A multi-model study of atmosphere predictability in coupled ocean–atmosphere systems

Youwei Ma[1] · Jianping Li[1,2,3] · Shaoqing Zhang[1,2,3,4] · Haoran Zhao[1]

Abstract

Of great importance for guiding numerical weather and climate predictions, understanding predictability of the atmosphere in the ocean − atmosphere coupled system is the first and critical step to understand predictability of the Earth system. However, previous predictability studies based on prefect model assumption usually depend on a certain model. Here we apply the predictability study with the Nonlinear Local Lyapunov Exponent and Attractor Radius to the products of multiple re-analyses and forecast models in several operational centers to realize general predictability of the atmosphere in the Earth system. We first investigated the predictability characteristics of the atmosphere in NCEP, ECMWF and UKMO coupled systems and some of their uncoupled counterparts and other uncoupled systems. Although the ECMWF Integrated Forecast System shows higher skills in geopotential height over the tropics, there is no certain model providing the most precise forecast for all variables on all levels and the multi-model ensemble not always outperforms a single model. Improved low-frequency signals from the air − sea and stratosphere − troposphere interactions that extend predictability of the atmosphere in coupled system suggests the significance of air − sea coupling and stratosphere simulation in practical forecast development, although uncertainties exist in the model representation for physical processes in air − sea interactions and upper troposphere. These inspire further exploration on predictability of ocean and stratosphere as well as sea − ice and land processes to advance our understanding of interactions of Earth system components, thus enhancing weather − climate prediction skills.

Keywords Atmosphere predictability · Coupled ocean − atmosphere system · Multi-model study · Nonlinear local lyapunov exponent · Attractor radius

1 Introduction

The predictability study is a continuous and important field in predictions of the atmosphere and ocean as dynamical stochastic systems. Since the results of Lorenz (1963, 1965) and Thompson (1957) were published, one has realized that chaotic systems are highly sensitive to a small initial error that can grow nonlinearly during the forecast, leading a so-called predictability limit, by which the initial signals could not be identified and making the prediction meaningless. The upper limit of predicting reliable high-impact weather events has been estimated as $1 − 2$ weeks (Lorenz 1969). Instead of this general estimate for the global atmosphere, the predictability limit varies with geographic locations and has also become an assessment indictor of numerical prediction models (Mu et al. 2017; Duan and Zhao 2015).

Previous work employed various methods to estimate the predictability limits of the atmosphere with aids of observations and numerical models. Lorenz (1969) proposed a

✉ Jianping Li
ljp@ouc.edu.cn

✉ Shaoqing Zhang
szhang@ouc.edu.cn

[1] The College of Oceanic and Atmospheric Sciences, Ocean University of China, Qingdao, China

[2] Key Laboratory of Physical Oceanography, Institute for Advanced Ocean Study, Frontiers Science Center for Deep Ocean Multispheres and Earth System (FDOMES), Ocean University of China, Qingdao, China

[3] Ocean Dynamics and Climate Function Lab, Pilot National Laboratory for Marine Science and Technology (QNLM), Qingdao, China

[4] International Laboratory for High-Resolution Earth System Model and Prediction (iHESP), Qingdao, China

concept called nature-analog over a large area (e.g., the north hemisphere) to estimate atmosphere predictability while it requires a long time series of $10-100$ year data (Van den Dool 1994). Oseledec (1968) used the Lyapunov exponent to measure the average exponential rates of divergence or convergence of nearby orbits on a strange attractor. Then the largest Lyapunov exponent and various local or finite-time Lyapunov exponents were used to measure the average predictability limit (Shimada and Nagashima 1979; Benettin et al. 1980; Sano and Sawada 1985; Wolf et al. 1985; Yoden and Normura 1993; Kazantsev 1999; Ziehmann et al. 2000). These methods only can be applied to circumstances of linear error growing, while the nonlinear local Lyapunov exponent (NLLE) was proposed to measure average growth of initial error taking account of the nonlinear behavior in chaotic systems (Li et al. 2006; Ding and Li 2007; Ding and Li 2007; Li and Ding 2009; Li and Wang 2008). Ding and Li (2007) demonstrated the superiority of the nonlinear Lyapunov exponent in determining the predictability limits of chaotic systems in comparison with linear counterpart: the linear and nonlinear average growth of errors are similar during short initial time while the nonlinear error growth departs from the linear one with the increasing time and finally saturates instead of constant exponential growth of linear counterpart. Furthermore, regardless the various magnitudes of initial errors, the predictability limits estimated by nonlinear Lyapunov exponent are longer than ones estimated by linear counterparts. With time series of observational data, Li and Ding (2011) derived an algorithm to calculate the predictability limit for the atmospheric chaotic system without its governing equation. On the other hand, the predictability limit of a forecast model is defined as 95% of the saturated root mean square error (RMSE) between forecasts and observations (Dalcher and Kalnay 1987; Simmons and Hollingsworth 2002; Buizza 2010), but the error growth is largely influenced by model deficiencies (Orrell et al. 2001). To overcome this defect, Li et al. (2018) proposed a unified theory about the global and local predictability limit of a chaotic system consisting of a forecast model that includes three fundamental intrinsic properties: attractor radius (AR), global attractor radius (GAR), and global average distance (GAD).

Unlike using a single model and under a prefect model assumption (Rowell 1998; Kumar et al.2003; Reichler and Roads 2004), here we use multiple models from several operational centers including the Climate Forecast System version 2 (CFSv2) from the National Centers for Environmental Prediction (NCEP) (Saha et al. 2010), Integrated Forecasting System (IFS) cycle 43R1 and 43R3 from the European Centre for Medium-Range Weather Forecasts (ECMWF) (Vitart 2014) and HadGEM3 in GloSea5 from the United Kingdom Met Office (UKMO) (MacLachlan et al. 2015). We also use the NCEP CFSR and ECMWF

Interim reanalysis products to represent different realizations of the "real world". The motivation of our study is to explore the estimated predictability limit distribution of "real world" and forecast models then generalize predictability limits of coupled models to mitigate the dependency of the individual prediction models.

This paper is organized as follows. After the introduction, Sect. 2 briefly describes the methods of the NLLE and AR as well as the data used throughout this study. Section 3 presents the predictability limits obtained by NLLE and AR characteristics in multiple coupled reanalysis datasets. The predictability characteristics in multiple coupled prediction systems are provided in Sect. 4. Finally, summary and discussions are given in Sect. 5.

2 Methodology

2.1 Nonlinear Local Lyapunov exponent and attractor radius

2.1.1 Nonlinear local Lyapunov exponent (NLLE)

To estimate the predictability limits, nonlinear patterns in the behavior of error growth need to be considered (Lacarra and Talagrand 1988; Mu 2000). The NLLE, a new method to estimate the average rate of initial error growth based on a nonlinear dynamical model without linearizing the governing equations, is recently proposed and applied to investigate the predictability limits for an n-dimensional chaotic system or a single variable (Li et al. 2006; Ding and Li 2007; Ding and Li 2007; Li and Ding 2009; Li and Wang 2008). For an n-dimensional chaotic system, the nonlinear equation of error growth could be integrated along the reference solution x from $t = t_0$ to $t_0 + \tau$,

$$\delta_1 = \eta(x_0, \delta_0, \tau)\delta_0, \tag{1}$$

where $\delta_1 = \delta(t_0 + \tau)$, $x_0 = x(t_0)$, $\delta_0 = \delta(t_0)$, and $\eta(x_0, \delta_0, \tau)$ is the nonlinear propagator. Then the NLLE is defined as:

$$\lambda(x_0, \delta_0, \tau) = \frac{1}{\tau} ln \frac{\|\delta_1\|}{\|\delta_0\|}, \tag{2}$$

where $\lambda(x_0, \delta_0, \tau)$ is the function of initial state x_0 in phase space, the initial error δ_0 and time τ. the NLLE of a single variable of the system is defined as:

$$\xi_i(x_0, \delta_0, \tau) = \frac{1}{\tau} ln \frac{|\delta_i(t_0 + \tau)|}{|\delta_i(t_0)|} \tag{3}$$

The mean NLLE $\overline{\xi_i}$ and relative growth of the initial error (RGIE) $\overline{\Phi_i}$ of the variables $\overline{x_i}$ can be obtained by:

$$\overline{\xi_i}(\delta_0, \tau) = \langle \xi_i(x_0, \delta_0, \tau) \rangle_N, (N \to \infty) \tag{4}$$

$$\overline{\Phi_i}(\delta_0, \tau) = exp\left[\overline{\xi_i}(\delta_0, \tau)\tau\right] \tag{5}$$

where $\langle \rangle_N$ denotes the ensemble average of samples with a large size.

To apply the NLLE to the study of atmospheric predictability, Li and Ding (2011) proposed a new algorithm searching for local dynamic analogs (LDA) using observational data. They obtained initial distance between the reference point and other points and corresponding evolutionary distance of two trajectories of these two points over a short time. Mathematically, initial distance d_i between the reference point $x(t_0)$ and other points in the time series $x(t_j)$ except temporal nearby points is given by

$$d_i = \left| x(t_0) - x(t_j) \right| \tag{6}$$

and the evolutionary distance d_e within a short period $\tau = K\Delta$, Δ is the sampling interval of time series, between $x(t_0)$ and $x(t_j)$ is given by

$$d_e = \sqrt{\frac{1}{K+1} \sum_{i=0}^{K} [x(t_i) - x(t_{j+i})]^2}. \tag{7}$$

Then we can base on the minimum of total distance $d_t = d_i + d_e$ to find out the local analog $x(t_k)$ (i.e. the nearest initial state in phase space) so the initial distance between $x(t_0)$ and $x(t_k)$ can be written as:

$$L(t_0) = \left| x(t_0) - x(t_k) \right|, \tag{8}$$

when $x(t_0)$ and $x(t_k)$ travel along the reference trajectory and analogous trajectory, the distance become

$$L(t_i) = \left| x(t_i) - x(t_{k+i}) \right|, \tag{9}$$

so, the NLLE during the time interval $t_i - t_0$ could be calculated as

$$\xi_1(t_i) = \frac{1}{t_i - t_0} ln\frac{L(t_i)}{L(t_0)}, \tag{10}$$

and $i = 1, 2, 3, \ldots, M$ where M is the total number of time series.

Changing the reference points along the observation time series we can construct a series of NLLE and the corresponding ensemble mean NLLE, as well as the approximation of the RGIE. We can estimate the predictability limits of variable x by investigating the time when RGIE reaches 98% of its saturation level. Reaching a saturation value of RGIE in a chaotic system represents the mean distance between two random points on an attractor so

that the initial value information is totally untraceable (Li and Ding 2015).

2.1.2 Optimal local dynamic analog (OLDA)

The LDA estimated by the above algorithm is the most analogous state against the reference state based on available observation so we call it as an estimated OLDA ($\widetilde{OLDA}$), which depends on the length of observation time series. With more observations, theoretically, there exists an **OLDA**:

$$OLDA = \lim_{N \to \infty} \widetilde{OLDA}(N), \tag{11}$$

where N is the length of observational data. The estimated initial error between the reference trajectory and $\widetilde{OLDA}$ is defined as $\widetilde{\delta_0}(N)$ while the optimal initial error between the reference trajectory and $OLDA$ is represented as δ_0 i.e.

$$\delta_0 = \lim_{N \to \infty} \widetilde{\delta_0}(N) \text{ and } \widetilde{\delta_0}(N_1) \leq \widetilde{\delta_0}(N_2), N_1 > N_2 \tag{12}$$

There are two types of predictability limits, estimated predictability limit $\widetilde{T_P}$ and optimal predictability limit T_P determined by the $\widetilde{OLDA}$ and $OLDA$ respectively, i.e.

$$T_P = \lim_{N \to \infty} \widetilde{T_P}(N) \text{ and } \widetilde{T_P}(N_1) \geq \widetilde{T_P}(N_2), N_1 > N_2 \tag{13}$$

With extended observational data, the attractor is gradually filled so it is more possible for NLLE method to find the $OLDA$. Furthermore, because the average divergence rate of $OLDA$ is lower than that of $\widetilde{OLDA}$, the $\widetilde{T_P}$ is lower than that T_P. To verify this point, we first run Lorenz63 model for 10^7 time steps using a fourth-order Runge–Kutta time forwarding scheme ($\Delta t = 0.01$) and the model is spun up for first 10^6 time steps starting from $(x, y, z) = (0, 1, 0)$. We further apply the NLLE method with different lengths of time series to estimate the corresponding initial error and predictability limits, shown in Fig. 1. When data are more abundant, the initial error decreases and maintains a small value, indicating that $\widetilde{OLDA}$ gradually approach to $OLDA$ with increasing predictability limits.

2.1.3 Attractor radius

Besides, Li et al. (2018) proposed three intrinsic properties of chaotic system called the AR, GAR and GAD between two attractors as more objective metrics. Consider x to be the state column vector on a compact attractor Θ, the expectation of the root mean square distance between all states on an attractor and the center of the attractor of an n-sphere is defined as the AR (R_E) as:

$$R_E = \sqrt{E(\|x - E(x)\|^2)}, x \in \Theta \tag{14}$$

Fig. 1 Estimated predictability limits and ensemble average of initial error between analogous trajectories and reference trajectories vary with length of time series from 1×10^4 to 5×10^6 in Lorenz63 model obtained by the LDA algorithm
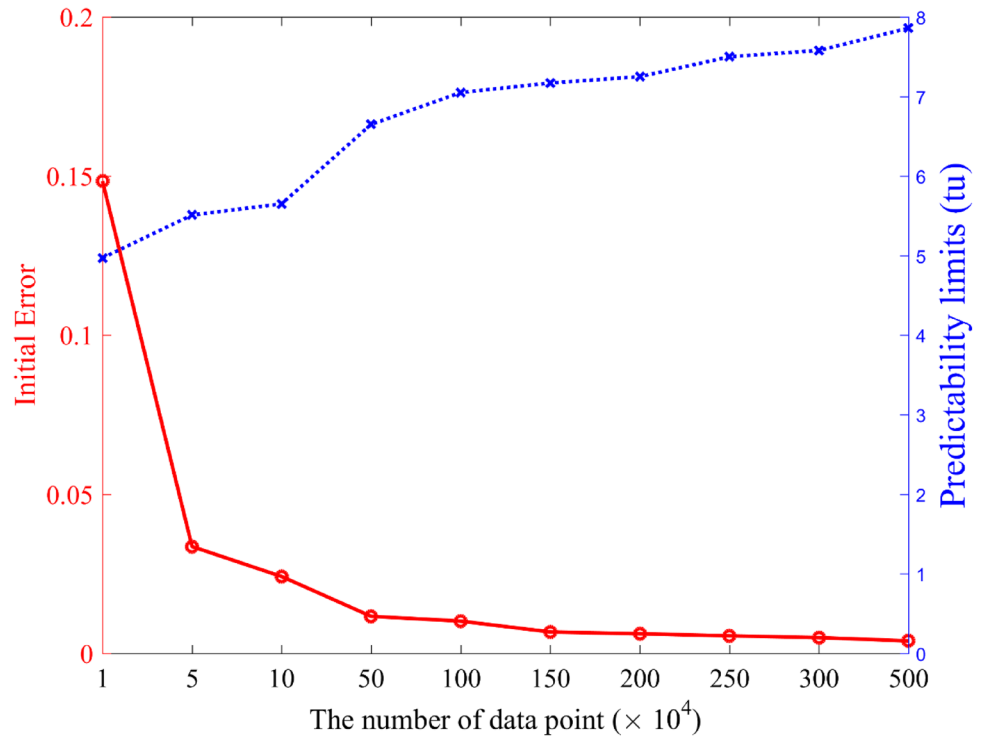
where E is the expectation of the time series and $\|\cdot\|$ is the L2-norm of a vector. AR has a similar form as the standard deviation in statistics. The GAR is defined as the expectation of the root mean square distance between any two random points on an attractor:

$$R_G = \sqrt{E(R_L^2)} = \sqrt{E(\|x - y\|^2)}, x, y \in \Theta \qquad (15)$$

The GAD between two attractors, Θ_1 and Θ_2 is written as:

$$R_G(\Theta_1, \Theta_2) = \sqrt{R_E^2(\Theta_1) + R_E^2(\Theta_2) + d^2(x_E, y_E)}. \qquad (16)$$

A constant relationship between the GAR and AR exists as

$$R_G = \sqrt{2}R_E \qquad (17)$$

simplifying the calculation of GAR. In terms of forecast model, because of the model error, ARs of models and reanalysis are different and the GAD (i.e. the saturation of model RMSE) is not same as the GAR. GAD is not suitable to be the threshold to measure predictability because it contains the different model errors in different models according to its definition. We let $R_e = min(R_E(\Theta_1), R_E(\Theta_2)), R_g = min(R_G(\Theta_1), R_G(\Theta_2))$. The practical (potential) predictability limit can be described as the prediction ability based on the current available (optimum) procedures. Therefore, the global practical predictability limit (global potential predictability limit) is defined

here as the time when the global ensemble average of RMSEs reaches the AR (GAR) for the first time.

2.2 Data, models and forecast systems

Reforecast data of coupled models used in this paper are all archived in S2S prediction experiment dataset (Vitart et al. 2017). We only use 00Z control forecast data to calculate the errors. The Climate Forecast System Version 2 (CFSv2) (Saha et al. 2014) from National Centers for Environmental Prediction (NCEP) of runs out to 45 days initialized for 4 times (00, 06, 12 and 18 UTC) per day over the 11-year (1999 to 2010) based on Climate Forecast System Reanalysis (CFSR), and 11-member ensemble reforecasts of ECMWF Integrated Forecast System (IFS) ensemble prediction system. IFS makes forecasts twice a week (Monday and Thursday 00Z) up to 46 days produced on the fly for last 20 years and the reforecast atmosphere component initial condition is Interim (T255L60). Details of ECMWF-IFS could be found at https://confluence.ecmwf.int/display/S2S/ECMWF+Model + Description + CY43R3. The UK Met Office (UKMO) HadGEM3 reforecast (Williams et al. 2015) is produced on the fly and initialized from Interim (atmosphere component) 4 times a month on 1st, 9th, 17th and 25th extending to 60 days for 23 years (1993 − 2015) with 7 ensembles per year. In Sect. 4.1, we introduce an uncoupled ensemble including GEPS reforecast from Environment and Climate Change Canada (ECCC) produced on the fly once a week up to 32 days with 4 members initialized from Interim,

ISAC-CNR model reforecast up to 31 days runs every 5 days from The Institute of Atmospheric Sciences and Climate (CNR-ISAC) and GEFS Reforecast Version 2 consists of an 11-member ensemble of forecasts, produced every day from 00 UTC. Only control run of the ensemble model above is employed. The NCEP climate forecast system reanalysis (CFSR) and ERA-Interim are approximately equivalent to the observed atmosphere state to calculate predictability limits of atmosphere using NLLE method and to verify the forecasts using AR method. Both CFSR and Interim run 4 times a day (00, 06, 12 and 18 UTC) for 39 years $(1979-2017)$ and all the data at February 29th are eliminated. The annual cycle of time series is removed when apply to calculate the AR and GAR. Details of the reforecast are summarized in Table 1.

## 3 The characteristics of NLLE and AR in multiple coupled reanalysis products

The predictability limits of the real atmosphere obtained by the NLLE method represent the saturated "error" due to intrinsic variability in an observational dataset. The AR is the metric measuring the predictability limits of a forecast model. This section focuses on analyzing the distribution of the predictability limits of the atmosphere estimated by the NLLE method. We also analyze the AR of CFSR and Interim to get the sense of the different reflections of atmosphere in different reanalysis datasets.

CFSR has 360×181 grid points, and there are $39a \times 365$(days/a)$\times 4$(times/day) time reference points available at each grid points (following Li and Ding 2011). Since we only search the local dynamic analog in a season about ± 45 days, totally 4(times/day)$\times 91$(days)$\times 38a$ data points are available for analog searching. Except for a little lower spatial resolution, the Interim data are same as the CFSR in terms of number of reference points and points available

for analog searching. Figure 2 shows the spatial distribution of predictability limits of geopotential height (GHT) by the NLLE method. At 200 hPa level, the predictability limits of the CFSR and Interim have nearly identical zonal patterns that the high values of $13-16$ days appear over the tropics, Arctic and Antarctic, followed by 8–12 days over mid-high latitudes in the Northern Hemisphere and the lowest predictability limits (4–6 days) are distributed over middle latitudes in the South Hemisphere. Reichler and Roads (2004) had a similar conclusion that the tropics and Antarctic are main regions with high predictability for which the dominant sources are both boundary conditions (e.g., signal coming from ocean and sea-ice components) and initial conditions. The high-latitude blocking and high-frequency baroclinic wave activity (Dalcher and Kalnay 1987; Renwick and Wallace 1996) may contribute to the lower predictability limits over mid latitudes. Compared to 200 hPa, the predictability limits on 500 and 850 hPa have the similar distribution but with lower value. The predictability limits in global domain have a barotropic-like structure which decreases progressively from the upper troposphere to the lower troposphere. The predictability limit is approximately the function of pressure just like the isobaric surface varies parallel to the isopycnic surface and the density is the function of pressure in barotropic structure. There are thought-provoking parallels between our results and previous researches, e.g., Li and Ding (2011). A possible comprehension is that above the planetary boundary layer, synoptic-scale flow of the free atmosphere not directly retarded by surface friction may contribute to the higher predictability limit at upper levels. Additionally, stratosphere or stratosphere-troposphere coupling may be a source of predictability at upper troposphere (Baldwin and Dunerton 2001; Baldwin et al. 2003; Kuroda, 2008; Douville 2009; Li and Ding 2011; Hitchcock and Simpson 2014; Lim et al. 2019). It is noted that the zonal mean of GHT predictability limits over mid latitudes in the Southern Hemisphere is lower than that in the North

Table 1 Reforecast data needed in this research and their properties

	NCEP-CFSv2	ECMWF	UKMO- GloSea5
Model	Coupled models		
Reforecast	Fixed	On the fly	On the fly
variables	GHT, temperature, u, v components of wind at 200, 500 and 850 hPa		
Time period	1999.1–2010.12	1996.1–2017.6	1993.1–2015.12
Run frequency	4 times a day at 00, 06, 12, 18 UTC	Twice a week at Monday and Thursday	4 times a month on 1st, 9th, 17th, 25th
Selected initial time	00Z		
Forecast length	45 days	46 days	60 days
Initial condition	CFSR	Interim	Interim
Spatial resolution of output data	$1° \times 1°$	$1.5° \times 1.5°$	$1.5° \times 1.5°$
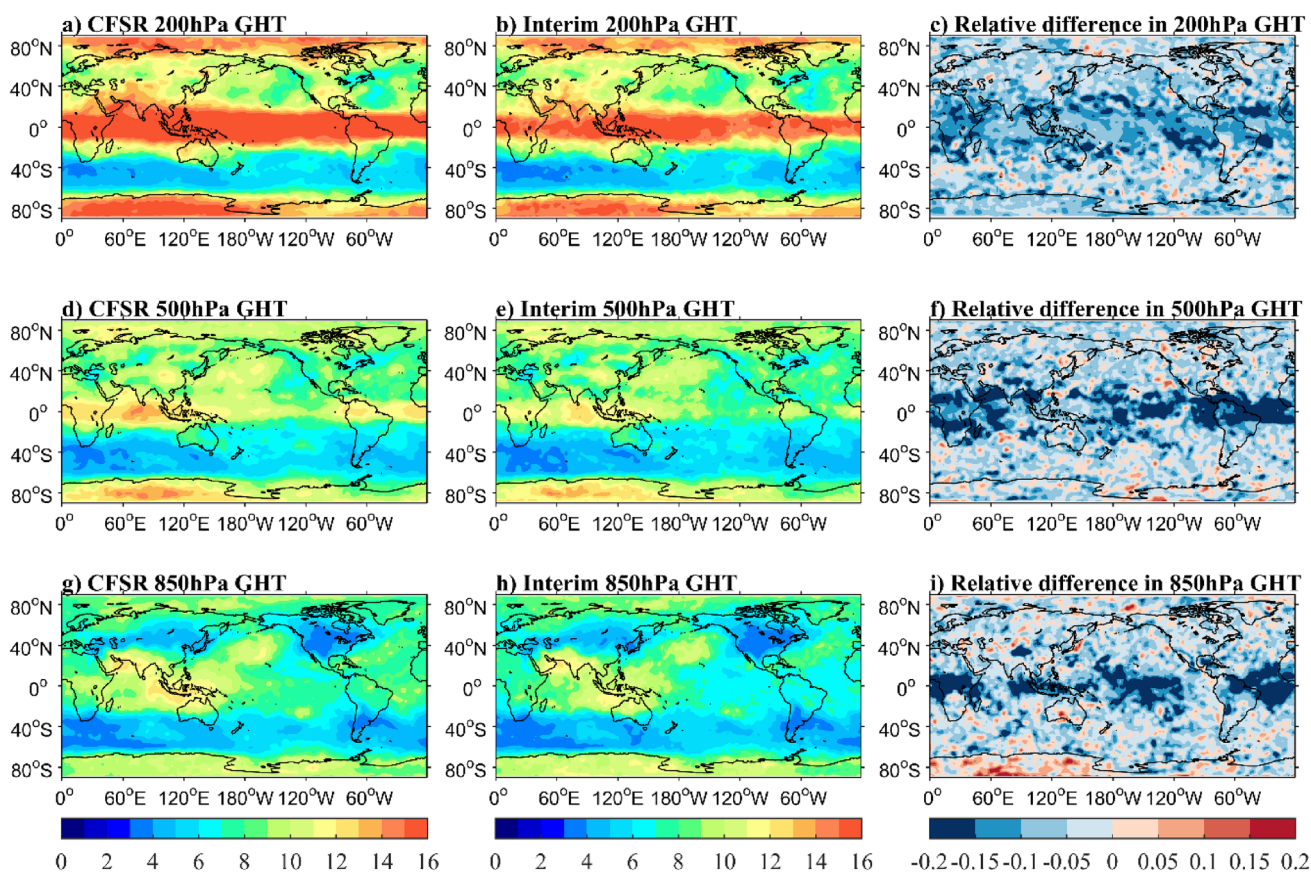
Fig. 2 Spatial distributions of predictability limits (in days) using GHT in CFSR (left panels) and Interim (middle panels) by NLLE method and the corresponding relative difference (right panels) of predictability limits between two reanalysis datasets. Upper, middle, lower panels are at 200, 500 and 850 hPa, respectively

Hemisphere at 3 levels (not shown). There might be some dynamical mechanisms in Southern Hemisphere responsible for the GHT predictability and they need further research. Meanwhile, the misrepresentation of reanalysis due to scarce observations cannot be overlooked. The relative difference of the GHT predictability limits between CFSR and Interim (right panels of Fig. 2) is quite small in most of the regions globally. The largest zonal mean of relative differences is around 10% over the tropics. We see that potential predictability limits of Interim are lower than ones of CFSR over tropical areas on different levels indicating that the error between reference trajectory and analogous trajectory in Interim diverges faster than in the CFSR.

Predictability limits of temperature in both CFSR and Interim show peaks over tropics and polar areas while valleys over mid latitudes (not shown). Specifically, they have high values above 10, 8 and 8 days over both the tropics and polar regions followed by 3−9, 3−7 and 3−7 days over mid latitudes at 200, 500 and 850 hPa, respectively. The predictability limit of wind speed shows a different zonal mean distribution with high values above 8, 6 and 5 days at the Northern Hemisphere subtropics followed by 4–7, 3–5

and 2–5 days over both the polar regions and mid latitudes at 200, 500 and 850 hPa, respectively. There still exist the barotropic-like structures in predictability limits of both temperature and wind speed. The zonal mean predictability limits of temperature and wind speed are lower than those of GHT. Furthermore, those of wind speed is the lowest among these 3 variables. The predictability limit difference of temperature and wind speed between two reanalysis datasets still exist on 3 levels as illustrated in Fig. 3. The predictability limit relative difference of temperature also shows a zonal distribution that at 200 hPa, it exceeds 15% over the tropics while is lower than 10% over other areas; at 500 hPa, it is approximately below 10% globally; at 850 hPa, the higher value is located in the tropics and Antarctica. Specifically, there is a lower potential predictability limit area for temperature in Interim over the tropics at 200 and 850 hPa while a higher limit area over western tropical Pacific at 500 hPa and over Antarctica at 850 hPa. As shown in the right panels of Fig. 3, zonal mean of wind speed predictability limits shows a similar distribution to that of temperature but over tropical areas Interim has higher potential predictability limits of wind speed than CFSR. The relative difference

Fig. 3 Difference (Interim minus CFSR) in temperature (left panels) and wind speed (right panels) predictability limits (in days) between CFSR and Interim. Upper, middle, lower panels are at 200, 500 and 850 hPa, respectively
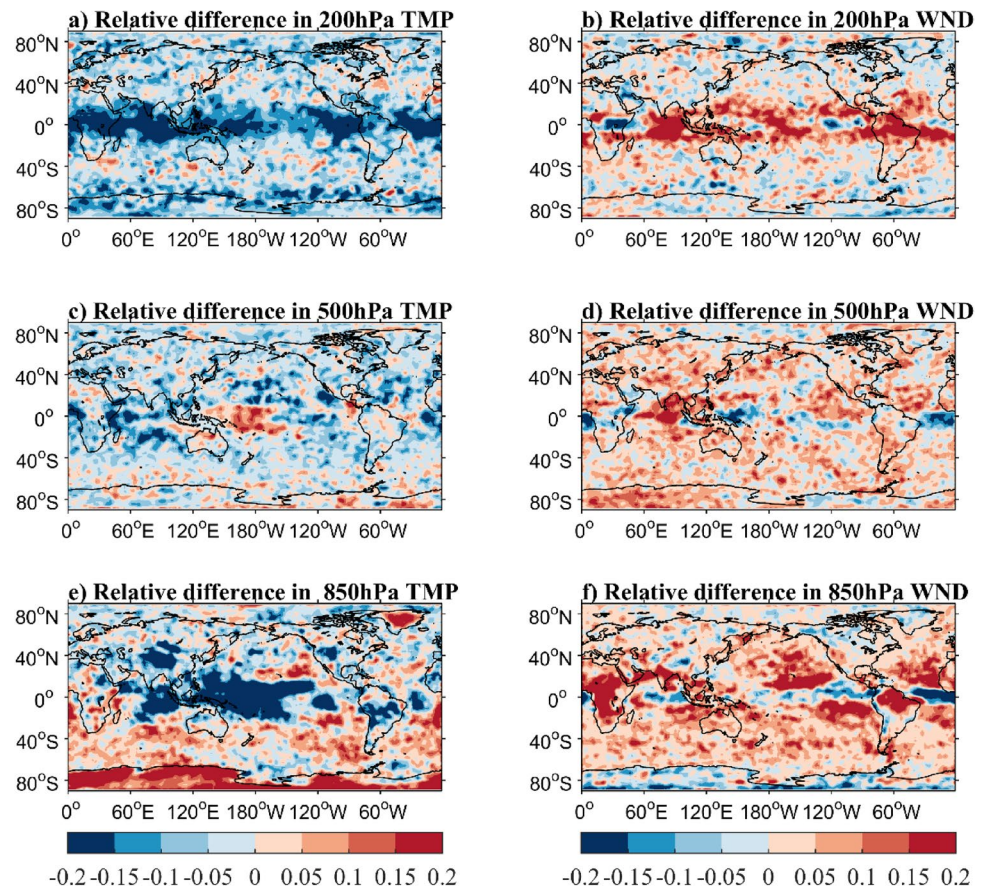
of predictability limits of wind speed between Interim and CFSR is little larger than those of GHT over the tropics at 200 and 850 hPa, but in other areas at 3 levels they may not have obvious differences. It is seen that although the differences in potential predictability between Interim and CFSR exist in all variables on 3 levels over various areas, these two sets of reanalysis data demonstrate generally similar distribution of potential predictability limits.

Figure 4 illustrates the spatial distribution of GHT AR from CFSR and Interim and differences between them on various pressure levels. The 200 hPa GHT AR in trough and ridge systems and Southern Hemisphere westerlies is much higher than that in the tropics. Furthermore, the AR of both CFSR and Interim exists a relatively high value center in the central-east Pacific compared to other tropical regions. This high value center may refer to the westerly duct, theorized by Webster and Holton (1982), that mid-latitude disturbance propagates into the tropics through westerly duct located in the upper-tropospheric eastern Pacific. The GHT ARs at 500 and 850 hPa have the similar zonal distribution patterns but lower values than 200 hPa. The ARs in Interim are lower than those in CFSR over most areas while the differences of ARs between CFSR and Interim are small.

For temperature and wind speed, the spatial distribution of the differences of ARs of CFSR and Interim on 3 pressure levels are shown in Fig. 5. The differences in temperature at 200 and 500 hPa are overall small but temperature in Interim varies more significantly over the westerlies and high-altitude areas than CFSR. Likewise, the wind speed ARs in Interim are a little larger in the westerly and polar regions than CFSR.

Both the ECMWF-IFS and HadGEM3 use ERA-Interim states as the initial conditions while the CFSv2 predictions use CFSR states as initial conditions. Analyses above show that no matter the differences of predictability limits between the CFSR and Interim calculated by the NLLE or AR itself are relatively small for GHT, temperature and wind speed, suggesting that these 2 sets of reanalysis data can describe the roughly similar chaotic properties of the atmosphere. We should note that although Interim are only coupled with ocean–wave model (Dee et al. 2011) while CFSR has the ocean and sea–ice components, the sensitivity of predictability limits to different reanalysis datasets can be roughly omitted as the results shown above. It can be comprehended that the sources of ocean signal for reanalysis are both ocean model and observation. Thus, the ocean signal from observation may make up the shortcoming of Interim as an uncoupled reanalysis. Additionally, the AR values calculated from CFSR and Interim reanalysis data are basically identical, so we will use the AR of
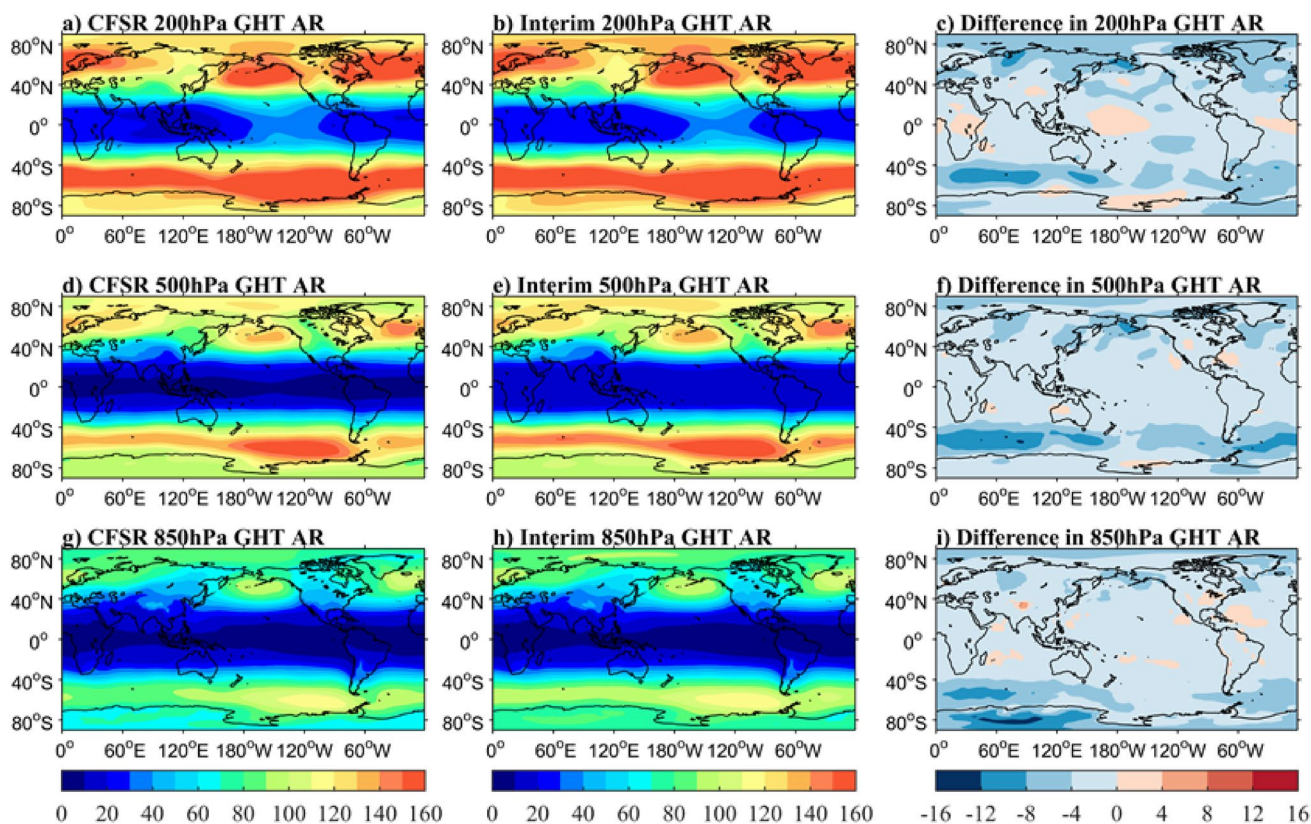
Fig. 4 The spatial distribution of AR of GHT obtained from CFSR (left panels) and Interim (middle panels) and the difference between two reanalysis datasets (right panels). The upper, middle and lower panels are at 200, 500 and 850 hPa, respectively. Unit: m

two reanalyses as unified metrics to measure the model predictability limits next.

4 Predictability characteristics in multiple coupled prediction systems

4.1 The characteristic predictability of coupled prediction systems

Based on the characteristics of atmosphere potential predictability estimated through NLLE above, we now discuss the predictability of coupled forecast models from ECMWF, NCEP and UKMO which are world well-known operational forecast centers.

Figure 6 shows the practical predictability limits of GHT, temperature and wind speed from ECMWF-IFS. The GHT (left panels) has longer predictability limits than other variables on all pressure levels. In the upper troposphere (200 hPa), predictability limits are with the maximum in the most of the tropical belt around 30 days. While predictability limits at 500 and 850 hPa are divided into two parts, Indo–Pacific warm pool and tropical Atlantic, where the predictability limits exceed 14 days. We will discuss more

on this phenomenon to understand these conspicuous high values.

It is noticed that compared to Fig. 2, the predictability limits of ECWMF-IFS over the tropics are surprisingly longer than the upper limits calculated by the NLLE using observation. Theoretically, potential predictability limits of "real" atmosphere calculated by NLLE from reanalysis data should be longer than practical predictability limits of models (Li and Ding 2015) due to model error and initial error. It is understandable that the real atmosphere possesses the potential predictability limits due to its chaotic characteristic regardless of the initial error. Despite we assumed that there is a perfect model, the predictability limit of this model is, at best, equal to the potential predictability limit. However, the model is actually an imperfect representation of the real atmosphere, its states will depart from the real atmosphere states with time. Although sometimes the initial errors may counteract with the model error at initial time, this kind of opposite effect may not always perpetuate since the evolution of the initial errors is intangible. Therefore, as long as the initial error and the model error make the estimation of model emerge a slight deviation from real atmosphere, the error would generally start growing. For example, in current operational forecast, the model errors in

Fig. 5 Difference in AR of temperature (left panels) and wind speed (right panels) between CFSR and Interim. Upper, middle and lower panels are at 200, 500 and 850 hPa, respectively
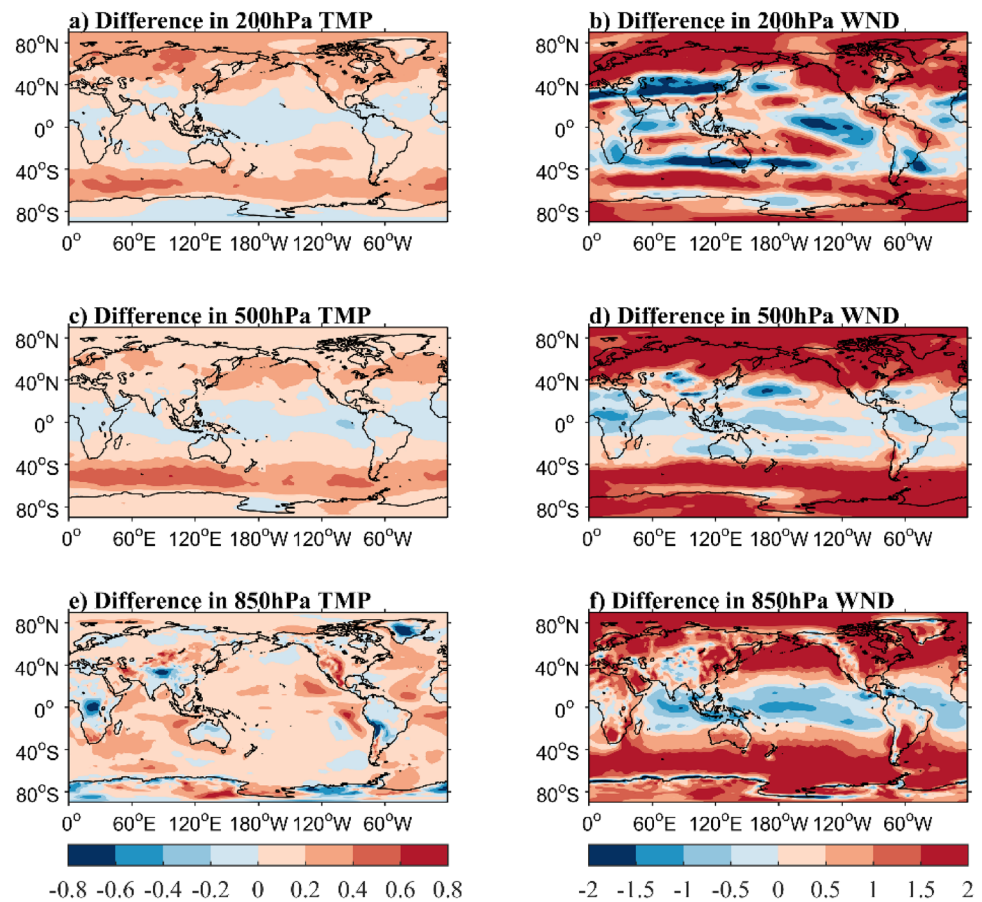
an imperfect coupled model result from the dynamic cores, couplers, numerical schemes, physical parameterization schemes and empirical parameters. Initial errors are from the misfitting forecast initial conditions obtained by assimilation. The data assimilation includes two steps: the analysis step and the forecast step. In analysis step, the observation information is combined with the model first guess and it will bring the misrepresentation of observation, sampling error and uncertainty of assimilation schemes into the initial condition. While in forecast step, model initialized by the analysis field produce the forecast to the next observation as the next first guess and the model errors will also blend into the initial condition (Zhang et al. 2020). Thus, the errors between the forecast and observation grow faster than errors between reference trajectory and estimated analogous trajectory in observation itself. Additionally, Li et al. (2018) verified this point using Lorenz63 model as real system with an imperfect forecast model. It showed that RMSE between the Lorenz63 model and the imperfect Lorenz63 model converges to the GAD and the practical predictability limits of imperfect model are shorter than potential predictability of Lorenz63 model.

To comprehend this inconsistency, we select the point (120° E, 0°) ((120° E, 45° N)) where predictability limits

of model are longer (shorter) than $\widetilde{T_P}(N)$ and apply NLLE method to obtain $\widetilde{T_P}(N)$ varying with length of time series. As shown in Fig. 7, at equator, the predictability limit doubles from 9 to 18 days, however at the point (120°E, 45°N) it increases slightly with extended data. Constrained by current length of observational data we may not be able to estimate *OLDA* in the tropics so the $\widetilde{T_P}$ over the tropics in Fig. 2 are shorter than the T_P which should still be longer than practical predictability limits of model forecast, but has not been available yet. While in mid-latitudes, $\widetilde{T_P}$ is longer than practical predictability limits of forecast reflecting that $\widetilde{T_P}$ in mid-latitude area is closer to T_P than $\widetilde{T_P}$ in the tropics. Thus, different geographic locations not only exist different $\widetilde{T_P}(N)$, but also require different length of observational data N to estimate the *OLDA*. What kinds of physical processes lead to the demand of longer observational data over the tropics needs to be addressed more in the future studies.

The reason of conspicuous high values in Fig. 6 need to further understand. Referring to the previous work (Reichler and Roads 2004, 2005), we learn that the low-frequency periodic oscillations such as MJO (Madden and Julian 1994), downward propagated stratospheric anomalies (Thompson et al. 2002), atmospheric teleconnection patterns represented
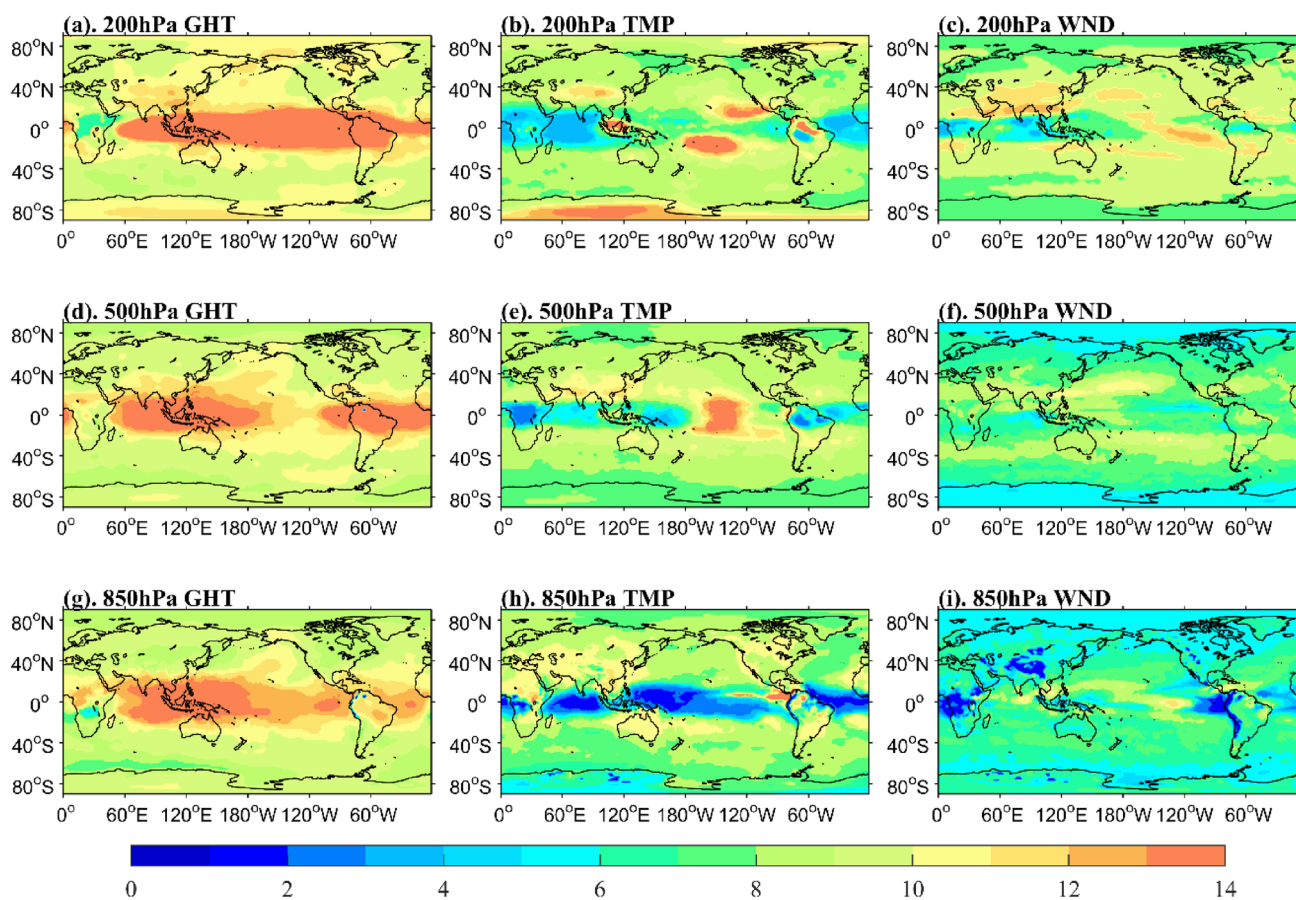
Fig. 6 Spatial distribution of practical predictability limits of GHT (left panels), temperature (middle panels) and wind speed (right panels) from ECWMF 46-day reforecast running twice a week (1996.01–2017.6). Upper, middle and lower panels are at 200, 500 and 850 hPa, respectively

in the 200 hPa height variability (Kumar et al. 2011) as well as boundary forcing from sea surface temperatures underneath (Shukla 1998) could extend the predictability by both
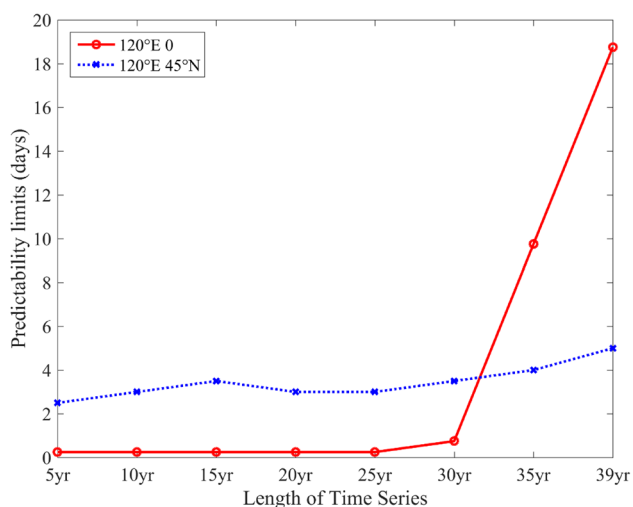
Fig. 7 Estimated predictability limits (in days) of 200 hPa GHT in Interim vary with length of time series from 5 to 39 years at 120° E 0 (red line) and 120° E 45° N (blue line)

initial-value and boundary-value. The Ensemble Empirical Mode Decomposition (EEMD) (e.g., Huang et al. 1998), a reliable method to process time series with instability and nonlinearity is applied to reveal the multi-time scale features and extract significant periods of the observational series in order to unveil this phenomenon.

Figure 8 shows the variance contribution of periods in different modes of GHT at 850 and 200 hPa statistically significant at the 95% confidence level using EEMD method. At 850 hPa, intraseasonal oscillations (ISOs) play a dominant role by which the variance contribution of the short periods (20–40 days) is more than 20%. But at 200 hPa, the variance contribution of long periods (2–4 years) is over 40% so the 200 hPa GHT is more influenced by low-frequency signals than the 850 hPa GHT. The sources of predictability of the atmosphere in a coupled earth system include the propagation of the atmospheric initial signals, and the interactions between coupled components such as the atmosphere, ocean, land processes and sea ice etc., which may impose longer timescale information (Bauer et al. 2015). Through the air–sea interaction or stratosphere–troposphere interaction, ocean and long-time scale oscillation within the atmosphere
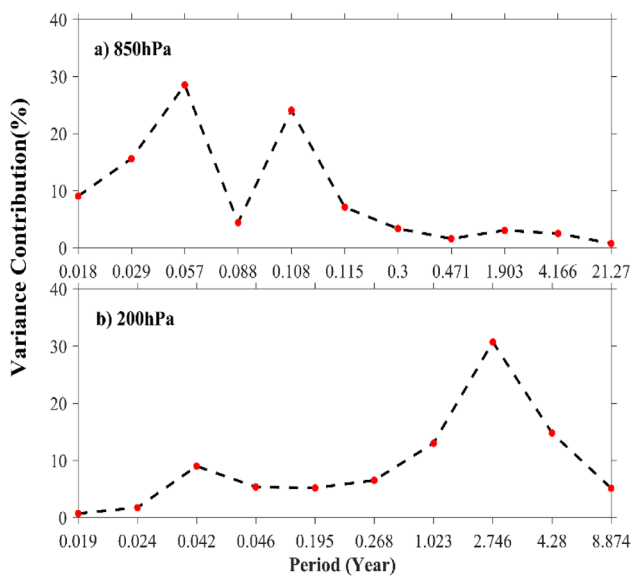
Fig. 8 The periods of area mean GHT over tropical Pacific (20° S–20° N, 120° E–80° W) from Jan 1979 to Dec 2017 at **a** 200 hPa and **b** 850 hPa. These periods are statistically significant at the 95% confidence level obtained by the EEMD method from Interim

can suppress the high frequency variation signals and add the slow-varying signals to the atmosphere, extending the predictability limit. The quasi-biennial oscillation (QBO) in the zonal wind of the tropical stratosphere is an important predictability source and can be well forecasted for many months in advance (Boer and Hamilton 2008) and it is possible to modulate deep convection in the tropics (e.g., Collimore et al. 1998, 2003) so we speculate that maybe the high value of predictability limits over the tropics in ECMWF are related to QBO. Meanwhile, some researches show that ECMWF-IFS has relatively higher predicting skill compared to other models in S2S datasets (e.g., Lim et al. 2019; Garfinkel et al. 2018). Žagar et al. (2007) proved that the phase of the QBO has an impact on the background error variances and the recent model improvements, primarily in the model physics, have substantially reduced the errors in both wind and GHT throughout the tropical atmosphere.

To illustrate the sensitivity of predictability limits in models to the low-frequency signal, we use Butterworth of order 1 (the Butterworth filter is a method that can extract the significant period of timeseries and obtain its main modes) to filter the period over 2 years for CFSR and Interim at each grid point. The AR presented in Sect. 3 is recalculated using the filtered reanalysis data, shown in Fig. 9. Compared to Fig. 6, the high value areas of practical predictability limits of GHT in ECMWF-IFS are sensitive (insensitive) to the high-pass filter process over the tropics (mid-latitudes) and show a higher sensitivity than other 2 coupled models. Therefore, the low-frequency signals, e.g., QBO have

significant impact of predictability limits in tropical GHT forecast and it is extremely urgent and important to develop model physics of QBO well represented in current numerical models to extend predictability limits of GHT.

The predictability limits of GHT, temperature and wind speed of CFSv2 are presented in Fig. 10. The GHT predictability limits have low values in the tropics, about 3–8 days at 200 hPa, 4–8 days at 500 hPa and 6–10 days at 850 hPa, decreasing with altitudes. The GHT predictability limits in mid latitudes are similar on all 3 levels, about 8–12 days, as same as the ECMWF's. The predictability limits of temperature have a similar distribution to that of GHT at 200 and 500 hPa. There is a high predictability over central tropical Pacific at 850 hPa. The wind speed predictability limits agree with the ECMWF patterns on all levels that the predictability decays from the upper troposphere to lower troposphere.

Consistent with the CFSv2 case, the practical predictability limits of GHT of HadGEM3 have low values of 0–2 and 1–6 days over the tropics at 200 and 500 hPa, respectively, whereas at 850 hPa the maximum predictability limits are over the tropics (Fig. 11). The temperature predictability limits of HadGEM3 have the similar spatial distribution to CFSv2 but with the weaker magnitudes over the tropics at 200 and 500 hPa declining to 2–4 and 2–8 days, respectively. At 850 hPa, restricted by topography, the high-altitude areas exist low value in maps. The predictability limits of wind speed also show high values of 8–12, 6–11 and 4–8 days over prevailing westerlies at 200, 500 and 850 hPa, and show the lower value of 2–6 days over the tropics decreasing from upper troposphere to lower troposphere.

The impact of air–sea coupling on lengthening predictability over the tropics in the lower troposphere is discussed below. To extend the general characteristics of predictability limit between coupled and uncoupled models, the zonal mean profiles of predictability limits of temperature, 850 hPa GHT and mean sea level pressure of multi-model ensemble mean are shown in Fig. 12. The coupled multi-model ensemble includes CFSv2, IFS and HadGEM3 and the uncoupled includes GEPS, ISAC-CNR Model and GEFS. In genal, zonal mean predictability limits of all variables in coupled models is longer than those of uncoupled models and the mean predictability limits of coupled models improve remarkably in the tropics represented a near doubling in GHT and temperature. This general difference suggests that air–sea coupling in models contributes to improved synoptic-scale prediction skill in lower troposphere especially in the tropics. Zhao et al. (2021) compared atmosphere predictability limits among different configurations of the coupled conceptual model. It is showed that more accurate ocean boundary condition prolongs the atmosphere predictability limit. However, the improvement of prediction skill depends on variable. Compared to temperature, zonal
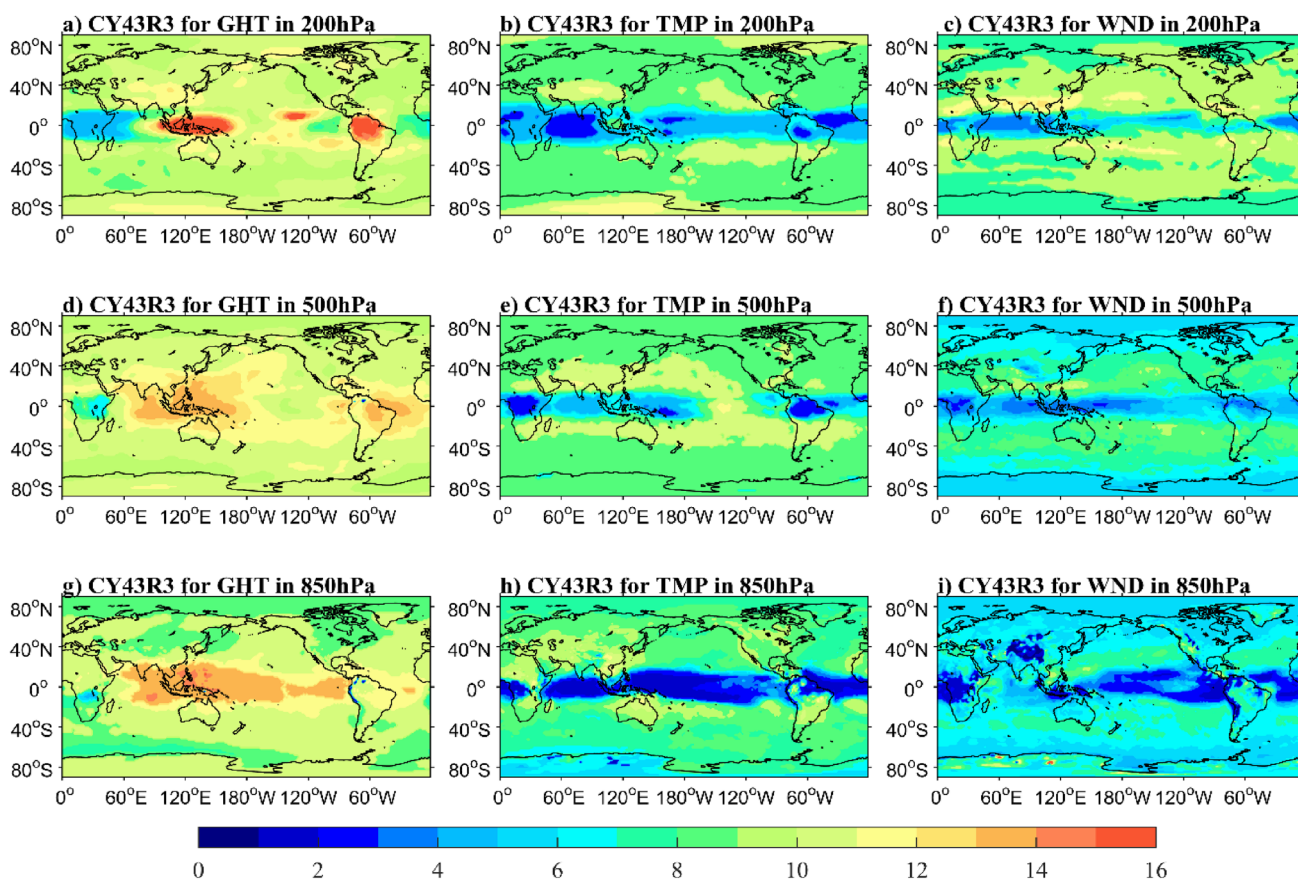
Fig. 9 Same as Fig. 6 but after the Butterworth 2-year high-pass filtering

mean predictability limits of 850 hPa GHT increase further in both the tropics and mid-latitudes. Moreover, for coupled models, the predictability limits of 850 hPa temperature are lower over the tropics than other regions, but the 850 hPa GHT of coupled models have almost equal limits over the tropics to other areas. This could be explained by air–sea interactions in coupled models. Through air–sea interactions, low-frequency signals imported to the atmosphere as the boundary condition could extend predictability of the atmosphere. Under such a circumstance, since the temperature only reflects the thermal properties at the current pressure level while the GHT contains all information of the column underneath, the GHT could be impacted more than the temperature. For example, the zonal mean predictability limits of sea level pressure have similar behavior as ones of the 850 hPa GHT, but a little different from the temperature's (Fig. 12c).

4.2 Generalized predictability and its uncertainty

Many previous relevant researches use certain model which only reflect predictability of one model. According to the Sect. 4.1, the predictability illustrates the model dependency.

We try to mitigate model dependency to obtain the generalized predictability using multi-model ensemble. Just like the higher skills of ensemble mean forecasts than those of deterministic forecast (e.g., Feng et al. 2019), many researches (e.g., Rajagopalan et al. 2002; Robertson et al. 2004; Doblas-Reyes et al. 2005; Stephenson et al. 2005) demonstrated multi-model ensemble combination (MMEC) increase prediction skill due to error cancellation and nonlinearity of the skill metrics applied (Hagedorn et al. 2005) so the generalized predictability is also roughly estimation of MMEC prediction skill.

Figure 13 shows the generalized predictability limits of coupled forecast models, CFSv2, ECMWF-IFS and HadGEM3 considered as a benchmark to reflect the mean characteristics of multi-model ensemble. The generalized predictability limits are the time by which the mean RMSE of 3 models reaches the average AR of CFSR and Interim at each grid point.

For the 200 hPa GHT, the high predictability limits of 8–10 days are over mid-latitudes while the low predictability limits of 3–8 days are in the tropics. There exists a high predictability dipole across the equator over central tropical Pacific. At 500 hPa, the predictability limits have similar
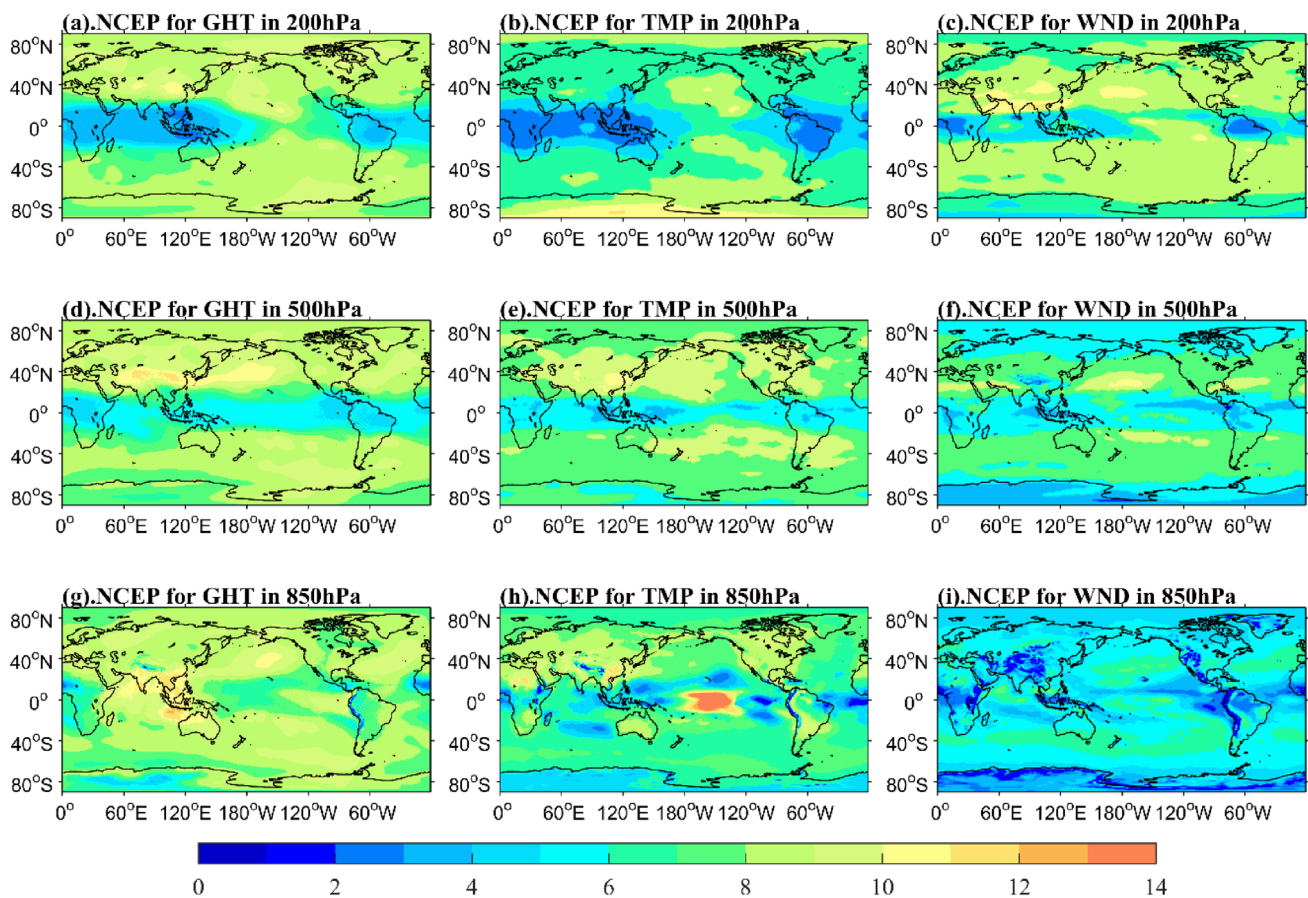
Fig. 10 Same as Fig. 6, but from CFSv2 45-day reforecast running every 6 h (00, 06, 12, 18 UTC) from 1999 to 2010

patterns as at 200 hPa but with much higher values in the tropics, while the 850 hPa predictability limits increase over warm pool.

For temperature, the predictability limits of 2–5 days and 6–10 days are observed in the tropics and mid-latitudes overall on all the pressure levels. For wind speed, the high (low) predictability limits are observed in westerlies (tropics) in all 3 levels but large values are in low troposphere increasing with height. In terms of variables, the GHT predictability limits over the tropics are the highest, while the wind speed's is the lowest and the temperature's is the middle.

The "generalized" predictability limits among 3 models over the tropics are far below the upper limits defined by the NLLE method, as shown by Figs. 2 and 13. This suggests that, in general, the current forecast skill in the tropics lag to that in mid-latitudes (Kanamitsu 1985; Reynolds et al. 1994), not only because of the spare observations but also due to the complex tropical dynamics. In the tropics, when the Coriolis parameter is small, maintenance of the quasi-geostrophic balance becomes difficult. Furthermore, the barotropic and convective instability and their interaction are dominated over the tropics and convective processes are

represented by the sub-grid scale parameterizations, which are major challenges in modeling and it is expected that a great enhancement of predictability over the tropics could be accompanied with the improvement on modeling of tropical convective processes.

5 The distinction of predictability in different prediction systems

The distinction of predictability between single model and generalized predictability reflects prediction skill of single models and a measure to demonstrate whether MMEC could outperform the single model extending the practical predictability. Figure 14 presents the distinction of GHT predictability limits in different forecast models. The ECMWF-IFS provides a more accurate forecasts than other two models, on average, beyond 5 days on 200 and 500 hPa, and 3–5 days at 850 hPa over the tropics. Comparing Fig. 6 with Fig. 2, we find that only the predictability limits of ECMWF-IFS have a similar pattern that has high values in tropical areas reaching to the upper limits as that in the observed atmosphere
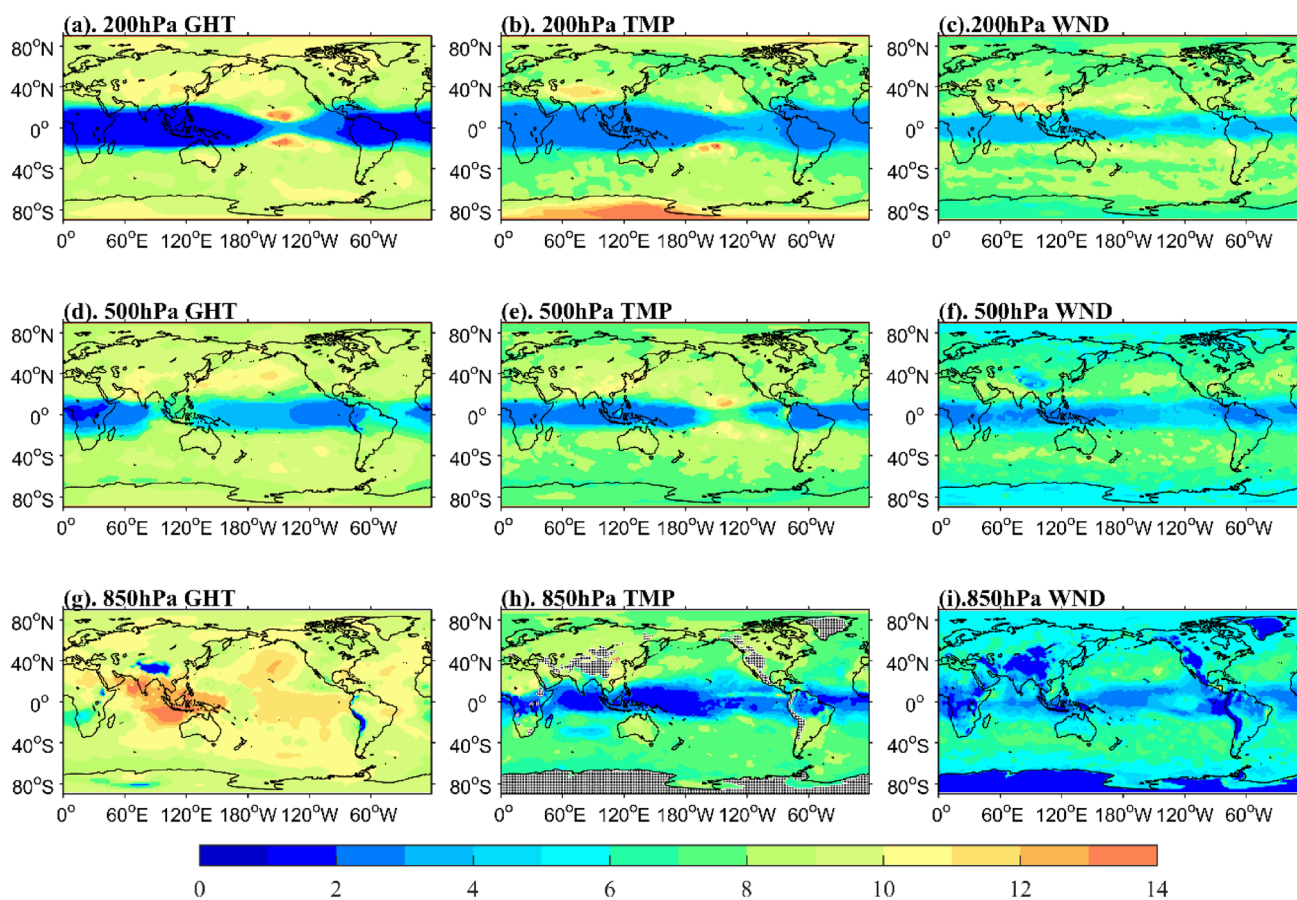
Fig. 11 Same as Fig. 5, but from HadGEM3 60-day reforecast running every day from 1993 to 2015. The shaded areas indicate the forecast data default

indicating that the forecasts of ECMWF-IFS are close to the maximum skill for deterministic forecasting without further reducing the initial state error (Froude et al. 2013). The predictability limits of CFSv2 have a distribution close to the ensemble mean states while the HadGEM3 has the limited forecast skill over the tropics at 200 hPa. It is also shown that the ECMWF-IFS has the higher forecast skill of GHT over the tropics on all 3 levels while the skills of CFSv2 and HadGEM3 are lower than MMEC. The 3 models share common predictability limit characteristics in mid-latitudes. It is indicated that MMEC can locally improve the GHT prediction skill in extratropic by capturing more underlying uncertainties but does not enhance GHT skill in the tropics as far as ECMWF-IFS concerned. Practically, we should pay more attention to ECMWF-IFS forecast in tropic and to MMEC results in extratropic.

The difference of temperature against the ensemble mean is shown in Fig. 15. We find that none of the 3 models can provide an excellent forecast on all pressure levels globally. For example, although the ECMWF-IFS has relatively accurate forecasts at 200 and 500 hPa in the tropics, its 850 hPa tropical (200 and 500 hPa extratropical) forecasts are not so

good as the CFSv2's (HadGEM3's). Unlike GHT, extratropical temperature forecast in HadGEM3 beats the MMEC while as for lower troposphere temperature in extratropic we still need refer to results in MMEC.

The difference of wind speed predictability limits between individual models and the ensemble mean is shown in Fig. 16. It is found that the CFSv2 and ECMWF-IFS are complementary in the tropics. For example, the predictability limits of CFSv2 at 200 hPa have high values over the tropical Indian Ocean and central east Pacific while the ECMWF-IFS's have low values there by coincidence. The HadGEM3 predictability limits are mostly close to the ensemble mean. Because of probably good simulation of QBO in the ECMWF-IFS, it has better 200 hPa wind speed forecasts but still need to refer to the CFSv2 in some areas. Wind speed has great local uncertainty so MMEC could enhance the prediction skill in most of areas.

In summary, within these 3 models, there is no certain model providing a more accurate forecast than other 2 models for all 3 variables on all levels. Different models have different data assimilation methods, numerical schemes and physical parameterizations that may lead to their better

Fig. 12 Zonal mean profile of predictability limits of 850 hPa temperature (**a**), 850 hPa GHT (**b**) and mean sea level pressure (**c**) obtained from coupled models (blue line) and uncoupled models (red line) ensemble mean with standard deviation in respective ensembles (filled curves). The coupled multi-model ensemble includes CFSv2, IFS and HadGEM3 and the uncoupled one includes GEPS, ISAC-CNR Model and GEFS
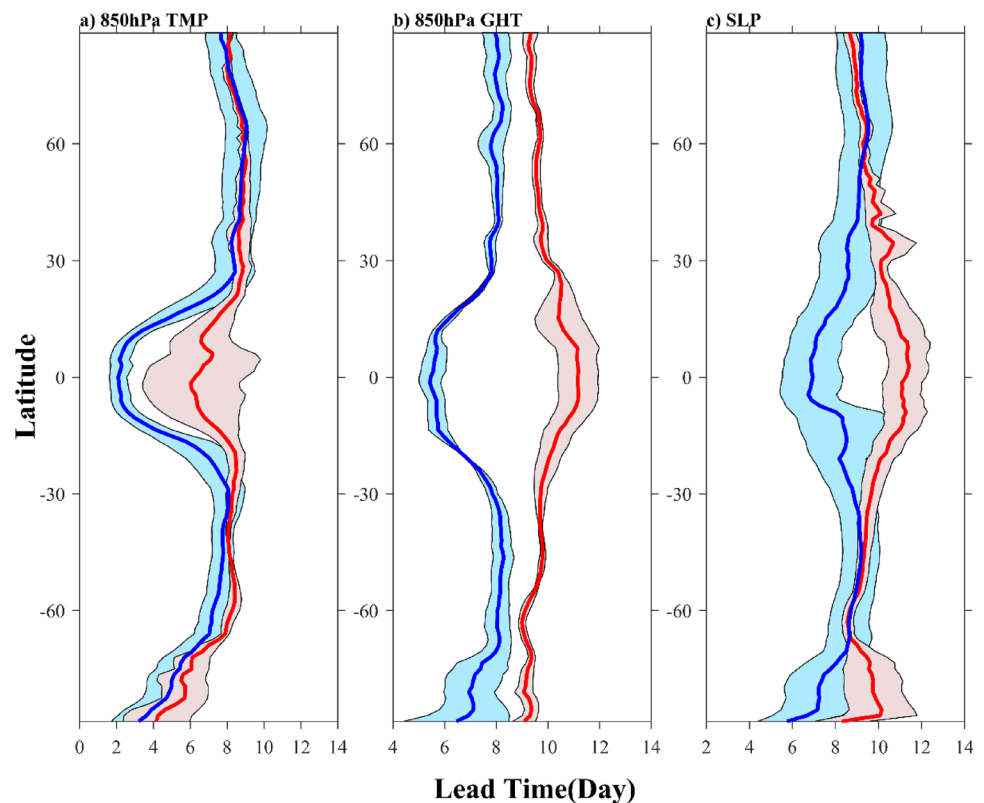
performances that have a latitude and variable dependency. Meanwhile, we understand that MMEC is not able to beat any single model in any location. According to Weigel et al. (2008), the effect of MMECC is to gradually widen the ensemble spread and to move the ensemble mean toward the "truth" and will improve the overall skill only if the participating single model ensembles are overconfident. Thus, we need to refer to the ensemble mean and each individual model property to understand the generalized predictability and its uncertainties.

6 Summary and discussions

The multiple reanalysis and coupled prediction systems have been used to study the generalized predictability of the atmosphere in the coupled ocean–atmosphere system. The upper predictability limits of the atmosphere are first analyzed through the NLLE and AR methods using the NCEP CFSR and ECMWF Interim reanalysis products. Then, three world-widely known coupled forecast systems, ECMWF-IFS, NCEP-CFSv2, and UKMO-HadGEM3 are selected to generalize the recognition of the atmosphere predictability in the coupled system. Based on the predictability limits of each coupled system, the distribution of generalized

predictability limits is used to evaluate the average forecast skill as a metric to measure the forecast ability of the atmosphere.

While different methods (i.e. NLLE and AR) and different reanalysis products (i.e. ECMWF Interim and NCEP CFSR) show nearly identical upper predictability properties of the atmosphere, it is found that the atmosphere upper predictability limits increase with height (maximum as 12, 14 and 16 days for 850, 500 and 200 hPa tropical GHTs respectively). The predictability of temperature and wind speed is a little lower than that of GHT.

Our results of multi-model studies show that the predictability of the atmosphere in coupled models is higher than in uncoupled models, especially for the GHTs of the low troposphere. Given that temperature only reflects the thermal properties in a current pressure level while GHT contains all information of the column, the air–sea interactions in coupled models as the atmosphere boundary conditions could extend the predictability of GHT more than temperature indicating a significant impact of air–sea coupling on predictability in lower troposphere especially over the tropics.

Given that different models have different data assimilation methods, numerical scheme and physics parameterization schemes, the distinction between the predictability characteristics of different models exist. None of them is able to provide more accurate forecasts for all variables on all pressure levels than others. Meanwhile MMEC cannot
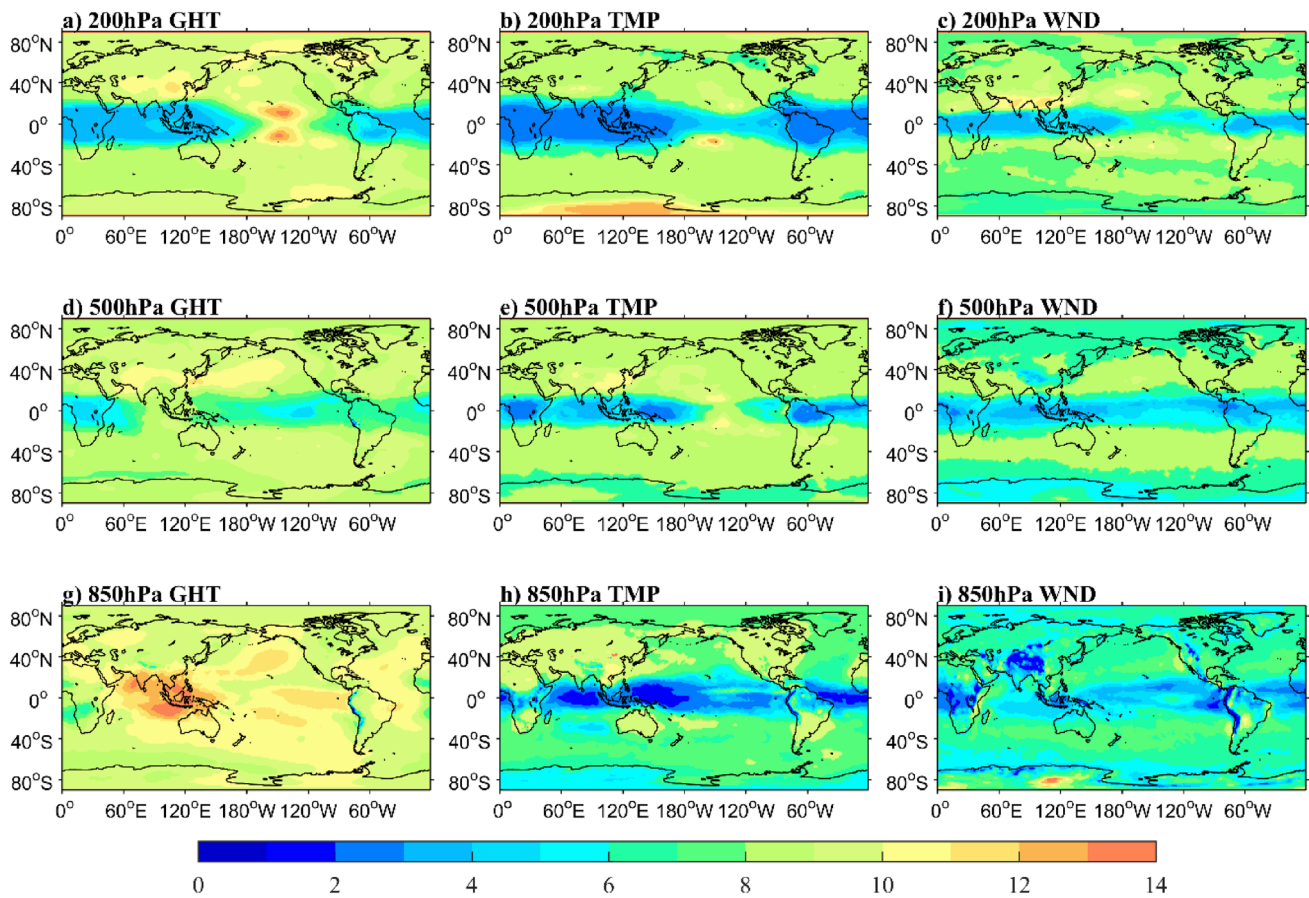
Fig. 13 Spatial distribution of temperature (left panels) and wind speed (right panels) mean practical predictability limits of IFS, CFSv2 and HadGEM3. Upper, middle and lower panels are at 200, 500 and 850 hPa, respectively

outperform the all participating models in all variables. We shall refer to multiple model results according to geographical location and variable in practical forecast. However, the generalized predictability as multi-model ensemble indeed shows that the predictability of the tropics is far below the upper limits calculated by the reanalysis data, suggesting that the coupled models has a great potential to increase the predictability of the tropics, thus greatly enhancing model forecast skills.

Challenges still exist and need to be resolved in further studies. The most urgent one is that what kind of physical

mechanisms that lead to the 2-year domain long period, ENSO, QBO, or other modes? How they affect the upper troposphere predictability so significantly? What kinds of dynamical mechanisms contribute to lower potential predictability over mid latitudes in Southern Hemisphere? The answers of these questions would greatly advance our understanding of atmosphere predictability and improve the coupled earth system model, enhancing the accuracy of atmosphere and ocean predictions. In addition, the predictability limits of the other components such as the ocean,
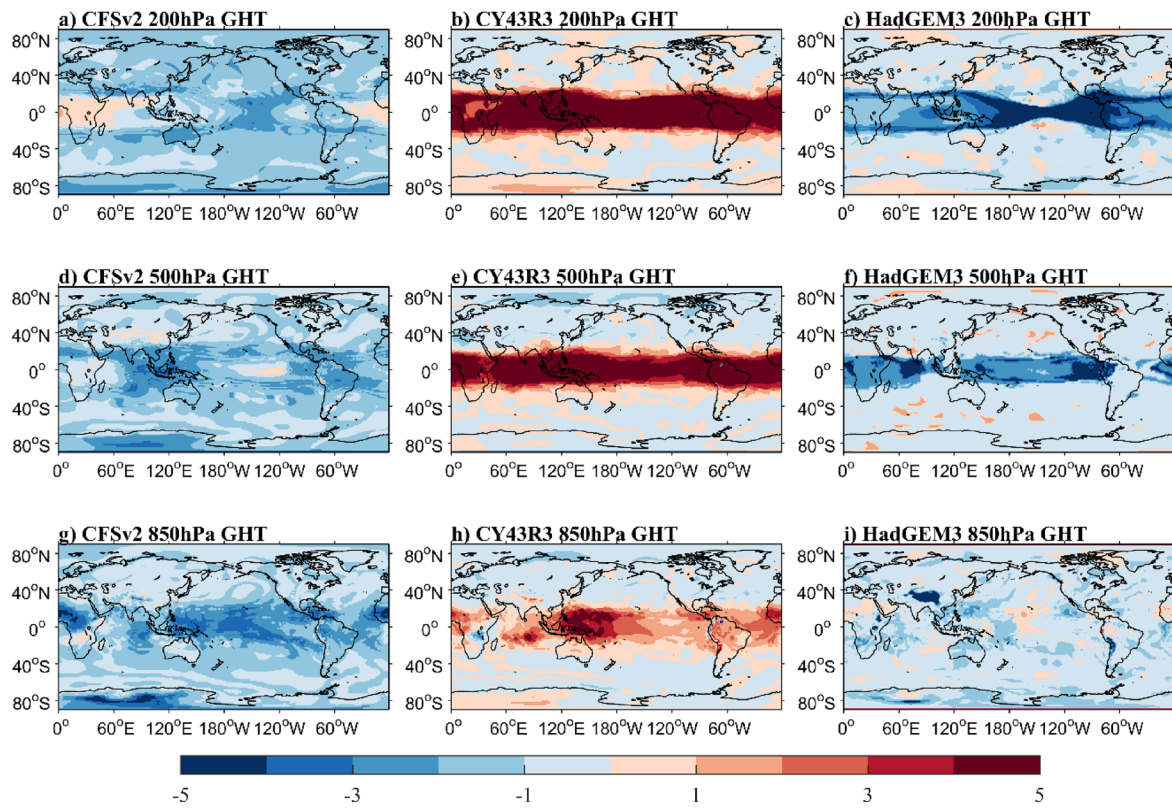
Fig. 14 Differences of GHT practical predictability limits between individual models and the ensemble mean. Results from CFSv2, IFS and HadGEM3 are shown from the left to the right. Upper, middle and lower panels are at 200, 500 and 850 hPa, respectively
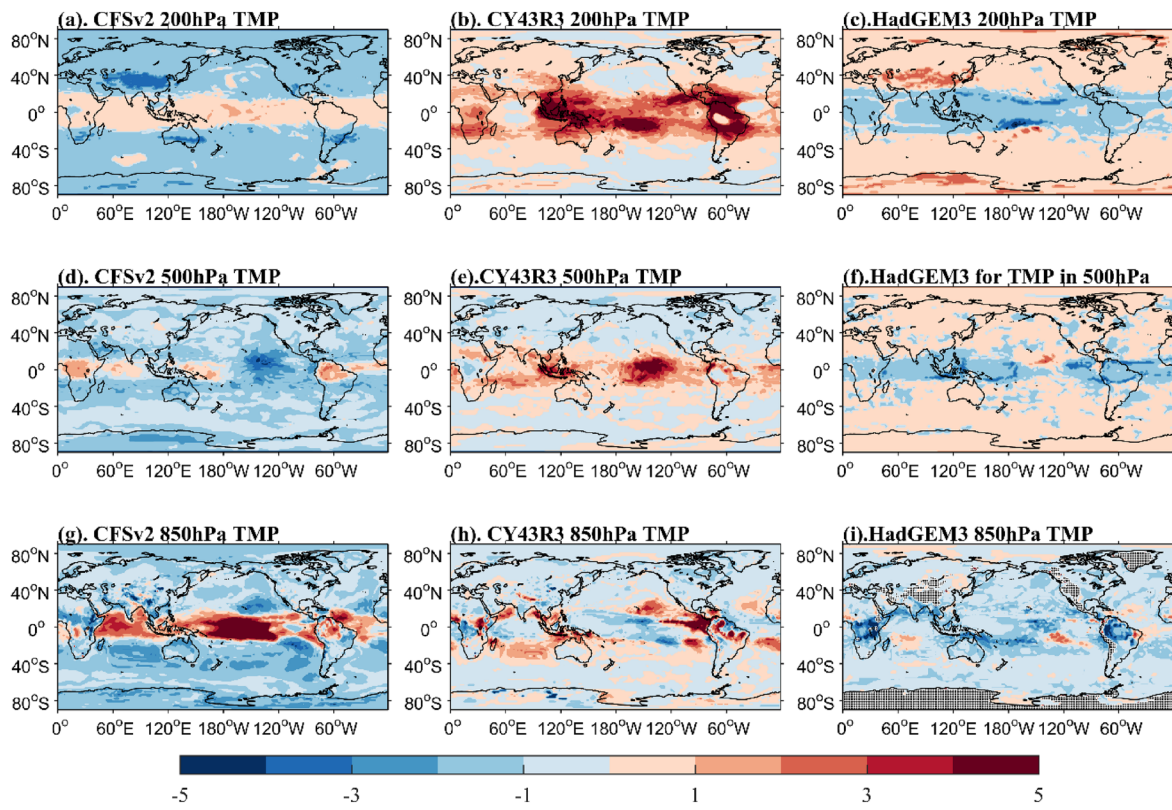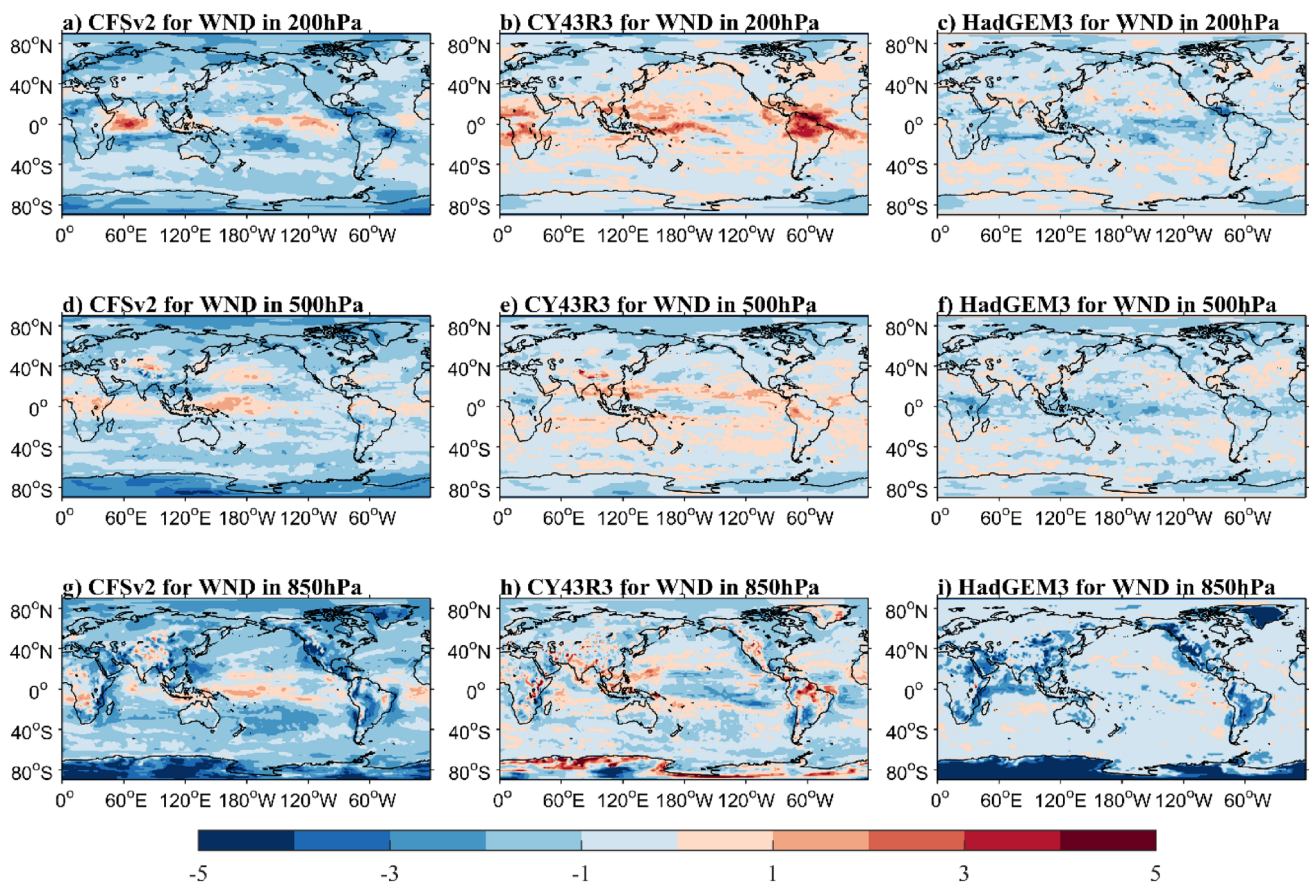
Fig. 15 Same as Fig. 14, but for temperature

Fig. 16 Same as Fig. 14, but for wind speed

sea ice and soil temperature and moisture are important and interesting research topics for the follow-up studies.

References

Baldwin MP, Dunkerton TJ (2001) Stratospheric harbingers of anomalous weather regimes. Science 294(5542):581–584

Baldwin MP, Stephenson DB, Thompson DW, Dunkerton TJ, Charlton AJ, O'Neill A (2003) Stratospheric memory and skill of extended-range weather forecasts. Science 301(5633):636–640

Bauer P, Thorpe A, Brunet G (2015) The quiet revolution of numerical weather prediction. Nature 525(7567):47–55. https://doi.org/10.1038/nature14956

Benettin G, Galgani L, Giorgilli A, Strelcyn J (1980) Lyapunov characteristics exponents for smooth dynamical systems and for Hamiltonian systems; a method for computing all of them. Meccanica 15(1):9–20. https://doi.org/10.1007/BF02128236

Boer GJ, Hamilton K (2008) QBO influence on extratropical predictive skill. Clim Dyn 31:987–1000. https://doi.org/10.1007/s00382-008-0379-5

Buizza R (2010) Horizontal resolution impact on short- and long-range forecast error. Q J R Meteorol Soc 136(649):1020–1035

Collimore CC, Hitchman MH, Martin DW (1998) Is there a quasibiennial oscillation in tropical deep convection? Geophys Res Lett 25(3):333–336

Collimore CC, Martin DW, Hitchman MH, Huesmann A, Waliser DE (2003) On the relationship between the QBO and tropical deep convection. J Clim 16:2552–2568

Dalcher A, Kalnay E (1987) Error growth and predictability in operational ECMWF forecasts. Tellus A 39(5):474–491

Dee DP, Uppala SM, Simmons AJ et al (2011) The Interim reanalysis: configuration and performance of the data assimilation system. Q J R Meteorol Soc 137:553–597. https://doi.org/10.1002/qj.828

Ding RQ, Li JP (2007) Nonlinear finite-time Lyapunov exponent and predictability. Phys Lett A 364:396–400

Doblas-Reyes FJ, Hagedorn R, Palmer TN (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting. Part II: calibration and combination. Tellus A 57(3):234–252

Douville H (2009) Stratospheric polar vortex influence on Northern Hemisphere winter climate variability. Geophys Res Lett 36:L18703. https://doi.org/10.1029/2009GL039334

Feng J, Li JP, Zhang J, Liu DQ, Ding RQ (2019) The relationship between deterministic and ensemble mean forecast errors revealed by global and local attractor radii. Adv Atmos Sci 36(3):271–278. https://doi.org/10.1007/s00376-018-8123-5

Froude LSR, Bengtsson L, Hodges KI (2013) Atmospheric predictability revisited. Tellus A: Dyn Meteorol Oceanogr 65(1):19022

Garfinkel CI, Schwartz C, Domeisen DIV, Son S-W, Butler AH, White IP (2018) Extratropical atmospheric predictability from the quasi-biennial oscillation in subseasonal forecast models. J Geophys Res Atmos 123:7855–7866. https://doi.org/10.1029/2018JD028724

Hagedorn R, Doblas-Reyes FJ, Palmer TN (2005) The rationale behind the success of multi-model ensembles in seasonal forecasting—I. Basic concept. Tellus A: Dyn Meteorol Oceanogr 57(3):219–233

Hitchcock P, Simpson IR (2014) The downward influence of stratospheric sudden warmings. J Atmos Sci 71:3856–3876. https://doi.org/10.1175/JAS-D-14-0012.1

Huang NE, Shen Z, Long SR et al (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and nonstationary time series analysis. Proc Roy Soc Lond 454A:903–995

Kazantsev E (1999) Local Lyapunov exponents of the quasigeostrophic ocean dynamics. Appl Math Comput 104:217–257

Kumar A, Schubert SD, Suarez MS (2003) Variability and predictability of 200-mb seasonal mean heights during summer and winter. J Geophys Res 108:4169. https://doi.org/10.1029/2002JD002728

Kumar A, Chen M, Wang W (2011) An analysis of prediction skill of monthly mean climate variability. Clim Dyn 37:1119–1131

Kuroda Y (2008) Role of the stratosphere on the predictability of medium-range weather forecast: a case study of winter 2003–2004. Geophys Res Lett 35:L19701. https://doi.org/10.1029/2008GL034902

Lacarra JF, Talagrand O (1988) Short-range evolution of small perturbations in a barotropic model. Tellus A 40:81–95

Li JP, Ding RQ (2009) Studies of predictability of single variable from multi-dimensional chaotic dynamical system (in Chinese with English abstract). Chin J Atmos Sci 33:551–556

Li JP, Ding RQ (2011) Temporal-spatial distribution of atmospheric predictability limit by local dynamical analogues. Mon Weather Rev 139:3265–3283

Li JP, Ding RQ (2015) Seasonal and interannual weather prediction. In: North G, Pyle J, Zhang F (eds) Encyclopedia of atmospheric sciences, vol 6, 2nd edn. Academic Press and Elsevier, Amsterdam, Boston, pp 303–312

Li JP, Wang S (2008) Some mathematical and numerical issues in geophysical fluid dynamics and climate dynamics. Commun Comput Phys 3:759–793

Li JP, Ding RQ, Chen BH (2006) Review and prospect on the predictability study of the atmosphere. In: Review and prospects of the developments of atmosphere sciences in early 21st century, China Meteorology Press, pp 96–104

Li JP, Feng J, Ding RQ (2018) Attractor radius and global attractor radius and their application to the quantification of predictability limits. Clim Dyn 51:2359–2374

Lim EP, Hendon HH, Boschat G, Hudson D, Thompson DWJ, Dowdy AJ, Arblaster JM (2019) Australian hot and dry extremes induced by weakenings of the stratospheric polar vortex. Nature Geosci 12:896–901. https://doi.org/10.1038/s41561-019-0456-x

Lorenz EN (1963) Deterministic nonperiodic flow. J Atmos Sci 20:130–141

Lorenz EN (1965) A study of the predictability of a 28-variable atmospheric model. Tellus 17:321–333

Lorenz EN (1969) The predictability of a flow which possesses many scales of motion. Tellus 21:289–307

MacLachlan C, Arribas A, Peterson KA et al (2015) Global seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system. Q J R Meteorol Soc 141:1072–1084. https://doi.org/10.1002/qj.2396

Mu M (2000) Nonlinear singular vectors and nonlinear singular values. Sci China 43D:375–385

Mu M, Duan W, Tang Y (2017) The predictability of atmospheric and oceanic motions: further understanding, prospects and exploration. Sci China Earth Sci 60:2001–2012. https://doi.org/10.1007/s11430-016-9101-x

Orrell D, Smith L, Barkmeijer J, Palmer TN (2001) Model error in weather forecasting. Nonlinear Process Geophys 8:357–371

Oseledec VI (1968) A multiplicative ergodic theorem: Lyapunov characteristic numbers for dynamical systems. Trans Moscow Math Soc 19:197–231

Rajagopalan B, Lall U, Zebiak SE (2002) Categorical climate forecasts through regularization and optimal combination of multiple GCM ensembles. Mon Weather Rev 130:1792–1811

Reichler T, Roads JO (2004) Time–space distribution of long-range atmospheric predictability. J Atmos Sci 61:249–263

Reichler T, Roads JO (2005) Long-range predictability in the tropics. Part I: monthly averages. J Clim 18:619–633

Renwick JA, Wallace JM (1996) Relationships between North Pacific Wintertime Blocking, El Niño, and the PNA Pattern. Mon Weather Rev 124:2071–2076

Robertson AW, Lall U, Zebiak SE, Goddard L (2004) Improved combination of multiple atmospheric GCM ensembles for seasonal prediction. Mon Weather Rev 132:2732–2744

Rowell DP (1998) Assessing potential seasonal predictability with an ensemble of multidecadal GCM simulations. J Clim 11:109–120

Saha S, Moorthi S, Pan H et al (2010) The NCEP climate forecast system reanalysis. Bull Am Meteorol Soc 91:1015–1057

Saha S, Moorthi S, Wu X et al (2014) The NCEP climate forecast system version 2. J Clim 27:2185–2208

Sano M, Sawada Y (1985) Measurement of the Lyapunov spectrum from a chaotic time series. Phys Rev Lett 55:1082–1085

Shimada I, Nagashima T (1979) A numerical approach to ergodic problem of dissipative dynamical systems. Prog Theor Phys 61:1605–1616

Simmons AJ, Hollingsworth A (2002) Some aspects of the improvement in skill of numerical weather prediction. Q J R Meteorol Soc 128:647–677

Stephenson DB, Coelho CAS, Doblas-Reyes FJ, Balmaseda M (2005) Forecast assimilation: a unified framework for the combination of multi-model weather and climate predictions. Tellus 57A:253–264

Thompson PD (1957) Uncertainty of initial state as a factor in the predictability of large-scale atmospheric flow pattern. Tellus 9:275–295

Thompson DWJ, Baldwin MP, Wallace JM (2002) Stratospheric connection to northern hemisphere wintertime weather: implications for prediction. J Clim 15(12):1421–1428

Van den Dool HM (1994) Searching for analogues, how long must we wait? Tellus A 46:314–324

Vitart F (2014) Evolution of ECMWF sub-seasonal forecast skill scores. Q J R Meteorol Soc 140:1889–1899. https://doi.org/10.1002/qj.2256

Vitart F, Ardilouze C, Bonet A et al (2017) The sub-seasonal to seasonal prediction (S2S) project database. Bull Am Meteorol Soc 2017:98

Webster PJ, Holton JR (1982) Cross-equatorial response to middle-latitude forcing in a zonally varying basic state. J Atmos Sci 39(4):722–733

Weigel AP, Liniger MA, Appenzeller C (2008) Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? Q J R Meteorol Soc 134:241–260. https://doi.org/10.1002/qj.210

Williams KD, Harris CM, Bodas-Salcedo A, Camp J, Comer RE, Copsey D, Fereday D, Graham T, Hill R, Hinton T, Hyder P, Ineson S, Masato G, Milton SF, Roberts MJ, Rowell DP, Sanchez C, Shelly A, Sinha B, Walters DN, West A, Woollings T, Xavier PK (2015) The Met Office Global Coupled model 2.0 (GC2) configuration. Geosci Model Dev 8:1509–1524. https://doi.org/10.5194/gmd-8-1509-2015

Wolf A, Swift JB, Swinney HL, Vastano JA (1985) Determining Lyapunov exponents from a time series. Phys D 16:285–317

Yoden S, Nomura M (1993) Finite-time Lyapunov stability analysis and its application to atmospheric predictability. J Atmos Sci 50:1531–1543

Žagar N, Andersson E, Fisher M, Untch A (2007) Influence of the Quasi-Biennial oscillation on the ECMWF model short-range-forecast errors in the tropical stratosphere. Q J R Meteorol Soc 133(628):1843–1853. https://doi.org/10.1002/qj.152

Zhang S, Liu Z, Zhang X, Wu X, Han G, Zhao Y, Yu X, Liu C, Liu Y, Wu S, Lu F, Li M, Deng X (2020) Coupled data assimilation and parameter estimation in coupled ocean–atmosphere models: a review. Clim Dyn 54(11–12):5127–5144

Zhao HR, Zhang SQ, Li JP, Ma YW (2021) A study of predictability of coupled ocean–atmosphere system using attractor radius and global attractor radius. Clim Dyn Online. https://doi.org/10.1007/s00382-020-05534-6

Ziehmann C, Smith LA, Kurths J (2000) Localized Lyapunov exponents and the prediction of predictability. Phys Lett A 4:237–251