters  $\gamma$ ,  $\omega$ ,  $\sigma$  in (1). However, by using the ensemble mean prediction, the value of  $\sigma$  essentially will not affect the prediction skill. The true signal is a known realization from the cOU process, we denote it as

$$u_{truth}(t; \theta) = \text{The true signal} = u(t; \theta) \quad (16)$$

with the underlying parameter  $\theta$  which labels the realization of the stochastic process. By (7), treating  $t - t_L$  as the current time, one single prediction for time  $t$  with lead time  $t_L$  will be another realization starting from  $u_{truth}(t - t_L; \theta)$  with underlying parameter  $\zeta$  as

$$\begin{aligned} u_{pred}(t, t_L; \zeta, \theta) &= \text{The prediction of the value of } u \text{ at time } t, \\ &\quad \text{given the value of } u \text{ at time } t - t_L, \\ &\quad \text{for ensemble member label } \zeta \\ &= e^{(-\gamma+i\omega)t_L} u_{truth}(t - t_L; \theta) + \sigma \int_{t-t_L}^t e^{(-\gamma+i\omega)(t-s)} dW(s; \zeta). \end{aligned} \quad (17)$$

Note that, technically speaking,  $W(s; \zeta)$  here is from another new Wiener process which could be denoted as  $\tilde{W}(s; \zeta)$ , although the tilde will be left off in order to ease notation. Also note that the prediction  $u_{pred}$  can be separated into components from the  $\theta$  realization and  $\zeta$  realization according to the Markovian nature of the process, as the future is independent from the past.

At the same time  $u_{truth}(t; \theta)$  is also a realization starting from  $u_{truth}(t - t_L; \theta)$  with underlying parameter  $\theta$  as

$$u_{truth}(t; \theta) = e^{(-\gamma+i\omega)t_L} u_{truth}(t - t_L; \theta) + \sigma \int_{t-t_L}^t e^{(-\gamma+i\omega)(t-s)} dW(s; \theta). \quad (18)$$

Then the ensemble mean of the predictions in the form of (17) serves as the ensemble forecasting for time  $t$  with ensemble size  $\infty$  which is

$$\begin{aligned} u_{pred}(t, t_L; \theta) &= \mathbb{E}_{\zeta}[u_{pred}(t, t_L; \zeta, \theta)] \\ &= e^{(-\gamma+i\omega)t_L} u_{truth}(t - t_L; \theta). \end{aligned} \quad (19)$$

First, we predict the signal at single time points with the perfect cOU model and lead time  $t_L$ . The mean square error ( $MSE_u$ ) and the Pearson correlation coefficient ( $\rho_u$ ) between the prediction and the true signal can be calculated as

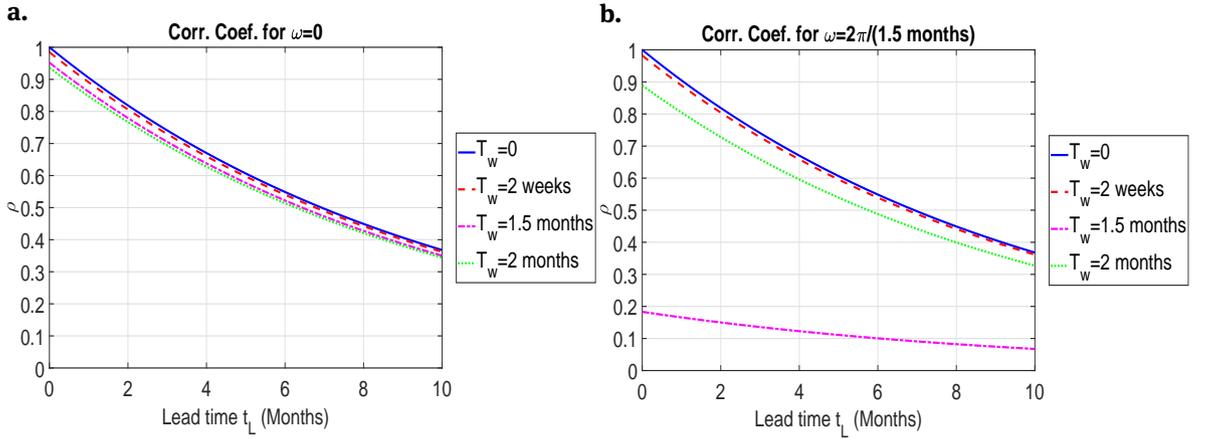
$$\begin{aligned} MSE_u(t_L) &= \mathbb{E}_{\theta}[|u_{truth}(t; \theta) - u_{pred}(t, t_L; \theta)|^2] \\ &= \mathbb{E}_{\theta}[\sigma^2 \int_{t-t_L}^t e^{(-\gamma+i\omega)(t-s)} dW(s; \theta)]^2 \end{aligned}$$

$$= \frac{\sigma^2}{2\gamma}(1 - e^{-2\gamma t_L}), \quad (20)$$

where  $\frac{\sigma^2}{2\gamma}$  is the variance of the true signal.

$$\begin{aligned} \rho_u(t_L) &= \frac{\mathbb{E}_\theta[(u_{pred}(t, t_L; \theta) - \mathbb{E}_\theta[u_{pred}(t, t_L; \theta)])(u_{truth}(t; \theta) - \mathbb{E}_\theta[u_{truth}(t; \theta)])^*]}{(\text{var}[u_{pred}(t, t_L; \theta)])^{1/2}(\text{var}[u_{truth}(t; \theta)])^{1/2}} \\ &= \frac{\mathbb{E}_\theta[u_{pred}(t, t_L; \theta)u_{truth}^*(t; \theta)]}{(\text{var}[u_{pred}(t, t_L; \theta)])^{1/2}(\text{var}[u_{truth}(t; \theta)])^{1/2}} \\ &= \frac{e^{(-\gamma+i\omega)t_L} \mathbb{E}_\theta[u_{truth}(t - t_L; \theta)u_{truth}^*(t; \theta)]}{e^{-\gamma t_L} (\text{var}[u_{truth}(t - t_L; \theta)])^{1/2} (\text{var}[u_{truth}(t; \theta)])^{1/2}} \\ &= \frac{e^{(-\gamma+i\omega)t_L} e^{(-\gamma-i\omega)t_L} \mathbb{E}_\theta[u_{truth}(t - t_L; \theta)u_{truth}^*(t - t_L; \theta)]}{e^{-\gamma t_L} \frac{\sigma^2}{2\gamma}} \\ &= e^{-\gamma t_L}. \end{aligned} \quad (21)$$

With a perfect model, the mean square error will keep increasing as the lead time increases but will never exceed the variance of the underlying true signal. As lead time grows to infinity,  $MSE_u$  is approaching the variance of the true signal. The correlation coefficient  $\rho_u(t_L)$  keeps decreasing with respect to the increasing in the lead time and will drop below 0.5 if the lead time  $t_L$  is larger than  $\ln(2)/\gamma$ , approximately 0.69 of the decorrelation time  $1/\gamma$ . The characteristics of the forecasting skills at a single time point for a cOU process are shown with the blue lines in Fig. 4. Since the conclusions from MSE ratios are consistent with those from the correlation coefficients throughout this paper, we show only MSE formulas but not plots everywhere.

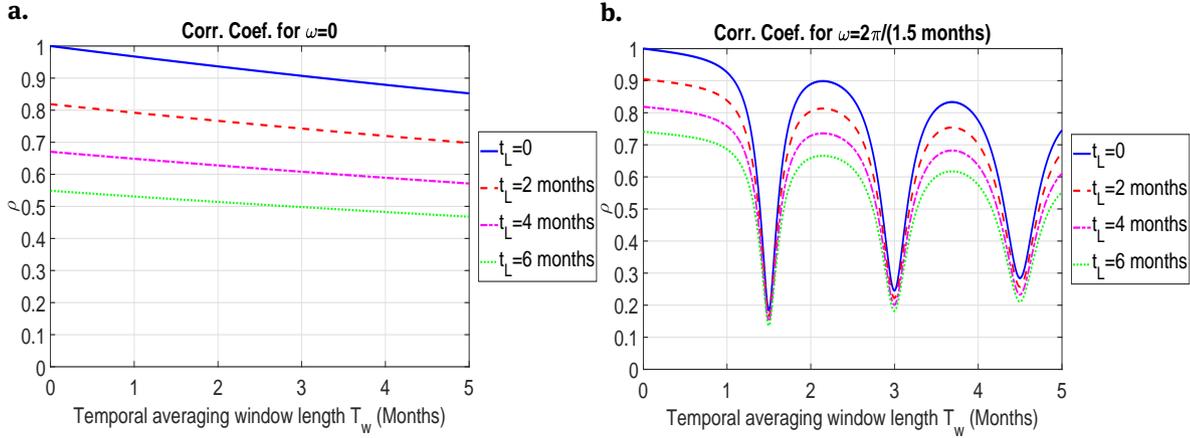


**Figure 4:** Panel **a.** forecasting skill (correlation coefficient) for different temporal averaging windows of a cOU process with  $\omega = 0$  under the default definitions clarified in the section 2.3; Panel **b.** forecasting skill (correlation coefficient) for different temporal averaging windows of a cOU process with  $\omega = 2\pi/(1.5 \text{ months})$  under the default definitions clarified in the section 2.3.

### 3.2 Forecasting with Temporal Averaging

Then we move on to assess the forecasting skill with a temporal averaging window. The forecasting skill is achieved by comparing the averaged true cOU signal at time  $t$ , namely

$$\hat{u}_{truth}(t, T_w; \theta) = \frac{1}{T_w} \int_t^{t+T_w} u_{truth}(s; \theta) ds$$



**Figure 5:** Panel a. forecasting skill for different lead times of a cOU process with  $\omega = 0$  under the default definitions clarified in the section 2.3; Panel b. forecasting skill for different lead times of a cOU process with  $\omega = 2\pi/(1.5 \text{ months})$  under the default definitions clarified in the section 2.3.

$$= \frac{1}{T_w} \int_t^{t+T_w} e^{(-\gamma+i\omega)(s-t+t_L)} u_{\text{truth}}(t-t_L; \theta) + \sigma \int_{t-t_L}^s e^{(-\gamma+i\omega)(s-s')} dW(s'; \theta) ds \quad (22)$$

and the ensemble prediction at time  $t$  with lead time  $t_L$  by the default definition clarified in the section 2.3 for this averaged true signal. Seeing  $t-t_L$  as the current time, a single prediction with lead time  $t_L$  for this averaged signal with underlying ensemble label  $\zeta$  is specified as

$$\begin{aligned} \hat{u}_{\text{pred}}(t, t_L, T_w; \zeta, \theta) &= \frac{1}{T_w} \int_t^{t+T_w} u_{\text{pred}}(s, s-(t-t_L); \zeta, \theta) ds \\ &= \frac{1}{T_w} \int_t^{t+T_w} e^{(-\gamma+i\omega)(s-t+t_L)} u_{\text{truth}}(t-t_L; \theta) + \sigma \int_{t-t_L}^s e^{(-\gamma+i\omega)(s-s')} dW(s'; \zeta) ds. \end{aligned} \quad (23)$$

Then the ensemble forecasting is the mean of the above forecasting with respect to  $\zeta$

$$\begin{aligned} \hat{u}_{\text{pred}}(t, t_L, T_w; \theta) &= \mathbb{E}_{\zeta}[\hat{u}_{\text{pred}}(t, t_L, T_w; \zeta, \theta)] \\ &= \frac{1}{T_w} \int_t^{t+T_w} e^{(-\gamma+i\omega)(s-t+t_L)} u(t-t_L; \theta) ds. \end{aligned} \quad (24)$$

Having all these ready, we are proceeding to compute the forecasting skill in this temporal averaging case.

$$\begin{aligned} \widehat{MSE}_u(t_L, T_w) &= \mathbb{E}_{\theta}[\widehat{u}_{\text{truth}}(t, T_w; \theta) - \hat{u}_{\text{pred}}(t, t_L, T_w; \theta)]^2 \\ &= \mathbb{E}_{\theta}[\frac{1}{T_w} \int_t^{t+T_w} \sigma \int_{t-t_L}^s e^{(-\gamma+i\omega)(s-s')} dW(s'; \theta) ds]^2 \\ &= \text{var}[\widehat{u}_{\text{truth}}(t, T_w; \theta)] - \frac{\sigma^2}{2\gamma} \frac{e^{-2\gamma t_L}}{(\gamma^2 + \omega^2) T_w^2} [e^{-2\gamma T_w} - 2e^{-\gamma T_w} \cos(\omega T_w) + 1], \end{aligned} \quad (25)$$

where  $\text{var}[\widehat{u}_{truth}(t, T_w; \theta)]$  is given in (52).

We do not present the details of the calculations that lead to (25), since they are relatively long and they provide limited additional insight. To check the accuracy of the formula in (25), we conducted Monte Carlo numerical simulations of the stochastic process in order to find statistical estimates of the forecast skill, and the formula in (25) was in agreement with the statistical estimates (not shown). This same scenario will repeat itself numerous times in the remainder of the paper; namely, many details of calculations will be omitted, but the accuracy of the formulas was verified using numerical statistical estimates.

From the formula in (25), one can see the following properties. With  $\cos(\omega T_w) \leq 1$ , the term  $e^{-2\gamma T_w} - 2e^{-\gamma T_w} \cos(\omega T_w) + 1 \geq (e^{-\gamma T_w} - 1)^2 \geq 0$ , which indicates that  $\widehat{MSE}_u(t_L, T_w) \leq \text{var}[\widehat{u}_{truth}(t, T_w; \theta)]$  always holds. The mean square error is always controlled under the variance of the temporal averaged signal if forecasting with a perfect model. It can be verified that as  $T_w \rightarrow \infty$ ,  $\widehat{MSE}_u(t_L, T_w) \rightarrow 0$  and as  $t_L \rightarrow \infty$ ,  $\widehat{MSE}_u(t_L, T_w) \rightarrow \text{var}[\widehat{u}_{truth}(t, T_w)]$ .

For the Pearson correlation coefficient,

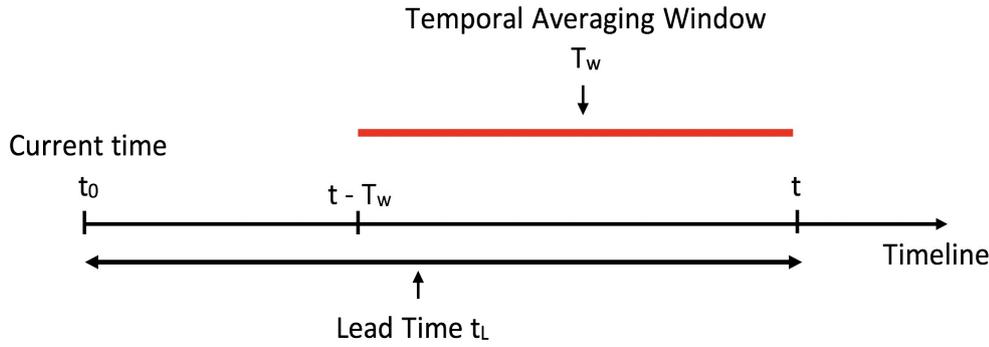
$$\begin{aligned} \widehat{\rho}_u(t_L, T_w) &= \frac{\mathbb{E}_\theta[\widehat{u}_{pred}(t, t_L, T_w; \theta)\widehat{u}_{truth}^*(t; \theta)]}{(\text{var}[\widehat{u}_{pred}(t, t_L, T_w; \theta)])^{1/2}(\text{var}[\widehat{u}_{truth}^*(t; \theta)])^{1/2}} \\ &= \left[ \frac{\text{var}[\widehat{u}_{pred}(t, t_L, T_w; \theta)]}{\text{var}[\widehat{u}_{truth}(t; \theta)]} \right]^{1/2} \\ &= \rho_u(t_L) \left[ \frac{(\gamma^2 + \omega^2)(e^{-2\gamma T_w} - 2e^{-\gamma T_w} \cos(\omega T_w) + 1)}{2\gamma T_w(\gamma^2 + \omega^2) - 2(\gamma^2 - \omega^2) + 2e^{-\gamma T_w} \left( (\gamma^2 - \omega^2) \cos(\omega T_w) - 2\gamma\omega \sin(\omega T_w) \right)} \right]^{1/2}. \quad (26) \end{aligned}$$

The performances of the forecasting skills for different temporal averaging windows and different lead time are demonstrated graphically in Fig. 4 and Fig. 5. (Note that a special behavior can sometimes arise when the temporal averaging window  $T_w$  is an integer multiple of the oscillation period,  $2\pi/\omega$ . For example, in Fig. 5, if  $T_w$  equals the oscillation period of 1.5 months, the prediction will have a very bad forecasting skill. This is mainly because the signal is almost completely averaged out when the temporal averaging window equals an integer multiple of the oscillation period; see also appendix A.)

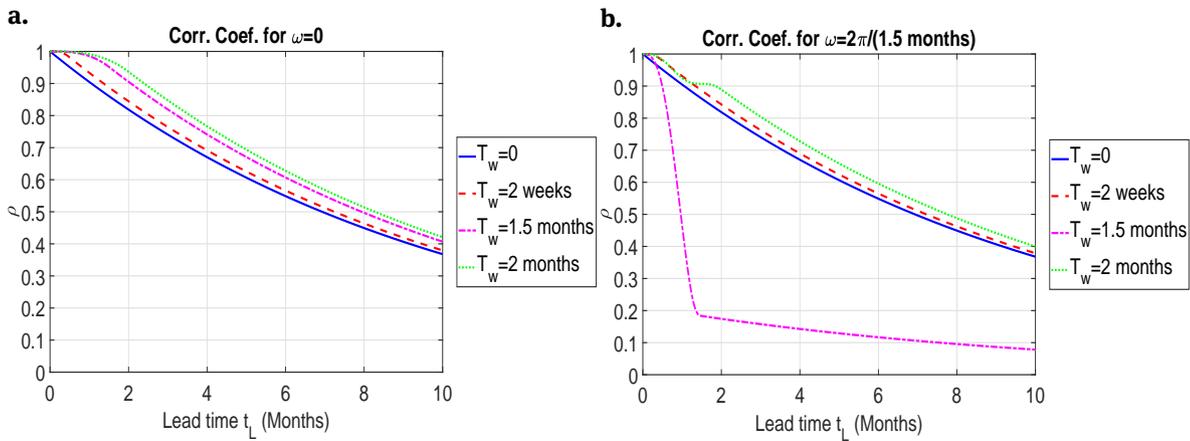
From these two figures 4 and 5, we can see clearly an interesting result: as the temporal averaging window increases, the correlation coefficient actually decreases, which means a worse forecasting for a larger temporal averaging window. This is counter-intuitive. When a temporal averaging window is applied, intuitively we would expect a lot of fast oscillations and unpredictable components would be averaged out, making the resulting signal much easier to predict. But the results from the formulas tell us a totally counter story that the temporal averaging makes the forecast skill worse!

## 4 Subtle Impacts of Definitions of Averaging Window and Lead Time

In the previous section, why did a wider temporal averaging window give a worse forecasting skill, counter to what one would expect from basic intuition? With this question in mind, we have tried to review all of the details of the procedure in order to determine the underlying cause. One key factor seems to be the relative definitions of the lead time and the temporal averaging window, and the relative definition seems to cause differences in the forecasting skills under different settings. To illustrate the different ways of making such definitions, here in this section we use an alternative definition: the time of the right endpoint of a temporal averaging interval is taken to be the “base time” of the averaged signal, which means a temporal averaged



**Figure 6:** Schematic of a new lead time and temporal averaging window definition in the analysis. The horizontal axis represents forecast time from the initial condition at  $t_0$ .



**Figure 7:** Panel a. forecasting skill for different temporal averaging windows of a cOU process with  $\omega = 0$  under the new definition clarified in the Fig. 6 and equation (27); Panel b. forecasting skill for different temporal averaging windows of a cOU process with  $\omega = 2\pi/(1.5 \text{ months})$  under the new definition clarified in the Fig. 6 and equation (27).

signal over a temporal averaging window with length  $T_w$  is defined as

$$\hat{u}_{truth}^R(t; \theta) = \frac{1}{T_w} \int_{t-T_w}^t u_{truth}(s; \theta) ds. \tag{27}$$

A diagram for this new definition is shown in Fig. 6. In this new definition, it needs  $T_w < t_L$  to make the entire temporal averaging window lie in the future. If  $T_w > t_L$ , then part of the averaging window lies in the past which means that only part of the signals in the averaging window that lies in the future needs to be predicted since the past is already known. Comparing the old and new definitions, the forecasting skill can be easily evaluated by replacing the old lead time  $t_L$  in (25) and (26) with  $t_L - T_w$  if  $t_L \geq T_w$ , namely  $\widehat{MSE}_u^R(t_L, T_w) = \widehat{MSE}_u(t_L - T_w, T_w)$  and  $\widehat{\rho}_u^R(t_L, T_w) = \widehat{\rho}_u(t_L - T_w, T_w)$ . If  $t_L < T_w$ , the prediction in this new definition will be made of a part of already known past values and a part of future forecasting. When  $t_L < T_w$ ,  $\widehat{MSE}_u^R(t_L, T_w) = \widehat{MSE}_u(0, t_L) \cdot t_L^2 / T_w^2$  while  $\widehat{\rho}_u^R$  needs to be calculated starting from the new definition. Here, we omit the calculation details and directly give out the formulas.

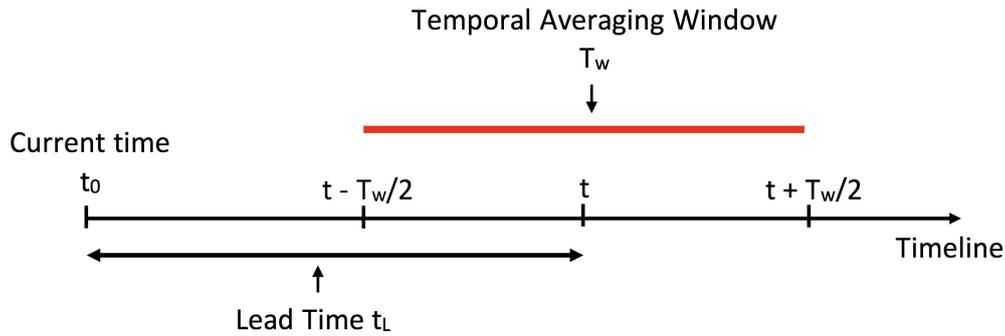
$$\widehat{MSE}_u^R(t_L, T_w) = \begin{cases} \text{var}[\hat{u}_{truth}(t, t_L; \theta)] \frac{t_L^2}{T_w^2} - \frac{\sigma^2}{2\gamma(\gamma^2 + \omega^2)T_w^2} [e^{-2\gamma t_L} - 2e^{-\gamma t_L} \cos(\omega t_L) + 1] & \text{if } t_L < T_w \\ \text{var}[\hat{u}_{truth}(t, T_w; \theta)] - \frac{\sigma^2}{2\gamma(\gamma^2 + \omega^2)T_w^2} [e^{2\gamma T_w} - 2e^{\gamma T_w} \cos(\omega T_w) + 1] & \text{if } t_L \geq T_w \end{cases} \tag{28}$$

$$\hat{\rho}_u^R(t_L, T_w) = \begin{cases} \left[ \frac{\text{var}[\frac{1}{T_w} (\int_{t+t_L-T_w}^t u_{truth}(s; \theta) ds + \int_t^{t+t_L} u_{pred}(s, s-t; \theta) ds)]}{\text{var}[\hat{u}_{truth}(t, T_w; \theta)]} \right]^{1/2} & \text{if } t_L < T_w \\ \rho_u(t_L) \left[ \frac{(\gamma^2 + \omega^2)(e^{2\gamma T_w} - 2e^{\gamma T_w} \cos(\omega T_w) + 1)}{2\gamma T_w (\gamma^2 + \omega^2) - 2(\gamma^2 - \omega^2) + 2e^{-\gamma T_w} ((\gamma^2 - \omega^2) \cos(\omega T_w) - 2\gamma \omega \sin(\omega T_w))} \right]^{1/2} & \text{if } t_L \geq T_w \end{cases} \quad (29)$$

where

$$\begin{aligned} & \text{var}[\frac{1}{T_w} (\int_{t+t_L-T_w}^t u_{truth}(s; \theta) ds + \int_t^{t+t_L} u_{pred}(s, s-t; \theta) ds)] \\ &= \frac{1}{T_w^2} \left[ (T_w - t_L)^2 \text{var}[\hat{u}_{truth}(t, T_w - t_L; \theta)] + \frac{\sigma^2}{2\gamma} \frac{1}{\gamma^2 + \omega^2} (e^{-2\gamma t_L} - 2e^{-\gamma t_L} \cos(\omega t_L) + 1) \right. \\ & \quad \left. + \frac{\sigma^2}{2\gamma} \text{Re}[\frac{2}{(\gamma + i\omega)^2} (1 - e^{(-\gamma - i\omega)t_L})(1 - e^{(-\gamma - i\omega)(T_w - t_L)})] \right] \end{aligned} \quad (30)$$

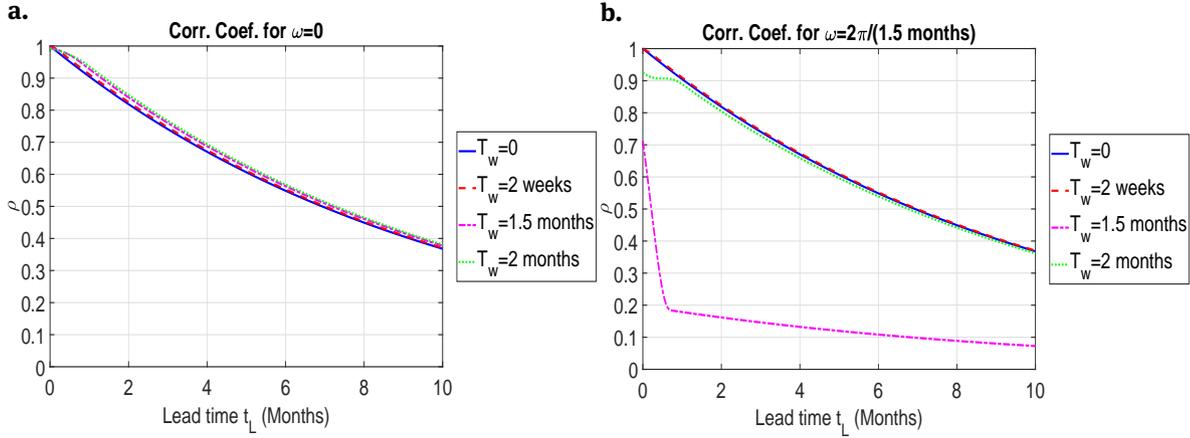
and  $\text{var}[\hat{u}_{truth}(t, T_w; \theta)]$  is given in (52).



**Figure 8:** Schematic of another new lead time and temporal averaging window definition in the analysis. The horizontal axis represents forecast time from the initial condition at  $t_0$ .

From Fig. 7, we care more about the case when  $t_L > T_w$  where the entire temporal averaging window falls in the future. Now, under the new definition, when  $\omega = 0$ , the forecasting skills are improving when a wider temporal averaging window is applied. For  $\omega \neq 0$ , things get a little bit more complicated. Forecasting is very bad when the signal is averaged over a window with length that is multiples of the oscillation period  $2\pi/\omega$ . But the highest skill in the second cycle for  $2\pi/\omega \leq T_w \leq 4\pi/\omega$  is even better than the highest skill in the first cycle when  $0 \leq T_w \leq 2\pi/\omega$  for positive lead time. These results show in a counter way as those under the default definition we discussed in the section 3.2. This is all caused by the different definitions of the averaging window.

Another choice of the temporal averaging window at time  $t$  is to put the window center exactly at the time point  $t$  as in the Fig. 8. With curiosity, we also dig into this case. The formulas of the new forecasting skills  $MSE_u^C(t_L, T_w)$  and  $\rho_u^C(t_L, T_w)$  can be derived directly based on (28) and (29) through the relationship  $MSE_u^C(t_L, T_w) = MSE_u^R(t_L + T_w/2, T_w)$  and  $\rho_u^C(t_L, T_w) = \rho_u^R(t_L + T_w/2, T_w)$  and their behaviors can be observed in Fig. 9. For  $\omega = 0$ , the forecasting skill is similar as but slightly better than the single time point predictions. For the  $\omega = 2\pi/1.5$  months, very bad forecasting skills stay the same at  $T_w$  is multiples of oscillation cycles.



**Figure 9:** Panel **a.** Pearson correlation coefficient of forecasting a cOU process with  $\omega = 0$  under the new definition clarified in the Fig. 8; Panel **b.** Pearson correlation coefficient of forecasting a cOU process with  $\omega = 2\pi/1.5$  months under the new definition clarified in the Fig. 8.

A summary figure, Fig. 10, is plotted to give a comprehensive view of the changes in forecast skill with different definitions of the temporal averaging window. As the center of the temporal averaging window over time  $t$  moves from the left side to the right side of time  $t$ , such as from the definition in Fig. 6 to Fig. 3, the forecast skill becomes worse. The centered definition in Fig. 8 has slightly better forecast skill than the baseline which is the single time point forecasting.

Note that two different  $\gamma$  values are chosen in Fig. 10,  $\gamma = 0.1$  in panel **a** and  $\gamma = 0.5$  in panel **b**. We can see that the differences in forecast skill among different locations of the temporal averaging windows are slight when  $\gamma = 0.1$  as the y-axis in panel **a** of Fig. 10 ranges from about 0.95 to 1.15 but are much more significant when  $\gamma = 0.5$ . This is because the correlation coefficient ratio  $\hat{\rho}(t_L, T_w)/\rho(t_L)$  under different definitions is depending on  $\gamma T_w$ , which can be seen as follows. If we think the baseline is the case of single time point forecasting and then define  $a$  as

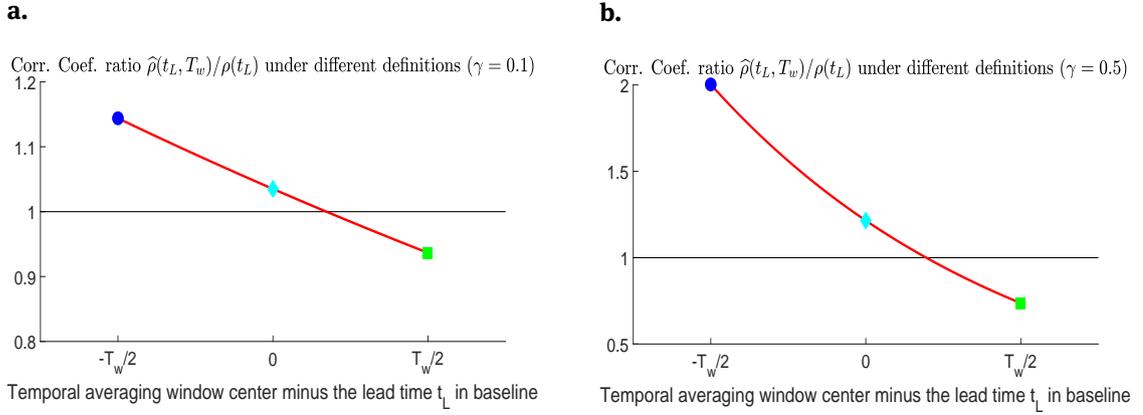
$$a = \frac{\text{The temporal averaging window center minus the lead time } t_L \text{ in baseline}}{T_w}. \quad (31)$$

Then  $a$  is a variable representing where the temporal averaging window is placed. Denote  $\hat{\rho}(t_L, T_w)$  under the definition of the temporal averaging window with  $a$  as  $\hat{\rho}(t_L, T_w, a)$ . When  $\omega = 0$ , the correlation coefficient ratio can be calculated as

$$\begin{aligned} \hat{\rho}(t_L, T_w, a)/\rho(t_L) &= \frac{\hat{\rho}(t_L - (T_w/2 - aT_w), T_w)}{\rho(t_L)} \\ &= (\exp(-\gamma T_w))^a \left[ \frac{\exp(\gamma T_w) + \exp(-\gamma T_w) - 2}{2\gamma T_w + 2\exp(-\gamma T_w) - 2} \right]^{1/2}. \end{aligned} \quad (32)$$

It is depending on  $\gamma T_w$  while independent from  $t_L$ .

In any case, to summarize, it is clear from Fig. 10 that the definition of the time averaging window, relative to the definition of the lead time, has a significant impact on the change in forecast skill. Certain definitions can even lead to the counter-intuitive result that time averaging causes a worse forecast skill. A “centered” time averaging window, which is centered about the target prediction time, appears to offer only a small increase in forecast skill, despite intuitive expectations that time averaging should offer larger increases in forecast skill.



**Figure 10:** Forecast skills (Pearson correlation coefficients) with temporal averaging under different definitions as in Fig. 13 compared to the baseline (single time point forecasting) for predicting cOU processes with  $\omega = 0$ .  $T_w = 2$  months is fixed and the correlation coefficient ratio is not depending on the lead time  $t_L$ . Only cases where the entire averaging window for forecasting lies in the future are considered here (in other words,  $t_L > T_w$ ). The blue dot stands for the definition in Fig. 6 with better forecast skill, the cyan diamond represents the definition of the similar skill as baseline (slightly better) in Fig. 8, the green square is used for the definition of the worse forecast skill in Fig. 3. Panel **a** is for the case  $\gamma = 0.1$  and panel **b** is for the case  $\gamma = 0.5$ .

## 5 Forecasting an Idealized Spatiotemporal Rainfall Model: Stochastic Heat Equation

A complex Ornstein-Uhlenbeck process is an interesting test case since it is a good presentation for a single wave. Nevertheless, spatial averaging issues cannot be conducted in a forecasting problem for a cOU signal. Meanwhile, in the real world, we are often trying to predict signals composed of numerous waves or modes. The cOU process is not enough for getting insights for those situations. Therefore, we proceed to study the forecasting problem for the signal arising from a stochastic heat equation in the form of (8). Now we return to the default definition of the time-averaging window from section 2.3 and take a look at the forecasting behaviors for this type of signals.

### 5.1 Forecasting at a Single Time Point and a Single Spatial Location

The stochastic PDE (8) can be solved analytically. For convenience, let's denote

$$Q(t, x) = q(t, x) - q^* - \tau F, \quad (33)$$

then  $\mathbb{E}[Q(t, x)] = 0$  for  $t \rightarrow \infty$ . Similar to the cOU process, the true signal  $q_{truth}(t, x; \theta)$  is a realization from (8), say it is with underlying parameter  $\theta$  as

$$\begin{aligned} q_{truth}(t, x; \theta) &= q(t, x; \theta) = Q(t, x; \theta) + q^* + \tau F \\ &= q^* + \tau F + \frac{e^{-t_L/\tau}}{(4\pi b_0 t_L)^{1/2}} \int_{\mathbb{R}} e^{-\frac{|x-y|^2}{4b_0 t_L}} Q(t - t_L, y; \theta) dy \\ &\quad + \int_0^{t_L} \frac{1}{(4\pi b_0 |t_L - s|)^{1/2}} \int_{\mathbb{R}} e^{-\frac{|x-y|^2}{4b_0(t-s)}} e^{-(t_L-s)/\tau} \dot{W}(s + t - t_L, y; \theta) dy ds. \end{aligned} \quad (34)$$

A single prediction  $q_{pred}(t, x, t_L; \zeta, \theta)$  for predicting  $q_{truth}(t, x; \theta)$  with lead time  $t_L$  is then a realization with another underlying parameter  $\zeta$ . In other words, it is

$$q_{pred}(t, x, t_L; \zeta, \theta) = q^* + \tau F + \frac{e^{-t_L/\tau}}{(4\pi b_0 t_L)^{1/2}} \int_{\mathbb{R}} e^{-\frac{|x-y|^2}{4b_0 t_L}} Q(t - t_L, y; \theta) dy \\ + \int_0^{t_L} \frac{1}{(4\pi b_0 |t_L - s|)^{1/2}} \int_{\mathbb{R}} e^{-\frac{|x-y|^2}{4b_0(t-s)}} e^{-(t_L-s)/\tau} \dot{W}(s + t - t_L, y; \zeta) dy ds. \quad (35)$$

With lead time  $t_L$ , the ensemble forecasting for time  $t$  is the mean of all the single predictions  $q_{pred}(t, x, t_L; \zeta, \theta)$  starting from  $q_{truth}(t - t_L, x; \theta) = Q(t - t_L, x; \theta) + q^* + \tau F$ , which is

$$q_{pred}(t, x, t_L; \theta) = \mathbb{E}_{\zeta}[q_{pred}(t, x, t_L; \zeta, \theta)] \\ = q^* + \tau F + \frac{e^{-t_L/\tau}}{(4\pi b_0 t_L)^{1/2}} \int_{\mathbb{R}} e^{-\frac{|x-y|^2}{4b_0 t_L}} Q(t - t_L, y; \theta) dy. \quad (36)$$

Now the forecasting skill can be achieved straightforwardly by comparing the  $q_{truth}(t, x; \theta)$  and  $q_{pred}(t, x, t_L; \theta)$  and direct calculation.

$$MSE_q(t_L) = \mathbb{E}_{\theta}[(q_{truth}(t, x; \theta) - q_{pred}(t, x, t_L; \theta))^2] \\ = \mathbb{E}_{\theta}[\left(\int_0^{t_L} \frac{1}{(4\pi b_0 |t_L - s|)^{1/2}} \int_{\mathbb{R}} e^{-\frac{|x-y|^2}{4b_0(t-s)}} e^{-(t_L-s)/\tau} \dot{W}(s + t - t_L, y; \theta) dy ds\right)^2] \\ = \frac{D_*^2}{4} \sqrt{\frac{\tau}{b_0}} \operatorname{erf}\left(\sqrt{\frac{2t_L}{\tau}}\right). \quad (37)$$

Note  $\frac{D_*^2}{4} \sqrt{\frac{\tau}{b_0}}$  is the variance of the true signal. Hence the mean square error is again controlled under the variance of the true signal with the perfect model.

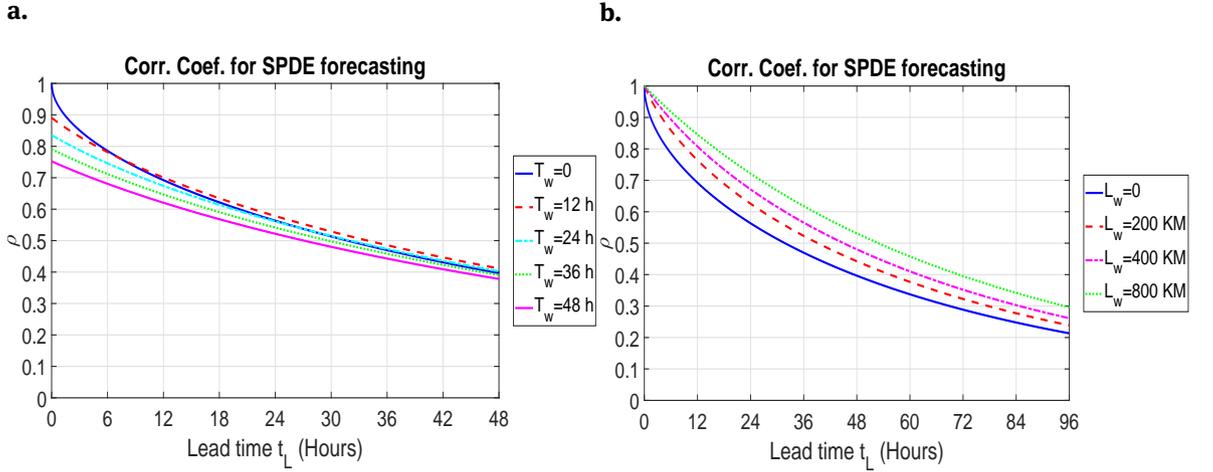
$$\rho_q(t_L) = \frac{\mathbb{E}_{\theta}[(q_{pred}(t, x, t_L; \theta) - \mathbb{E}_{\theta}[q_{pred}(t, x, t_L; \theta)])(q_{truth}(t, x; \theta) - \mathbb{E}_{\theta}[q_{truth}(t, x; \theta)])]}{(\operatorname{var}[q_{pred}(t, x, t_L; \theta)])^{1/2} (\operatorname{var}[q_{truth}(t, x; \theta)])^{1/2}} \\ = \left(\frac{\operatorname{var}[q_{pred}(t, x, t_L; \theta)]}{\operatorname{var}[q_{truth}(t, x; \theta)]}\right)^{1/2} \\ = [1 - \operatorname{erf}\left(\sqrt{\frac{2t_L}{\tau}}\right)]^{1/2}. \quad (38)$$

Essentially, the forecasting skill at a single time point and a single spatial location is only related to the ratio of  $t_L$  and  $\tau$  and doesn't involve any other parameters in the model. With  $t_L \leq 0.33\tau$ ,  $\rho_q(t_L) \geq 0.5$ . In our simulation, for  $\tau = 96h = 4 \text{ days}$  in the idealized rainfall model,  $t_L \leq 1.32344 \text{ days} = 31.76256 \text{ hours}$  will have a good forecast correlation coefficient  $\rho_q(t_L) \geq 0.5$ .

## 5.2 Forecasting at a Single Spatial Location with Temporal Averaging

Temporal averaging forecasting skill for a single spatial location for the signal from the stochastic heat equation is studied as well. The temporal averaged signal is

$$\widehat{q}_{truth}(t, x, T_w; \theta) = \frac{1}{T_w} \int_t^{t+T_w} q_{truth}(s, x; \theta) ds = \frac{1}{T_w} \int_t^{t+T_w} Q(s, x; \theta) ds + q^* + \tau F, \quad (39)$$



**Figure 11:** Panel **a.** forecasting skill for different temporal averaging windows of a SPDE process under the default definition clarified in the section 2.3; Panel **b.** forecasting skill for different spatial averaging windows of a SPDE process under the default definition clarified in the section 2.3.

where

$$\begin{aligned}
 Q(s, x; \theta) &= \frac{e^{-(s-(t-t_L))/\tau}}{[4\pi b_0(s-(t-t_L))]^{1/2}} \int_{\mathbb{R}} e^{-\frac{|x-y|^2}{4b_0(s-(t-t_L))}} Q(t-t_L, y; \theta) dy \\
 &+ \int_0^{s-(t-t_L)} \frac{1}{(4\pi b_0|s-(t-t_L)-\mu|)^{1/2}} \int_{\mathbb{R}} e^{-\frac{|x-y|^2}{4b_0(s-(t-t_L)-\mu)}} e^{-(s-(t-t_L)-\mu)/\tau} \xi(\mu+t-t_L, y; \theta) dy d\mu \quad (40)
 \end{aligned}$$

and the ensemble forecasting is achieved by taking the ensemble mean of all the single predictions

$$\widehat{q}_{pred}(t, x, t_L, T_w; \theta) = \frac{1}{T_w} \int_t^{t+T_w} \frac{e^{-a(s-(t-t_L))}}{[4\pi b(s-(t-t_L))]^{1/2}} \int_{\mathbb{R}} e^{-\frac{|x-y|^2}{4b(s-(t-t_L))}} Q(t-t_L, y; \theta) dy ds + q^* + \tau F, \quad (41)$$

$$\begin{aligned}
 \widehat{MSE}_q(t_L, T_w) &= \mathbb{E}_{\theta}[(\widehat{q}_{truth}(t, x, T_w; \theta) - \widehat{q}_{pred}(t, x, t_L, T_w; \theta))^2] \\
 &= \frac{D_x^2}{4} \sqrt{\frac{\tau}{b_0}} \frac{\tau^2}{T_w^2} \left[ \frac{2(t_L+T_w)}{\tau} \int_{(2t_L+T_w)/\tau}^{2(t_L+T_w)/\tau} \operatorname{erf}(\sqrt{x}) dx - \frac{2t_L}{\tau} \int_{2t_L/\tau}^{(2t_L+T_w)/\tau} \operatorname{erf}(\sqrt{x}) dx \right. \\
 &\quad - \frac{2T_w}{\tau} \int_0^{T_w/\tau} \operatorname{erf}(\sqrt{x}) dx + \int_{2t_L/\tau}^{(2t_L+T_w)/\tau} x \operatorname{erf}(\sqrt{x}) dx \\
 &\quad \left. - \int_{(2t_L+T_w)/\tau}^{2(t_L+T_w)/\tau} x \operatorname{erf}(\sqrt{x}) dx + 2 \int_0^{T_w/\tau} x \operatorname{erf}(\sqrt{x}) dx \right], \quad (42)
 \end{aligned}$$

where

$$\int_A^B \operatorname{erf}(\sqrt{x}) dx = \operatorname{Berf}(\sqrt{B}) - \operatorname{Aerf}(\sqrt{A}) - \frac{1}{2} \operatorname{erf}(\sqrt{B}) + \frac{1}{2} \operatorname{erf}(\sqrt{A})$$

$$+ \sqrt{\frac{B}{\pi}} \exp(-B) - \sqrt{\frac{A}{\pi}} \exp(-A) \quad (43)$$

and

$$\begin{aligned} \int_A^B x \operatorname{erf}(\sqrt{x}) dx &= \frac{1}{2} B^2 \operatorname{erf}(\sqrt{B}) - \frac{1}{2} A^2 \operatorname{erf}(\sqrt{A}) - \frac{3}{8} \operatorname{erf}(\sqrt{B}) + \frac{3}{8} \operatorname{erf}(\sqrt{A}) \\ &+ \frac{1}{2\sqrt{\pi}} B^{3/2} \exp(-B) - \frac{1}{2\sqrt{\pi}} A^{3/2} \exp(-A) \\ &+ \frac{3}{4\sqrt{\pi}} B^{1/2} \exp(-B) - \frac{3}{4\sqrt{\pi}} A^{1/2} \exp(-A). \end{aligned} \quad (44)$$

The correlation coefficient can again be proved to be the square root of the ratio between the two variances of the true averaged signal and the prediction as

$$\hat{\rho}_q(t_L, T_w) = \left( \frac{\operatorname{var}[\hat{q}_{pred}(t, x, t_L, T_w; \theta)]}{\operatorname{var}[\hat{q}_{truth}(t, x, T_w; \theta)]} \right)^{1/2}, \quad (45)$$

where

$$\begin{aligned} \operatorname{var}[\hat{q}_{pred}(t, x, t_L, T_w; \theta)] &= \frac{D_*^2}{4} \sqrt{\frac{\tau}{b_0}} \left\{ 1 - \frac{\tau^2}{T_w^2} \left[ \int_{2t_L/\tau}^{(2t_L+T_w)/\tau} (x \operatorname{erf}(\sqrt{x})) dx - \int_{(2t_L+T_w)/\tau}^{2(t_L+T_w)/\tau} x \operatorname{erf}(\sqrt{x}) dx \right. \right. \\ &\left. \left. - \frac{2t_L}{\tau} \int_{2t_L/\tau}^{(2t_L+T_w)/\tau} \operatorname{erf}(\sqrt{x}) dx + \frac{2(t_L+T_w)}{\tau} \int_{(2t_L+T_w)/\tau}^{2(t_L+T_w)/\tau} \operatorname{erf}(\sqrt{x}) dx \right] \right\} \end{aligned} \quad (46)$$

and  $\operatorname{var}[\hat{q}_{truth}(t, x, T_w; \theta)]$  is given in (53).

For positive lead time, the skill has slight improvements for very narrow temporal averaging window  $T_w$  and then starts to drop as the averaging window becomes wider and wider. Fig. 11 shows this characteristic when lead time  $t_L$  is larger than about 6 hours. The SHE signal is made up of a lot of cOU waves with different decorrelation times but no oscillations if we see it from Fourier space. Recalling the discussion from the cOU part, the fast decay waves with short decorrelation times are hard to predict and those slow decay waves with long decorrelation times are predictable. When a temporal averaging window is used, we average out the fast decayed waves which helps to improve the forecasting skill while at the same time the forecast skill for each individual predictable wave is dropping that may deteriorate the forecasting. The improvement part can be induced both from intuition and the correlation formula. When we do the temporal averaging window, we removed those unpredictable waves and diminish the overall variance of the signal, making it easier to predict. The variance contribution of the predictions for those fast decayed waves is always almost zero in  $\operatorname{var}[\hat{q}_{pred}(t, x, t_L, T_w; \theta)]$  while averaging them out definitely reduces the variance of the true signal  $\hat{q}_{truth}(t, x, T_w; \theta)$ . The Pearson correlation coefficient appears to be the ratio of these two values indicating the averaging out of those fast decayed waves will help the forecasting. The forecasting on those slow decayed waves is worse when a temporal averaging window under the default definition in section 2.3 is applied as we stated before. Hence, the overall change in the forecasting skill is a trade off between the improvements brought by averaging out the fast decayed waves and the decline caused by the worse forecasting for those slow decayed waves.

As we know before, whether the forecasting skill will become better or worse for a predictable wave when temporal averaging is used is highly depending on how we place the temporal averaging window. If a temporal averaging window is placed as in Fig. 6, then the forecasting for the SHE signal is keeping climbing when a wider temporal averaging window is applied since both the averaging out of the fast decayed signals and the averaging of the slow decayed signals are helping the forecasting. The same thing will happen if we invoke the averaging definition in Fig. 8 with the same reasons.

### 5.3 Forecasting at a Single Time Point with Spatial Averaging

For the spatial averaging, we define the spatial averaged signal at location  $x$  with spatial window length  $L_w$  by averaging all the signals over the interval  $[x - L_w/2, x + L_w/2]$  as in (14), namely the averaged truth signal  $\tilde{q}_{truth}(t, x, L_w; \theta)$  and its corresponding prediction  $\tilde{q}_{pred}(t, x, t_L, L_w; \theta)$  are

$$\tilde{q}_{truth}(t, x, L_w; \theta) = \frac{1}{L_w} \int_{x-L_w/2}^{x+L_w/2} q_{truth}(t, y; \theta) dy \quad (47)$$

and

$$\tilde{q}_{pred}(t, x, t_L, L_w; \theta) = \frac{1}{L_w} \int_{x-L_w/2}^{x+L_w/2} q_{pred}(t, y, t_L; \theta) dy. \quad (48)$$

We omit more details in this section and directly give the formulas for the forecasting skills as following.

$$\begin{aligned} \widetilde{MSE}_q(t_L, L_w) &= \frac{D_*^2}{4} \sqrt{\frac{\tau}{b_0}} \left\{ \frac{2\sqrt{b_0\tau}}{L_w} - \frac{2b_0\tau}{L_w^2} \operatorname{erf}(\sqrt{2t_L/\tau}) \right. \\ &\quad + \frac{b_0\tau}{L_w^2} \exp\left(-\frac{L_w}{\sqrt{b_0\tau}}\right) \left(1 - \operatorname{erf}\left(\frac{L_w}{\sqrt{8b_0t_L}} - \sqrt{2t_L/\tau}\right)\right) \\ &\quad - \frac{b_0\tau}{L_w^2} \exp\left(\frac{L_w}{\sqrt{b_0\tau}}\right) \left(1 - \operatorname{erf}\left(\frac{L_w}{\sqrt{8b_0t_L}} + \sqrt{2t_L/\tau}\right)\right) \\ &\quad \left. - 2\frac{\sqrt{b_0\tau}}{L_w} \exp(-2t_L/\tau) \left(\operatorname{erf}\left(\frac{L_w}{\sqrt{8b_0t_L}}\right) - \frac{\sqrt{8b_0t_L}}{\sqrt{\pi}L_w} (1 - \exp(-\frac{L_w^2}{8b_0t_L}))\right) \right\}, \quad (49) \end{aligned}$$

$$\tilde{\rho}_q(t_L, L_w) = \left( \frac{\operatorname{var}[\tilde{q}_{pred}(t, x, t_L, L_w; \theta)]}{\operatorname{var}[\tilde{q}_{truth}(t, x, L_w; \theta)]} \right)^{1/2}, \quad (50)$$

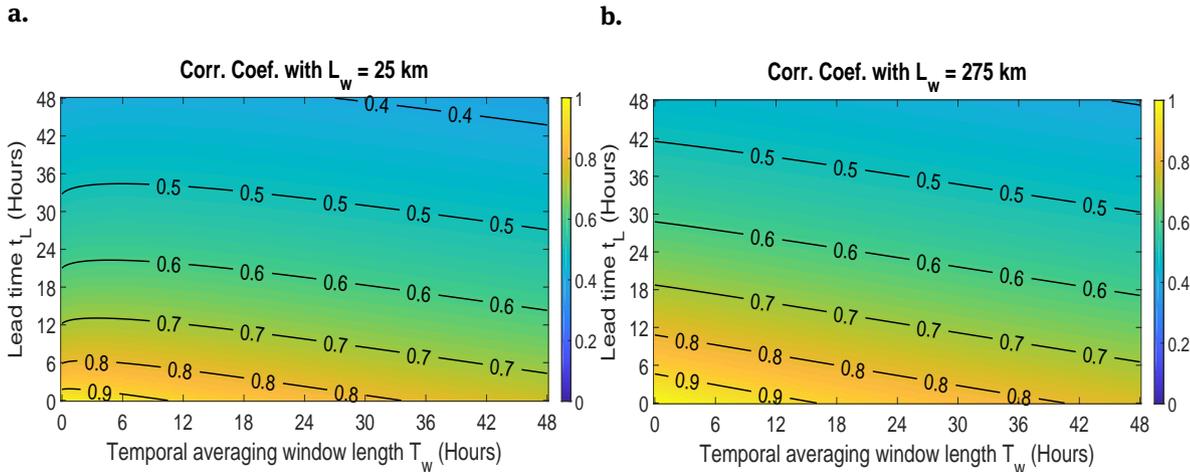
where

$$\begin{aligned} \operatorname{var}[\tilde{q}_{pred}(t, x, t_L, L_w; \theta)] &= \frac{D_*^2}{4} \sqrt{\frac{\tau}{b_0}} \left\{ \frac{b_0\tau}{L_w^2} \left[ \exp\left(-\frac{L_w}{\sqrt{b_0\tau}}\right) + \exp\left(\frac{L_w}{\sqrt{b_0\tau}}\right) - 2 \right] \right. \\ &\quad \frac{b_0\tau}{L_w^2} \exp\left(-\frac{L_w}{\sqrt{b_0\tau}}\right) \operatorname{erf}\left(\frac{L_w}{\sqrt{8b_0t_L}} - \sqrt{2t_L/\tau}\right) \\ &\quad + b_0\tau \exp\left(\frac{L_w}{\sqrt{b_0\tau}}\right) \operatorname{erf}\left(-\frac{L_w}{\sqrt{8b_0t_L}} - \sqrt{2t_L/\tau}\right) \\ &\quad - \frac{2b_0\tau}{L_w^2} \operatorname{erf}(-\sqrt{2t_L/\tau}) + \frac{2\sqrt{b_0\tau}}{L_w} \exp(-2t_L/\tau) \operatorname{erf}\left(\frac{L_w}{\sqrt{8b_0t_L}}\right) \\ &\quad \left. - \frac{2\sqrt{b_0\tau}}{\sqrt{\pi}L_w} \frac{\sqrt{8b_0t_L}}{L_w} \exp(-2t_L/\tau) (1 - \exp(-\frac{L_w^2}{8b_0t_L})) \right\} \quad (51) \end{aligned}$$

and  $\operatorname{var}[\tilde{q}_{truth}(t, x, L_w; \theta)]$  is given in (54).

The spatial averaging is helping to improve the forecasting skills unambiguously. A wider spatial averaging window produce better forecasting. Since the spatial averaging window is independent from the lead time, we would expect the improvements in forecasting brought by the spatial averaging no matter whether the averaging window at location  $x$  is placed more on the left side of  $x$ , centered at  $x$  or more on the right side of  $x$ .

## 5.4 Forecasting with Both Temporal and Spatial Averaging



**Figure 12:** Panel **a.** forecasting skill of a temporal averaged SPDE process under the default definition clarified in the section 2.3 with a fixed spatial averaging window  $L_w = 25$  km; Panel **b.** forecasting skill of a temporal averaged SPDE process under the default definition clarified in the section 2.3 with a fixed spatial averaging window  $L_w = 275$  km.

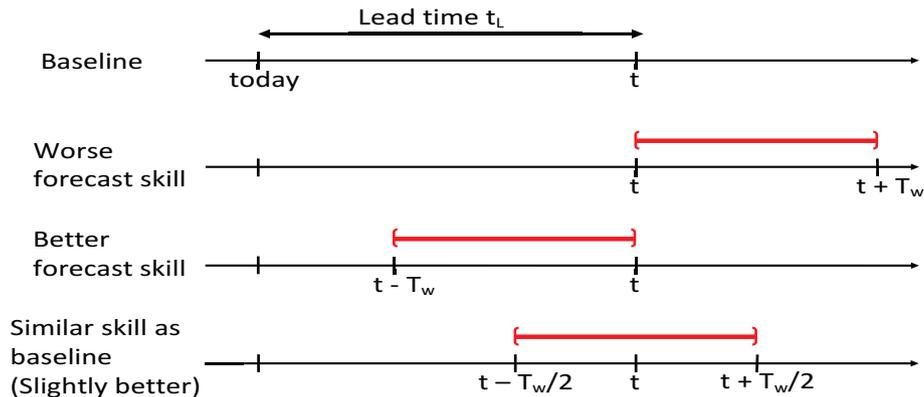
Regarding the formulas of forecast skills with both temporal and spatial averaging for predicting a SHE signal, see in appendix section B. Instead, we take a look at the patterns through Fig. 12 here. The forecasting performance is the combination effects of the temporal and spatial averaging where the spatial averaging always helps to improve the forecasting while the temporal averaging helps the forecasting for very narrow windows and then make the forecasting worse for wider windows. If the definitions shown in Fig. 6 or Fig. 8 are used, then the temporal averaging is also helping the forecasting so that the skills are always improving with wider spatial and temporal windows in those cases.

## 6 Conclusions

The effect of time averaging on forecast skill was seen to be significantly different depending on the definitions of the temporal averaging window and the lead time. Two main competing effects contribute to this difference. On the one hand, time averaging should increase forecasting skill by eliminating (“averaging out”) high-frequency, nearly unpredictable components from the signal; such an effect is consistent with some a priori, basic intuition. On the other hand, time averaging can decrease forecasting skill if the time-averaging window  $[t, t + T_w]$  lies in the future, beyond the time  $t$  of the prediction, since the signal is less predictable for times that are farther in the future. As a result, it is difficult to make a general statement about whether time averaging leads to an increase or decrease in forecasting skill. The answer depends on the relative definitions of the time-averaging window and the lead time; see Fig. 13 for a schematic diagram.

Spatial averaging was seen to always improve the forecast skill. It eliminates (“averages out”) high-frequency, nearly unpredictable components from the signal, and it leaves behind only the low-frequency, more-predictable components of the signal. Such behavior is consistent with a priori, basic intuition. No subtlety exists in the relative definition of lead time and a *spatial* averaging window.

What is the “correct” definition of lead time, relative to the time-averaging window? One might argue that the lead time should be defined with respect to the *end* of an averaging window,  $[t - T_w, t]$ . Such a definition seems sensible for a number of reasons. For instance, in terms of the initial data for the forecast, the time-averaged state over the past  $T_w$  time units would be known at the initial time. Also, this definition leads to



**Figure 13:** Schematics of comparison between different definitions of temporal averaging window and single time point predictions and brief conclusions for predicting cOU processes with  $\omega = 0$ .

enhanced prediction skill, which is in line with some basic intuition for the impact of averaging on prediction skill, and it seems desirable to have agreement with basic intuition. However, it is possible that the enhanced prediction skill is not entirely due to the averaging itself; instead, it is likely enhanced at least partially due to the fact that the time-averaging window  $[t - T_w, t]$  contains signal data *before* the target time  $t$ , and this “earlier” data should naturally be more predictable than the “later” data at time  $t$ . Consequently, it is unclear whether there is a “best” definition of lead time, relative to the time-averaging window, since several competing factors are intermingled.

The models here were chosen to be exactly solvable and to be somewhat representative of atmospheric, oceanic, and climate variability. Therefore, the results here could potentially be expected to hold to some degree even for more complicated dynamical models. However, it is possible that nonlinearities [5–9, 15, 23, 26, 34] and other factors could potentially impact the conclusions, and it would be interesting to investigate similar questions with more complicated models in the future. As one small step in this direction, we have added advection to the stochastic heat equation by replacing  $\partial_t q$  by  $\partial_t q + c \partial_x q$  with an advection speed of  $c \approx 15$  m/s; numerical results of forecasts with time averaging (not shown) are essentially the same as the main conclusions presented above. It would be interesting to investigate other models in the future.

**Acknowledgments:** The authors thank two anonymous reviewers for helpful comments. The research of S.N.S. is partially supported by a Sloan Research Fellowship from the Alfred P. Sloan Foundation and a Vilas Associates Award from the University of Wisconsin–Madison. Y.L. is partially supported as a graduate student researcher by the Vilas Associates Award.

## Appendix

### A Temporal and Spatial Averaging: Model Statistics

Before we evaluate the forecasting skill, we need to investigate a little bit in the change of variance of the signal through the temporal averaging and spatial averaging. Relative small mean square errors are expected for forecasting with temporal or spatial averaging since averaging will create a smaller variance for the averaged

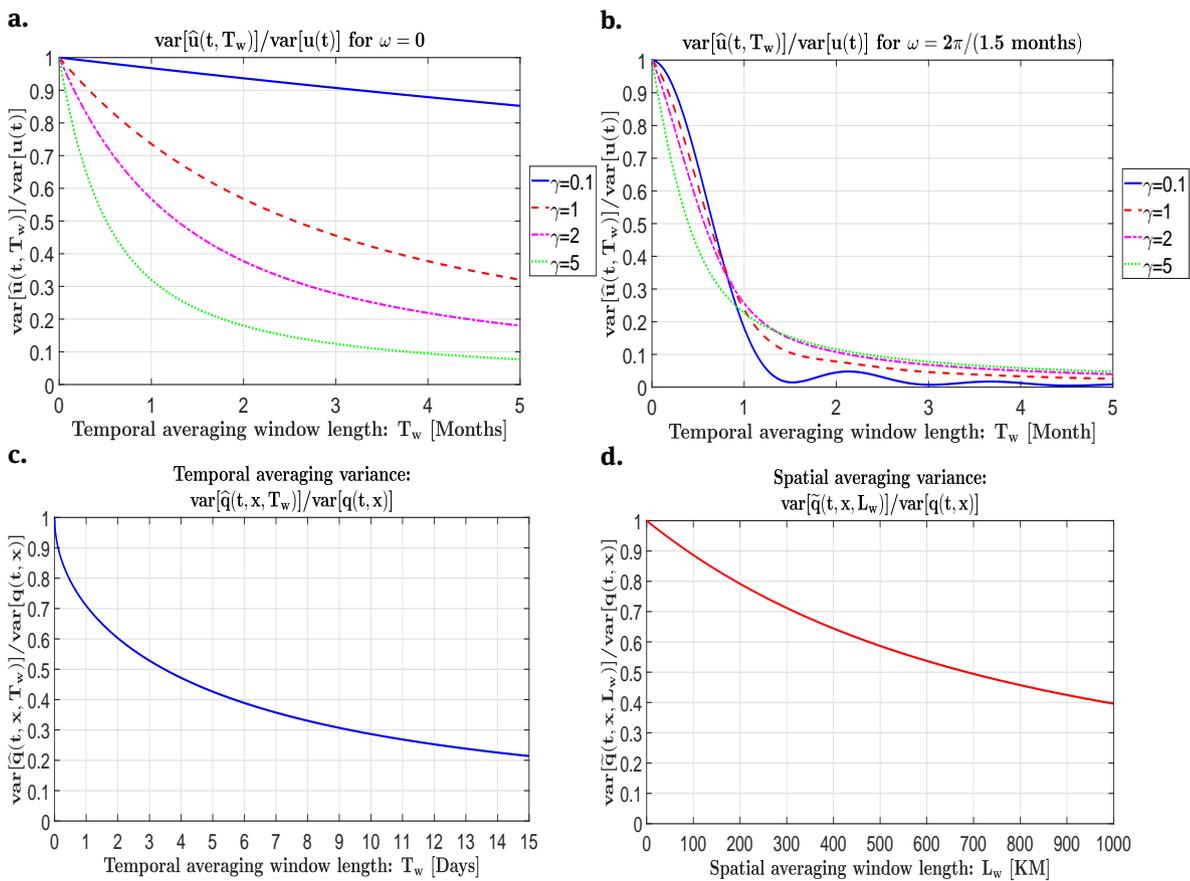
signal. The trends of the variance changes as we do the averaging are shown in Fig. 14. For a complex OU process from (1), after approaching the stationary distribution arbitrary closely, the averaged mean will be nearly 0 while the variance of the averaged signal over a temporal window with length  $T_w$  is close to the stationary value of

$$\begin{aligned} \text{var}[\hat{u}(t, T_w)] &= \frac{\sigma^2}{2\gamma} \cdot \frac{1}{(\gamma^2 + \omega^2)^2 T_w^2} \cdot [2\gamma T_w(\gamma^2 + \omega^2) - 2(\gamma^2 - \omega^2) \\ &\quad + 2e^{-\gamma T_w} ((\gamma^2 - \omega^2)\cos(\omega T_w) - 2\gamma\omega\sin(\omega T_w))], \end{aligned} \tag{52}$$

where  $\sigma^2/2\gamma$  is the variance of the unaveraged cOU signal and this variance has properties

$$\lim_{T_w \rightarrow 0} \text{var}[\hat{u}(t, T_w)] = \frac{\sigma^2}{2\gamma} = \text{var}[u(t)], \quad \lim_{T_w \rightarrow \infty} \text{var}[\hat{u}(t, T_w)] = 0,$$

which agree with one's intuition that the signal will be averaged out if the temporal averaging window is very



**Figure 14:** Panel **a.** variance change over temporal averaging for  $\omega = 0$  and different  $\gamma$  of a cOU process; Panel **b.** variance change over temporal averaging for  $\omega = 2\pi/(1.5 \text{ months})$  and different  $\gamma$  of a COU process ; Panel **c.** variance change over temporal averaging for a stochastic PDE process; Panel **d.** variance change over spatial averaging for a stochastic PDE process

large. From the panel **a** of Fig. 14, we can see that the variance of the cOU signal is monotone decreasing if  $\omega = 0$  as we increase the length of the temporal averaging window. When  $\omega \neq 0$ , the variance of the averaged cOU signal is decreasing at first if we enlarge the temporal averaging window and the signal is almost averaged out after a time period  $2\pi/\omega$  according to that its variance is almost below 15% of the original variance after a time period in panel **b** of Fig. 14. As the decorrelation time  $1/\gamma$  decreases ( $\gamma$  increases), the declining rate of the variance is changing from a cosine type decreasing to an exponential decreasing. In the

case  $\omega = 0$ , the variance decline is dominated by  $\gamma$  only. In the case  $\omega \neq 0$ , the variance always decreases to a very small amount after a time period regardless of the values of  $\gamma$ .

Regarding the stationary process stemming from the stochastic heat equation (8), both the impacts of the temporal and spatial averaging on the variance have to be considered. The averaged signal always has the same mean  $q^* + \tau F$ . Through the temporal averaging with a temporal window of length  $T_w$ , the variance of the temporal averaged signal becomes

$$\begin{aligned} \text{var}[\hat{q}(t, x, T_w)] &= \frac{D_*^2}{4} \sqrt{\frac{\tau}{b_0}} \left[ 1 - \text{erf}\left(\sqrt{\frac{T_w}{\tau}}\right) \left( \frac{3}{4} \left(\frac{\tau}{T_w}\right)^2 - \frac{\tau}{T_w} + 1 \right) \right. \\ &\quad \left. + \exp\left(-\frac{T_w}{\tau}\right) \left( \frac{3}{2\sqrt{\pi}} \left(\frac{\tau}{T_w}\right)^{3/2} - \frac{1}{\sqrt{\pi}} \left(\frac{\tau}{T_w}\right)^{1/2} \right) \right]. \end{aligned} \quad (53)$$

Through the spatial averaging with a spatial window of length  $L_w$ , the variance of the spatial averaged signal will be

$$\text{var}[\tilde{q}(t, x, L_w)] = \frac{D_*^2}{4} \sqrt{\frac{\tau}{b_0}} \cdot 2 \left[ \frac{\sqrt{b_0 \tau}}{L_w} - \frac{b_0 \tau}{L_w^2} \left( 1 - \exp\left(-\frac{L_w}{\sqrt{b_0 \tau}}\right) \right) \right], \quad (54)$$

where  $\frac{D_*^2}{4} \sqrt{\frac{\tau}{b_0}}$  is the variance of the unaveraged stochastic partial differential equation signal. The variance is decreasing monotonically as the averaging window increases in both the temporal averaging case and the spatial averaging case for the stochastic partial differential equation process by panel **c** and panel **d** of Fig. 14. By the analytic formulas (53) and (54), the variance of the temporal averaged SPDE signal compared to the original unaveraged variance is determined by the ratio of the temporal averaging window  $T_w$  and the relaxation time  $\tau$  only and not related to other parameters in the model (8). The variance of the spatial averaged SPDE signal compared to the original unaveraged variance is determined by the ratio of square root of  $b_0 \tau$  and the spatial averaging window  $L_w$  only.

## B Formulas of forecasting skill for a both temporal and spatial averaged SPDE

- Mean Square Error  $\overline{MSE}_q(t_L, T_w, L_w)$

$$\begin{aligned} \overline{MSE}_q(t_L, T_w, L_w) &= \frac{D_*^2}{4} \sqrt{\frac{\tau}{b_0}} \frac{1}{L_w^2 T_w^2} \frac{\tau^2}{2} \left[ \frac{2T_w}{\tau} (\mathcal{P}(T_w/\tau, L_w) - \mathcal{P}(0, L_w)) \right. \\ &\quad + \frac{2t_L}{\tau} (\mathcal{P}((2t_L + T_w)/\tau, L_w) - \mathcal{P}(2t_L/\tau, L_w)) \\ &\quad - \frac{2(t_L + T_w)}{\tau} (\mathcal{P}(2(t_L + T_w)/\tau, L_w) - \mathcal{P}((2t_L + T_w)/\tau, L_w)) \\ &\quad - 2(\mathcal{Q}(T_w/\tau, L_w) - \mathcal{Q}(0, L_w)) \\ &\quad - (\mathcal{Q}((2t_L + T_w)/\tau, L_w) - \mathcal{Q}(2t_L/\tau, L_w)) \\ &\quad \left. + (\mathcal{Q}(2(t_L + T_w), L_w) - \mathcal{Q}((2t_L + T_w)/\tau, L_w)) \right], \end{aligned} \quad (55)$$

where

$$\begin{aligned} \mathcal{P}(x, y) &= \sqrt{\frac{1}{b_0 \tau}} [(h_2(x, y) + g_2(x, y)) - y(h_1(x, y) + g_1(x, y))] \\ &\quad + (2x - 1)[(h_1(x, y) - g_1(x, y)) - y(h_0(x, y) - g_0(x, y))] \\ &\quad - 4\sqrt{b_0 \tau} \cdot x \cdot \exp(-x) \left( y \cdot \text{erf}\left(\frac{y}{2\sqrt{b_0 \tau x}}\right) - 2\sqrt{\frac{b_0 \tau x}{\pi}} \left( 1 - \exp\left(-\frac{y^2}{4b_0 \tau x}\right) \right) \right), \end{aligned} \quad (56)$$

$$\begin{aligned}
\Omega(x, y) = & -\frac{1}{4b_0\tau}[(h_3(x, y) - g_3(x, y)) - y(h_2(x, y) - g_2(x, y))] \\
& + \frac{3}{4} \frac{1}{\sqrt{b_0\tau}}[(h_2(x, y) + g_2(x, y)) - y(h_1(x, y) + g_1(x, y))] \\
& + (x^2 - \frac{3}{4})[(h_1(x, y) - g_1(x, y)) - y(h_0(x, y) - g_0(x, y))] \\
& - \sqrt{b_0\tau}(2x + 3)x \cdot \exp(-x) \left( y \cdot \operatorname{erf}\left(\frac{y}{2\sqrt{b_0\tau x}}\right) - 2\sqrt{\frac{b_0\tau x}{\pi}}(1 - \exp(-\frac{y^2}{4b_0\tau x})) \right),
\end{aligned} \tag{57}$$

$$h_0(x, y) = \int_0^y \exp\left(-\frac{z}{\sqrt{b_0\tau}}\right) \operatorname{erf}\left(\frac{z}{2\sqrt{b_0\tau x}} + \sqrt{x}\right) dz, \tag{58}$$

$$h_1(x, y) = \int_0^y z \cdot \exp\left(-\frac{z}{\sqrt{b_0\tau}}\right) \operatorname{erf}\left(\frac{z}{2\sqrt{b_0\tau x}} + \sqrt{x}\right) dz, \tag{59}$$

$$h_2(x, y) = \int_0^y z^2 \cdot \exp\left(-\frac{z}{\sqrt{b_0\tau}}\right) \operatorname{erf}\left(\frac{z}{2\sqrt{b_0\tau x}} + \sqrt{x}\right) dz, \tag{60}$$

$$h_3(x, y) = \int_0^y z^3 \cdot \exp\left(-\frac{z}{\sqrt{b_0\tau}}\right) \operatorname{erf}\left(\frac{z}{2\sqrt{b_0\tau x}} + \sqrt{x}\right) dz, \tag{61}$$

$$g_0(x, y) = \int_0^y \exp\left(-\frac{z}{\sqrt{b_0\tau}}\right) \operatorname{erf}\left(\frac{z}{2\sqrt{b_0\tau x}} - \sqrt{x}\right) dz, \tag{62}$$

$$g_1(x, y) = \int_0^y z \cdot \exp\left(-\frac{z}{\sqrt{b_0\tau}}\right) \operatorname{erf}\left(\frac{z}{2\sqrt{b_0\tau x}} - \sqrt{x}\right) dz, \tag{63}$$

$$g_2(x, y) = \int_0^y z^2 \cdot \exp\left(-\frac{z}{\sqrt{b_0\tau}}\right) \operatorname{erf}\left(\frac{z}{2\sqrt{b_0\tau x}} - \sqrt{x}\right) dz, \tag{64}$$

$$g_3(x, y) = \int_0^y z^3 \cdot \exp\left(-\frac{z}{\sqrt{b_0\tau}}\right) \operatorname{erf}\left(\frac{z}{2\sqrt{b_0\tau x}} - \sqrt{x}\right) dz. \tag{65}$$

- Correlation coefficient  $\bar{\rho}_q(t_L, T_w, L_w)$

$$\bar{\rho}_q(t_L, T_w, L_w) = \left( \frac{\operatorname{var}[\bar{q}_{\text{pred}}(t, x, t_L, T_w, L_w; \theta)]}{\operatorname{var}[\bar{q}_{\text{truth}}(t, x, T_w, L_w; \theta)]} \right)^{1/2}, \tag{66}$$

$$\begin{aligned}
\operatorname{var}[\bar{q}_{\text{truth}}(t, x, T_w, L_w; \theta)] = & \frac{D_*^2}{4} \sqrt{\frac{\tau}{b_0}} \frac{1}{L_w^2} \left\{ \frac{\tau^2}{T_w^2} \left[ \frac{T_w}{\tau} (\mathcal{P}(T_w/\tau, L_w) - \mathcal{P}(0, L_w)) \right. \right. \\
& \left. \left. (\Omega(T_w/\tau, L_w) - \Omega(0, L_w)) \right] \right. \\
& \left. + b_0\tau \left( \exp\left(\frac{L_w}{\sqrt{b_0\tau}}\right) + \exp\left(\frac{L_w}{\sqrt{b_0\tau}}\right) - 2 \right) \right\},
\end{aligned} \tag{67}$$

$$\begin{aligned} \text{var}[\bar{q}_{pred}(t, x, t_L, T_w, L_w; \theta)] &= \frac{D_*^2}{4} \sqrt{\frac{\tau}{b_0}} \frac{1}{L_w^2} \left\{ b_0 \tau \left[ \exp\left(\frac{L_w}{\sqrt{b_0 \tau}}\right) + \exp\left(-\frac{L_w}{\sqrt{b_0 \tau}}\right) - 2 \right] \right. \\ &\quad + \frac{1}{T_w^2} \left[ \int_{2t_L}^{2t_L+T_w} f_1(\alpha, L_w) d\alpha - \int_{2t_L+T_w}^{2(t_L+T_w)} f_1(\alpha, L_w) d\alpha \right. \\ &\quad \left. \left. - 2t_L \int_{2t_L}^{2t_L+T_w} f_0(\alpha, L_w) d\alpha + 2(t_L + T_w) \int_{2t_L+T_w}^{2(t_L+T_w)} f_0(\alpha, L_w) d\alpha \right] \right\}, \end{aligned} \quad (68)$$

where  $t_L$  is the lead time,  $T_w$  is the width of the temporal averaging window,  $L_w$  is the width of the spatial averaging window and

$$\begin{aligned} f_0(\alpha, \beta) &= b_0 \tau \exp\left(-\frac{\beta}{\sqrt{b_0 \tau}}\right) \text{erf}\left(\frac{\beta}{2\sqrt{b_0 \alpha}} - \sqrt{\alpha/\tau}\right) \\ &\quad - b_0 \tau \exp\left(\frac{\beta}{\sqrt{b_0 \tau}}\right) \text{erf}\left(\frac{\beta}{2\sqrt{b_0 \alpha}} + \sqrt{\alpha/\tau}\right) \\ &\quad + 2b_0 \tau \text{erf}(\sqrt{\alpha/\tau}) + 2\sqrt{b_0 \tau} \beta \exp(-\alpha/\tau) \text{erf}\left(\frac{\beta}{2\sqrt{b_0 \alpha}}\right) \\ &\quad - 4b_0 \sqrt{\frac{\tau \alpha}{\pi}} \exp(-\alpha/\tau) \left(1 - \exp\left(-\frac{\beta^2}{4b_0 \alpha}\right)\right), \end{aligned} \quad (69)$$

$$f_1(\alpha, \beta) = \alpha f_0(\alpha, \beta). \quad (70)$$

## C Calculations of some important statistics

By the change of the variable in (33), inserting into the SPDE (8) gives

$$\frac{\partial Q}{\partial t} = b_0 \Delta Q - \frac{1}{\tau} Q + D_* \dot{W}. \quad (71)$$

With one more step of changing the variable by  $v(t, x) = e^{t/\tau} Q(t, x)$ , the equation (71) is simplified to a simple and regular stochastic heat equation

$$\frac{\partial v}{\partial t} = b_0 \Delta v + D_* e^{t/\tau} \dot{W}. \quad (72)$$

$v$  then can be solved directly as

$$v(t, x) = \frac{1}{(4\pi b_0 t)^{1/2}} \int_{\mathbb{R}} e^{-\frac{|x-y|^2}{4b_0 t}} v(0, y) dy + \int_0^t \frac{1}{4\pi b |t-s|^{1/2}} \int_{\mathbb{R}} e^{-\frac{|x-y|^2}{4b(t-s)}} e^{s/\tau} \dot{W}(s, y) dy ds. \quad (73)$$

The important statistics of the  $q(t, x)$  and  $Q(t, x)$  can be calculated through this analytic solution (73) as the following.

- **Mean**

For a bounded initial value  $|q(0, y)| \leq K$  ( $K$  is a finite constant)

$$\begin{aligned} |\mathbb{E}[q(t, x)] - (q^* + \tau F)| &= |\mathbb{E}[Q(t, x)]| \\ &= |\mathbb{E}[e^{-t/\tau} v(t, x)]| \end{aligned}$$

$$\begin{aligned}
&= \left| e^{-t/\tau} \frac{1}{(4\pi b_0 t)^{1/2}} \int_{\mathbb{R}} e^{-\frac{|x-y|^2}{4b_0 t}} v(0, y) dy \right| \\
&= \left| e^{-t/\tau} \frac{1}{(4\pi b_0 t)^{1/2}} \int_{\mathbb{R}} e^{-\frac{|x-y|^2}{4b_0 t}} (q(0, y) - q^* - \tau F) dy \right| \\
&\leq e^{-t/\tau} \frac{1}{(4\pi b_0 t)^{1/2}} \int_{\mathbb{R}} e^{-\frac{|x-y|^2}{4b_0 t}} (K + q^* + \tau F) dy \\
&= (K + q^* + \tau F) e^{-t/\tau} \rightarrow 0 \quad \text{as } t \rightarrow \infty \\
&\Rightarrow \lim_{t \rightarrow \infty} \mathbb{E}[q(t, x)] = q^* + \tau F. \tag{74}
\end{aligned}$$

• **Covariance and variance**

$$\begin{aligned}
\text{cov}[q(s, x), q(t, y)] &= \text{cov}[Q(s, x), Q(t, y)] \\
&= \text{cov}[e^{-s/\tau} v(s, x), e^{-t/\tau} v(t, y)] \\
&= e^{-(s+t)/\tau} \text{cov}[v(s, 0), v(t, x-y)] \\
&= e^{-(s+t)/\tau} \frac{D_*^2}{4\pi b_0} \\
&\quad \mathbb{E} \left[ \int_0^t \int_0^s \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{1}{|t-r|^{1/2} |s-r'|^{1/2}} e^{-\frac{|x-y|^2}{4b_0(t-r)} - \frac{|y'|^2}{4b_0(s-r')}} \dot{W}(r, y) \dot{W}(r', y') e^{r/\tau} e^{r'/\tau} dy dy' dr' dr \right] \\
&= e^{-(s+t)/\tau} \frac{D_*^2}{4\pi b_0} \int_0^{s \wedge t} \int_{\mathbb{R}} \frac{1}{|t-r|^{1/2} |s-r|^{1/2}} e^{-\frac{|x-y|^2}{4b_0(t-r)} - \frac{|y|^2}{4b_0(s-r)}} e^{2r/\tau} dy dr \\
&= e^{-(s+t)/\tau} \frac{D_*^2}{(4\pi b_0)^{1/2}} \int_0^{s \wedge t} (s+t-2r)^{-1/2} \exp\left(-\frac{|x-y|^2}{4b_0(s+t-2r)}\right) e^{2r/\tau} dr \\
&= \frac{D_*^2}{4\sqrt{\pi b_0}} \int_{|s-t|}^{s+t} l^{-1/2} \exp\left(-\frac{|x-y|^2}{4b_0 l} - al\right) dl. \tag{75}
\end{aligned}$$

For the temporal covariance at a single location, take  $x = y$  in (75),

$$\begin{aligned}
\text{cov}[q(s, x), q(t, x)] &= \frac{D_*^2}{4\sqrt{\pi b_0}} \int_{|s-t|}^{s+t} l^{-1/2} \exp(-l/\tau) dl \\
(\text{let } \eta &= (l/\tau)^{1/2}) = \frac{D_*^2}{4\sqrt{\pi b_0}} \int_{\sqrt{|s-t|/\tau}}^{\sqrt{(s+t)/\tau}} 2\sqrt{\tau} \exp(-\eta^2) d\eta \\
&= \frac{D_*^2}{4} \sqrt{\frac{\tau}{b_0}} \cdot \frac{2}{\sqrt{\pi}} \int_{\sqrt{|s-t|/\tau}}^{\sqrt{(s+t)/\tau}} \exp(-\eta^2) d\eta \\
&= \frac{D_*^2}{4} \sqrt{\frac{\tau}{b_0}} (\text{erf}(\sqrt{(s+t)/\tau}) - \text{erf}(\sqrt{|s-t|/\tau})). \tag{76}
\end{aligned}$$

For the asymptotic spatial covariance at a single time point, take  $s = t$  in (75),

$$\lim_{t \rightarrow \infty} \text{cov}[q(t, x), q(t, y)] = \frac{D_*^2}{4\sqrt{\pi b_0}} \int_0^{\infty} l^{-1/2} \exp\left(-\frac{|x-y|^2}{4b_0 l} - l/\tau\right) dl$$

$$\begin{aligned}
(\text{Let } \eta = \sqrt{l}) &= \frac{D_*^2}{2\sqrt{\pi b_0}} \exp\left(-\frac{1}{\sqrt{b_0\tau}}|x-y|\right) \int_0^\infty \exp\left(-\left(\frac{|x-y|}{2\sqrt{b_0\eta}} - \frac{\eta}{\sqrt{\tau}}\right)^2\right) d\eta \\
&= \frac{D_*^2}{2\sqrt{\pi b_0}} \exp\left(-\sqrt{\frac{1}{b_0\tau}}|x-y|\right) \cdot \frac{\sqrt{\pi\tau}}{2} \\
&= \frac{D_*^2}{4} \sqrt{\frac{\tau}{b_0}} \exp\left(-\frac{1}{\sqrt{b_0\tau}}|x-y|\right). \tag{77}
\end{aligned}$$

For the stationary variance for  $q(t, x)$ , take  $x = y$  in (77),

$$\lim_{t \rightarrow \infty} \text{var}[q(t, x)] = \frac{D_*^2}{4} \sqrt{\frac{\tau}{b_0}}. \tag{78}$$

## References

- [1] R. Alexander, Z. Zhao, E. Székely, and D. Giannakis. Kernel analog forecasting of tropical intraseasonal oscillations. *J. Atmos. Sci.*, 74(4):1321–1342, 2017.
- [2] F. Atger. Verification of intense precipitation forecasts from single models and ensemble prediction systems. *Nonlinear Processes in Geophysics*, 8(6):401–417, 2001.
- [3] A. G. Barnston, M. H. Glantz, and Y. He. Predictive skill of statistical and dynamical climate models in SST forecasts during the 1997–98 El Niño episode and the 1998 La Niña onset. *Bull. Amer. Meteorol. Soc.*, 80(2):217–244, 1999.
- [4] B. Casati, G. Ross, and D. B. Stephenson. A new intensity-scale approach for the verification of spatial precipitation forecasts. *Meteorological Applications*, 11(2):141–154, 2004.
- [5] N. Chen and A. J. Majda. Predicting the cloud patterns for the Boreal summer intraseasonal oscillation through a low-order stochastic model. *Math. Clim. Weather Forecast.*, 1:1–20, 2016.
- [6] N. Chen, A. J. Majda, and D. Giannakis. Predicting the cloud patterns of the Madden–Julian Oscillation through a low-order nonlinear stochastic model. *Geophys. Res. Lett.*, 41(15):5612–5619, 2014.
- [7] S. Chen, A. J. Majda, and S. N. Stechmann. Multiscale asymptotics for the skeleton of the Madden–Julian oscillation and tropical–extratropical interactions. *Math. Clim. Weather Forecast.*, 1:43–69, 2015.
- [8] S. Chen, A. J. Majda, and S. N. Stechmann. Tropical–extratropical interactions with the MJO skeleton and climatological mean flow. *J. Atmos. Sci.*, 73(10):4101–4116, 2016.
- [9] S. Chen and S. N. Stechmann. Nonlinear traveling waves for the skeleton of the Madden–Julian oscillation. *Comm. Math. Sci.*, 14:571–592, 2016.
- [10] J. Dias, M. Gehne, G. N. Kiladis, N. Sakaeda, P. Bechtold, and T. Haiden. Equatorial waves and the skill of NCEP and ECMWF numerical weather prediction systems. *Mon. Wea. Rev.*, 146(6):1763–1784, 2018.
- [11] Elizabeth E Ebert. Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. *Meteorological applications*, 15(1):51–64, 2008.
- [12] Elizabeth E Ebert. Neighborhood verification: A strategy for rewarding close forecasts. *Weather and Forecasting*, 24(6):1498–1510, 2009.
- [13] S. Hottovy and S. N. Stechmann. A spatiotemporal stochastic model for tropical precipitation and water vapor dynamics. *J. Atmos. Sci.*, 72:4721–4738, 2015.
- [14] M. A. Janiga, C. Schreck, J. A. Ridout, M. Flatau, N. Barton, E. J. Metzger, and C. Reynolds. Subseasonal forecasts of convectively coupled equatorial waves and the MJO: Activity and predictive skill. *Mon. Wea. Rev.*, page in press, 2018.
- [15] B. Khouider and A. J. Majda. A simple multicloud parameterization for convectively coupled tropical waves. Part II: Nonlinear simulations. *J. Atmos. Sci.*, 64:381–400, 2007.
- [16] B. Khouider, A. J. Majda, and S. N. Stechmann. Climate science in the tropics: waves, vortices and PDEs. *Nonlinearity*, 26(1):R1–R68, 2013.
- [17] M. Latif, D. Anderson, T. Barnett, M. Cane, R. Kleeman, A. Leetmaa, J. O'Brien, A. Rosati, and E. Schneider. A review of the predictability and prediction of ENSO. *J. Geophys. Res.: Oceans*, 103(C7):14375–14393, 1998.
- [18] W. K. M. Lau and D. E. Waliser, editors. *Intraseasonal Variability in the Atmosphere–Ocean Climate System*. Springer, Berlin, 2nd edition, 2012.
- [19] R. A. Madden and P. R. Julian. Detection of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci.*, 28(5):702–708, 1971.
- [20] R. A. Madden and P. R. Julian. Description of global-scale circulation cells in the Tropics with a 40–50 day period. *J. Atmos. Sci.*, 29:1109–1123, September 1972.
- [21] A. J. Majda and J. Harlim. *Filtering Turbulent Complex Systems*. Cambridge University Press, 2012.

- [22] A. J. Majda and S. N. Stechmann. The skeleton of tropical intraseasonal oscillations. *Proc. Natl. Acad. Sci. USA*, 106(21):8417–8422, 2009.
- [23] A. J. Majda and S. N. Stechmann. Nonlinear dynamics and regional variations in the MJO skeleton. *J. Atmos. Sci.*, 68:3053–3071, 2011.
- [24] Chiara Marsigli, Andrea Montani, and Tiziana Paccangnella. A spatial verification method applied to the evaluation of high-resolution ensemble forecasts. *Meteorological Applications*, 15(1):125–143, 2008.
- [25] J. D. Neelin, D. S. Battisti, A. C. Hirst, F.-F. Jin, Y. Wakata, T. Yamagata, and S. E. Zebiak. ENSO theory. *J. Geophys. Res.: Oceans*, 103(C7):14261–14290, 1998.
- [26] H. R. Ogrosky, S. N. Stechmann, and S. Hottovy. Instability and nonlinear evolution of the MJO in a model with vertically-varying convective adjustment. *J. Adv. Model. Earth Syst.*, page submitted, 2018.
- [27] C. Penland and T. Magorian. Prediction of Niño 3 sea surface temperatures using linear inverse modeling. *J. Climate*, 6(6):1067–1076, 1993.
- [28] Nigel Roberts. Assessing the spatial and temporal variation in the skill of precipitation forecasts from an nwp model. *Meteorological Applications*, 15(1):163–169, 2008.
- [29] Nigel M Roberts and Humphrey W Lean. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review*, 136(1):78–97, 2008.
- [30] M. J. Rodwell and F. J. Doblas-Reyes. Medium-range, monthly, and seasonal prediction for Europe and the use of forecast information. *J. Climate*, 19(23):6025–6046, 2006.
- [31] S. N. Stechmann and S. Hottovy. Cloud regimes as phase transitions. *Geophys. Res. Lett.*, 43:6579–6587, 2016.
- [32] S. N. Stechmann and S. Hottovy. Unified spectrum of tropical rainfall and waves in a simple stochastic model. *Geophys. Res. Lett.*, 44:10,713–10,724, 2017.
- [33] S. Thual, A. J. Majda, N. Chen, and S. N. Stechmann. Simple stochastic model for El Niño with westerly wind bursts. *Proc. Natl. Acad. Sci.*, 113(37):10245–10250, 2016.
- [34] S. Thual, A. J. Majda, and S. N. Stechmann. A stochastic skeleton model for the MJO. *J. Atmos. Sci.*, 71:697–715, 2014.
- [35] M. Wheeler and G. N. Kiladis. Convectively coupled equatorial waves: analysis of clouds and temperature in the wavenumber–frequency domain. *J. Atmos. Sci.*, 56(3):374–399, 1999.
- [36] M. C. Wheeler, H. Zhu, A. H. Sobel, D. Hudson, and F. Vitart. Seamless precipitation prediction skill comparison between two global models. *Q. J. Roy. Meteorol. Soc.*, 143(702):374–383, 2017.
- [37] Y. Ying and F. Zhang. Practical and intrinsic predictability of multiscale weather and convectively coupled equatorial waves during the active phase of an MJO. *J. Atmos. Sci.*, 74(11):3771–3785, 2017.
- [38] Y. Ying and F. Zhang. Potentials in improving predictability of multiscale tropical weather systems evaluated through ensemble assimilation of simulated satellite-based observations. *J. Atmos. Sci.*, 75(5):1675–1698, 2018.
- [39] C. Zhang. Madden–Julian Oscillation. *Rev. Geophys.*, 43:RG2003, June 2005.
- [40] H. Zhu, M. C. Wheeler, A. H. Sobel, and D. Hudson. Seamless precipitation prediction skill in the tropics and extratropics from a global model. *Mon. Wea. Rev.*, 142(4):1556–1569, 2014.
- [41] H. Zhu and M. C. Wheeler. Seamless precipitation prediction skill in the tropics and extratropics from a global model. *Monthly Weather Review*, 142:1556–1569, 2014.