



RESEARCH ARTICLE

10.1029/2021MS002570

Key Points:

- Initial shocks, drifts, and biases affect climate predictions at all time scales
- As they usually lead to a substantial reduction of the prediction skill, further efforts toward their understanding are needed
- The newly developed data set presented in this article is expected to serve as a useful tool for their study

Correspondence to:

R. I. Saurral,
saurral@cima.fcen.uba.ar

Citation:

Saurral, R. I., Merryfield, W. J., Tolstykh, M. A., Lee, W.-S., Doblas-Reyes, F. J., García-Serrano, J., et al. (2021). A data set for intercomparing the transient behavior of dynamical model-based subseasonal to decadal climate predictions. *Journal of Advances in Modeling Earth Systems*, 13, e2021MS002570. <https://doi.org/10.1029/2021MS002570>

Received 28 APR 2021

Accepted 23 AUG 2021

A Data Set for Intercomparing the Transient Behavior of Dynamical Model-Based Subseasonal to Decadal Climate Predictions

Ramiro I. Saurral^{1,2,3} , William J. Merryfield⁴ , Mikhail A. Tolstykh^{5,6,7} , Woo-Sung Lee⁴, Francisco J. Doblas-Reyes^{8,9}, Javier García-Serrano¹⁰, François Massonnet¹¹, Gerald A. Meehl¹² , and Haiyan Teng¹³

¹Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Ciencias de la Atmósfera y los Océanos, Buenos Aires, Argentina, ²CONICET – Universidad de Buenos Aires, Centro de Investigaciones del Mar y la Atmósfera (CIMA), Buenos Aires, Argentina, ³CNRS – IRD – CONICET – UBA, Instituto Franco-Argentino para el Estudio del Clima y sus Impactos (IRL 3351 IFAECI), Buenos Aires, Argentina, ⁴Canadian Centre for Climate Modeling and Analysis, Environment and Climate Change Canada, Victoria, British Columbia, Canada, ⁵Marchuk Institute of Numerical Mathematics, Russian Academy of Sciences, Moscow, Russia, ⁶Hydrometcentre of Russia, Moscow, Russia, ⁷Moscow Institute of Physics and Technology, Dolgoprudny, Russia, ⁸Barcelona Supercomputing Center (BSC), Barcelona, Spain, ⁹Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain, ¹⁰Group of Meteorology, Universitat de Barcelona (UB), Barcelona, Spain, ¹¹Université Catholique de Louvain, Georges Lemaître Centre for Earth and Climate Research, Earth and Life Institute, Louvain-la-Neuve, Belgium, ¹²National Center for Atmospheric Research, Climate and Global Dynamics Laboratory, Boulder, CO, USA, ¹³Pacific Northwest National Laboratory, Richland, WA, USA

Abstract Climate predictions using coupled models in different time scales, from intraseasonal to decadal, are usually affected by initial shocks, drifts, and biases, which reduce the prediction skill. These arise from inconsistencies between different components of the coupled models and from the tendency of the model state to evolve from the prescribed initial conditions toward its own climatology over the course of the prediction. Aiming to provide tools and further insight into the mechanisms responsible for initial shocks, drifts, and biases, this paper presents a novel data set developed within the Long Range Forecast Transient Intercomparison Project, LRFTIP. This data set has been constructed by averaging hindcasts over available prediction years and ensemble members to form a hindcast climatology, that is a function of spatial variables and lead time, and thus results in a useful tool for characterizing and assessing the evolution of errors as well as the physical mechanisms responsible for them. A discussion on such errors at the different time scales is provided along with plausible ways forward in the field of climate predictions.

Plain Language Summary The prediction of climate from weeks to years is usually made using complex numerical models that need to have an appropriate representation of the different components of the climate system (atmosphere, oceans, land areas, etc.) as well as of their interactions. Most of these simulations, however, are affected by errors that originate in part from the numerical methods and approximations used to solve the system of equations of the atmospheric and oceanic systems, and in part by an incorrect representation of the interaction between components. Naturally, these errors lead to a significant decrease in the skill of the predictions, and therefore, must be taken into account. This study presents a new data set made from climate model predictions that have been manipulated in order to be useful for the assessment and further understanding of the physical mechanisms behind the aforementioned errors. The expected increase in our knowledge in the field will help us obtain more skillful climate predictions in the future.

1. Introduction

The evergrowing field of climate predictions on subseasonal to multidecadal time scales can provide useful information for decisionmaking on economic activities such as agriculture, water management, and renewable energy (Cassou et al., 2018; Merryfield et al., 2020; Robertson et al., 2015). Such forecasts, derived from numerical models that simulate the different components of the climate system as well as their

interactions, are generally affected by different types of errors and biases (e.g., Hawkins et al., 2014; Meehl et al., 2014; Mulholland et al., 2015), which may severely reduce predictive skills (Doblas-Reyes et al., 2011; Kim et al., 2017; Shukla et al., 2018; Smith et al., 2013; Yeager et al., 2018). These errors are commonly described as *initial shock* and *model drift*. The first term refers to the fast adjustment of the model immediately after initialization caused by inconsistencies between different components of the coupled model (Balmaseda & Anderson, 2009; Mulholland et al., 2015), whereas the *model drift* defines the tendency of the model state to evolve from the prescribed initial conditions toward the model's own climatology over the course of the prediction (Mulholland et al., 2015), which differs from that of observations due to biases arising from systematic model errors (Zadra et al., 2018). Conceptually, shocks have been viewed as including initial imbalances that excite rapidly varying degrees of freedom such as gravity waves on the fast manifold, whereas drift represents balanced evolution on the slow manifold (Ballabrera-Poy et al., 2009; Cotter, 2013; Puri, 1981). As yet, however, there are no widely accepted definitions or diagnostics that can precisely identify and distinguish shocks from drifts in complex climate models. Therefore, these phenomena are here collectively referred to as *transient behaviors* of the initialized prediction.

The purpose of this paper is to describe a community resource developed to enable intercomparative studies across a range of prediction systems of transient behavior, which bears on several crucial aspects of climate prediction. One such aspect is that systematic errors associated with model transients must be corrected in order to reduce errors in predictions. These corrections are applied *a posteriori* to the raw outputs from the models and range from a simple correction of the mean error to more sophisticated corrections that take into account other properties of the distributions or the joint distribution of variables so as to maintain physical consistency. Bias reduction is typically accomplished by using retrospective forecasts or “hindcasts” (e.g., Arribas et al., 2011; Choudhury et al., 2017; García-Serrano & Doblas-Reyes, 2012; Kharin et al., 2012; Saha et al., 2014) to evaluate systematic errors as a function of the lead time of the prediction, which are robustly estimated for typical ensemble sizes (Manzanas, 2020). Such ensemble hindcasts provide the basis for the data set described here.

Another important aspect of transient behavior in initialized predictions relates to strategies for avoiding it. The method used to initialize dynamical forecasts is clearly important, with the full-field approach usually leading to larger shocks and drifts than anomaly initialization (e.g., Hazeleger et al., 2013; Kim et al., 2012; Smith et al., 2013; Volpi et al., 2017). Additional initialization approaches have been developed that aim to minimize transients. These include reduced-dimension implementations of four-dimensional variational data assimilation (He et al., 2017, 2020) and anomaly initialization (Polkova et al., 2019), anomaly initialization taking into account biased model responses to external forcings (Chikamoto et al., 2019) and various formulations of coupled data assimilation (e.g., Mulholland et al., 2015). An alternative approach is to adjust the forecast model runs by introducing additive bias correction terms in the model equations, either as two-dimensional flux corrections at the atmosphere-ocean interface (Magnusson et al., 2013; Polkova et al., 2014), or as three-dimensional corrections derived from analysis increments for atmospheric state variables (Chang et al., 2019; Kharin & Scinocca, 2012).

Analysis of the physical mechanisms associated with model transients is motivated by their potential to provide insights for improving initialization procedures, as well as clues to the origins of model systematic errors and hence guidance for reducing them. Such analyses have focused on a range of time scales, processes, and locations. The Transpose-AMIP II experiment examined the development of systematic errors in the atmospheric components of several climate models on weather prediction time scales out to five days. This type of approach has been used to elucidate the development of shortwave radiation biases over the Southern Ocean (Williams et al., 2013), marine stratocumulus clouds (Brient et al., 2019), and the Indian Summer Monsoon (Keane et al., 2019). On subseasonal time scales, Shonk et al. (2018) examined the causal sequence leading to a spurious northward drift of the Pacific Intertropical Convergence Zone during the first 60 days of ECMWF System 4 hindcasts, and Shonk et al. (2019) applied a similar approach to explain the development of model biases in the tropical Atlantic.

At longer seasonal to decadal time scales, studies have largely focused on the development of sea surface temperature (SST) biases caused by model misrepresentations of the Atlantic Meridional Overturning Circulation and other ocean circulations (Huang et al., 2015; Wang et al., 2015), oceanic heat uptake (e.g., Hawkins et al., 2014), wind stress (Voldoire et al., 2019; Woelfle et al., 2018), meridional heat

transport (Sánchez-Gómez et al., 2016), and combinations of atmospheric and oceanic processes (Exarchou et al., 2018; Siongco et al., 2020; Vannière et al., 2013). Once developed, the maintenance of SST biases is often attributable to ocean-atmosphere coupling (e.g., Toniazzo & Woolnough, 2014).

Additional studies have related biases that develop following initialization to those in long-term simulations unconstrained by initialization (Hermanson et al., 2018; Ma et al., 2014, 2020) and inconsistencies between sea ice and ocean initial conditions (Cruz-García et al., 2021).

A further approach to identifying sources of model error leading to transient behavior and long-term biases is to assess systematic tendencies in model equations integrated from observationally constrained states. This can be accomplished either by averaging initial tendencies near the start of a set of short-term forecasts (Rodwell & Palmer, 2007) or by estimating systematic tendencies from averages of analysis increments when data assimilation is applied (Batté & Doblas-Reyes, 2015; Bhargava et al., 2018; Crawford et al., 2020). Because models are not allowed to evolve far from observational states, these methods primarily highlight errors attributable locally to model parameterizations of “fast physics.” Such an approach can be extended to diagnose nonlocal, longer time scale influences of model error by applying tendency bias corrections as forcing terms in specified regions. For example, Schubert et al. (2019) used this method to quantify the relative roles of *remote* (teleconnection-driven) versus *local* influences on the total bias.

Most efforts thus far to describe, diagnose, and reduce systematic errors leading to transient behavior in subseasonal to decadal predictions have focused on individual prediction systems and have thus generated conclusions that, while informative, are specific to the models considered. By contrast, major efforts to intercompare the behavior of climate simulations (Flato et al., 2013) and predictions (Tompkins et al., 2017) across diverse sets of models have led to broader perspectives. In a similar spirit, this paper introduces a recently developed data set that aims to facilitate multi-model intercomparisons of transient behavior across numerous subseasonal, seasonal, and decadal prediction systems, leading to a better understanding of the transient behavior of forecast models and the mechanisms involved in the origin and growth of such errors. An overview of this data set is provided in Section 2, and several intercomparisons of transient behavior in subseasonal, seasonal, and decadal prediction models derived from it are presented in Section 3. Section 4 contains concluding remarks and recommendations for enabling further progress in characterizing and understanding physical mechanisms of transient behavior in climate predictions.

2. The Long-Range Forecast Transient Intercomparison Project Data Set

Aiming at increasing and improving our knowledge of the physical processes related to the origin and growth of initial transient behavior (i.e., shocks and drifts, leading ultimately to biases), the World Climate Research Programme (WCRP)'s Working Group on Subseasonal to Interdecadal Prediction (WGSIP) has led the Long-Range Forecast Transient Intercomparison Project, LRFTIP. Its purpose is to facilitate multi-model intercomparison studies of the transient behavior of coupled long-range forecast models evolving from observation-based initial conditions, the term “long-range forecast” standing for predictions from 30 days up to years according to the World Meteorological Organization definition. The main goals of LRFTIP can be summarized as follows:

1. Developing a publicly available multi-model online archive of hindcast climatologies and related diagnostics from systems contributing to projects such as the Subseasonal to Seasonal Prediction Project (S2S; Vitart et al., 2012), the Climate-system Historical Forecasting Project (CHFP; Tompkins et al., 2017), and the Decadal Climate Prediction Project (DCPP; Boer et al., 2016).
2. Consolidating a standard set of diagnostics in order to consistently characterize transient behavior in initialized prediction models.
3. Addressing scientific questions, including the influence of different initialization methods on transient behavior of climate model components and the identification of any impacts on forecast quality.
4. Including hindcast initialization experiments, using the same model but different initialization methods, in contributions to items 1–3.

Table 1
Current List of Model Names and Number of Atmospheric (ATM) and Oceanic (OCE) Variables With Daily Data From Subseasonal Hindcasts

Model	ATM daily	OCE daily
ECCC-S2S	7–9	1–1
ECMWF-S2S	7–11	1–1
JMA-S2S	2–2	1–1
Météo-France S2S	7–11	1–1
NCEP-S2S	7–11	1–1
UKMO-S2S	4–6	1–1

Note. The number of variables in the reference or analysis data set (in the hindcast climatology) are indicated in first place (second place).

tion as has been shown for the Atlantic Meridional Overturning Circulation (Bilbao et al., 2021) or the Niño 3.4 index (Teng et al., 2017). This problem has been illustrated for globally averaged temperature for the Decadal Prediction Large Ensemble (DPLE) with the Community Earth System Model version 1 (CESM1) whereby the initialized hindcasts drift toward the uninitialized model climate, but drifts early in the hindcast period are larger than later in the period (Meehl et al., 2021, Figure 4). Thus, when removing bias by subtracting the drifted model climatology from individual drifted hindcasts, errors will be introduced

LRFTIP has developed a data set that is expected to be useful to the scientific community in order to address such goals. It encompasses a large set of subseasonal to decadal hindcasts that have been processed in the following manner: For a particular model and start date, hindcasts are averaged over available years and ensemble members to form a hindcast climatology that is a function of spatial variables and lead time. When available, climatologies for the same model are also constructed for the free-running model (ideally, from CMIP historical simulations, averaging over multiple ensemble members) and for hindcast initial conditions represented by assimilating model runs, analyses used for initialization or other observation-based reference states. The objective of the data set is to facilitate studies aimed at better characterizing and intercomparing systematic transient evolution away from the initial conditions of a prediction model toward the model's own climate. Notwithstanding, drifts are typically nonstationary, particularly at longer time scales, meaning that they could be shaped by the prevailing conditions during initialization

particularly for hindcasts early in the period and late in the hindcast period. This poses a challenge for interpreting hindcast skill measures, and studies on interannual or decadal time scales making use of the data set described here might need to take this issue into consideration.

Table 2
List of Model Names and Number of Atmospheric (ATM), Oceanic (OCE), Land (LND) and Sea-Ice (SIC) Variables With Daily (dly) and Monthly (mly) Data From Seasonal Hindcasts

Model	ATM dly	ATM mly	OCE mly	LND mly	SIC mly
CanCM3	18–22	22–22	12–12	0–4	1–2
CanCM4	18–22	22–22	12–12	0–4	1–2
CanCM4i	21–26	22–26	11–13	2–3 ^a	1–2 ^b
ECMWF-S4	-	18–20	-	0–1	-
GEM-NEMO	-	12–12	1–1	1–1	1–1
JMAMRI-CGCM1	12–16	20–20	6–7	-	-
JMAMRI-CGCM2	-	19–21	6–6	-	-
MIROC5_v1.0	12–18	16–18	6–7	-	-
MPI-ESM-LR	-	20–22	-	-	-
POAMA_p24	-	12–14	-	-	-
ARPEGE	-	7–7	-	-	-
CFS_SHFP	3–3	8–8	-	-	-
CMAM	3–5	7–7	-	-	-
GloSea4	0–1	7–7	-	-	-
GloSea5	0–1	7–7	-	-	-
ENSEMBLES ^c	16–20	20–20	-	0–1	-

Note. The number of variables in the reference or analysis data set (in the seasonal hindcast climatology) are indicated in first place (second place). Note that ENSEMBLES is not a model itself but a set of 5 models.

^aDaily values also available. ^bDaily values also available for hindcast climatology only. ^c5 models as detailed in Doblas-Reyes et al. (2010).

The LRFTIP data archive covers three different time scales: *subseasonal*, with daily data out to 30/60 days; *seasonal*, with daily data out to 30/60 days plus monthly means through the forecast range (3–6 months); and *decadal*, which includes daily, monthly, and annual means, as available. Files within the data set are CF-compliant NetCDF4 and variable naming follows the CMIP/ESGF conventions. The archive can be accessed at <http://crd-data-donnees-rcd.ec.gc.ca/CCCMA/products/LRFTIP/>.

As of July 2021, the data set includes hindcast climatologies of 6 subseasonal forecast models from the S2S project, data archives from 21 seasonal forecast models from CHFP and ENSEMBLES (Doblas-Reyes et al., 2010), and output from 16 decadal model predictions from CMIP5, CMIP6, and ENSEMBLES. Detailed information on the model names, numbers of atmospheric, oceanic, land and sea ice variables, and temporal frequencies available can be found in Tables 1–3 for subseasonal, seasonal, and decadal hindcasts, respectively.

A wide range of scientific questions can be addressed with the LRFTIP data set. These potentially involve characterizing and better understanding the sources of initial shock, describing the evolution of drifts toward the preferred states of the models and assessing the relative roles of the oceans and the atmosphere in the maintenance of the long-term biases. Examples of such topics are presented in the next section following a discussion of different methods that may be used to quantify transient behavior. These topics include an analysis of initial shocks in subseasonal predictions of total cloud cover (TCC), SST, and precipitation (PR), followed by drifts in seasonal predictions of SST and PR, and finishing with the evolution of SST errors in decadal forecasts.

Table 3

List of Model Names and Number of Atmospheric (ATM), Oceanic (OCE), Land (LND) and Sea-Ice (SIC) Variables With Daily (dly), Monthly (mly) and Yearly (yly) Data From Decadal Hindcasts

Model	ATM dly	ATM mly/yly	OCE mly/yly	LND mly/ yly	SIC mly/ yly
CCSM4	-	0-24-24	0-8-9	0-3-3	0-2-2
MF-ENSEMBLES	-	0-20-0	0-11-0	0-1-0	-
CFSv2	-	0-26-0	0-7-0	0-1-0	0-2-0
CanCM4	25-16-6	25-26-17	13-13-13	4-3-2	2-2-2
CanESM5	21-21-21	26-26-26	11-11-11	3-3-3	2-2-2
CNRM-CM5	0-6-6	0-26-26	0-10-10	0-3-3	0-2-2
ECMWF-ENSEMBLES	-	0-20-0	0-11-0	0-1-0	-
GFDL-CM2p1	-	0-18-18	0-11-11	0-2-2	0-2-2
HadCM3	0-9-9	0-25-25	0-6-6	0-3-3	0-2-2
IFM-ENSEMBLES	-	0-20-0	0-11-0	0-1-0	-
MIROC5	0-20-20	0-26-26	0-7-0	0-3-3	0-2-2
MRI-CGCM3	0-6-0	0-26-0	0-10-0	0-3-0	0-2-0
UKMO-DePreSys	-	0-20-0	-	0-1-0	-
UKMO-HadGEM2	-	0-20-0	-	0-1-0	-
EC-Earth	-	0-19-17	0-7-3	-	0-2-2
BCC-CSM1.1	0-6-6	0-26-26	0-7-7	-	0-2-2

Note. The number of variables in the reference or analysis data set, in the decadal hindcast climatology, and in the historical run climatology are indicated in first, second, and third place, respectively.

3. Identifying Initial Shocks, Drifts and Biases in Subseasonal to Decadal Predictions

3.1. Quantifying Transient Behavior

The purpose of the LRFTIP data set is to enable quantification and intercomparisons of transient behavior in initialized climate predictions. However, there is no unique way to characterize such transient evolution. This nonuniqueness is illustrated in Figure 1a, which schematically depicts model transient behavior for a particular variable as a function of lead time t_L . At the start of the prediction, the model climatology (MOD) coincides with that of the initial conditions (IC), which may depart from observed climatology (OBS) if anomaly initialization is used, or if initial states are otherwise not strongly constrained by observations. MOD subsequently evolves away from IC, and for large t_L asymptotically approaches the climatology of the free-running model represented here by that of uninitialized historical simulations (HIST) subject to the same external forcings as the predictions. It should be noted that for simplicity, this diagram omits complicating factors such as time dependences of climatologies other than MOD, the possibility that MOD does not monotonically approach HIST (e.g., Hermanson et al., 2018) and so forth.

For scalar quantities such as global mean temperature, it may be straightforward to depict transient evolution in plots such as Figure 1a. However, for spatial fields it is more convenient to depict transient evolution as maps representing differences from some reference climatology. Which differences are shown may depend on what aspect of transient behavior the analysis is attempting to capture, and what reference data are available. The standard practice of characterizing model errors in terms of climatological differences from observations (dependent in this case on t_L) may be appropriate if IC is close to OBS and the necessary observational data are available, noting OBS should be constructed using the same

temporal sampling as MOD. We label such MOD minus OBS differences, that is, the evolving model bias, as DIFF-OBS (Figure 1b).

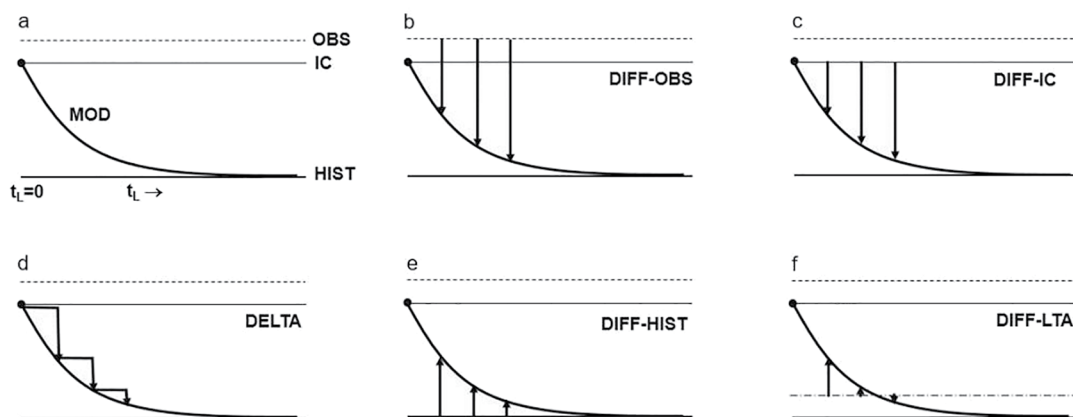


Figure 1. (a) Schematic depiction of systematic transient behavior in initialized climate predictions. The vertical axis represents values for some variable as a function of increasing lead time t_L (horizontal axis). Model climatology (MOD) is the evolving hindcast climatology as a function of t_L , whereas observed climatology (OBS), initial conditions (IC), and historical simulations (HIST) represent corresponding climatologies of observations, initial conditions, and uninitialized historical simulations, respectively, in which seasonal or other variations are not shown for simplicity; (b) differences between MOD and OBS, corresponding to the t_L -dependent model bias; (c) differences between MOD and IC; (d) successive differences between MOD at different t_L ; (e) differences between MOD and HIST; and (f) differences between MOD and its lead time average LTA (dot-dashed).

If IC is not close to OBS or OBS is not available, then it may be desirable to consider the evolving departure from initial conditions, that is, MOD minus IC, labeled as DIFF-IC in Figure 1c. However, this requires the availability of IC, which is not always the case for hindcast data sets even though some, such as the DCP6 CMIP6 archive, do request data reflecting model initial states (Boer et al., 2016). An alternative way to quantify transient evolution particularly if OBS and/or IC are not available is to consider differences in MOD at successive values of t_L (Figure 1d). Although this method has the limitation that such differences cannot be separated from changes rooted in the seasonal cycle, it can be useful for examining large transients during the first few days of hindcasts when daily or higher-frequency data are available and is labeled here as DELTA.

Alternatively, when long temporal ranges are considered as for decadal predictions, the approach to asymptotic model climatology, that is, MOD minus HIST, may be of greater interest than the departure from OBS or IC. This measure, labeled DIFF-HIST in Figure 1e, requires the availability of ensembles of uninitialized simulations that are congruent with the initialized hindcasts, as is the case for some CMIP5 and CMIP6 decadal prediction systems. In instances where these simulations are unavailable or such analysis is otherwise inconvenient, differencing from the MOD lead-time average (LTA) over the entire hindcast range may be approximately representative of convergence to HIST if maximum lead time is sufficiently long that LTA is much closer to HIST than to IC. This approach is labeled DIFF-LTA and is represented in Figure 1f.

Each diagnostic illustrated in Figures 1b–1f can be constructed in two ways. The first, referred to here as Method 1, is to consider t_L increasing from a common initial time t_i , often the first day of a particular calendar month or sometimes a fixed set of start dates (e.g., Mulholland et al., 2016). In this case, the target interval $t_T \rightarrow t_T + \Delta t_T$ where Δt_T is typically one week, month, or season evolves forward with t_L according to $t_T = t_i + t_L$. Although this approach is common, one disadvantage is that because t_T changes with t_L , simply tracking the evolution of MOD with t_L as with the DELTA approach in Figure 1d does not separate transient evolution from the climatological seasonal cycle. As a result, the DELTA approach is only useful when initialization transients greatly exceed seasonal changes as may occur for very short lead times of a few days or if a particular month or season is considered in consecutive years of a multi-year prediction. In the second way of constructing these diagnostics, referred to here as Method 2, the target month (or other averaging period) is held fixed and t_L is varied by adjusting t_i . For example, t_L can be advanced N months by moving the initialization date N months earlier, as in Figure 1 of Manzananas (2020). An advantage is that transient behavior as a function of t_L does not have to be separated from the climatological seasonal cycle. However, a disadvantage is that to examine drift at intervals of Δt_L requires hindcasts that are initialized at the same interval. This precludes, for example, the consideration of monthly Δt_L in decadal hindcasts initialized once per year or daily Δt_L in seasonal hindcasts initialized once per month.

The examples described below sample some of the measures of transient behavior depicted in Figure 1, in all cases using Method 1.

3.2. Initial Transients in Subseasonal Predictions of TCC, SST and PR

We first focus on the origin of initial shocks in subseasonal hindcasts of TCC from the S2S project. The analysis of this variable is of particular interest in this context, given that it has a large dependence on model physics. The ERA-Interim reanalysis (Dee et al., 2011) is used as a proxy of observations against which the predictions are compared as a function of lead time. Subseasonal hindcasts within the LRFTIP data set span prediction ranges of 30–60 days.

Figure 2 shows the zonal mean TCC in hindcasts initialized in May for the first 7 forecast days. Note that ERA-Interim shows very little day-to-day variation in the zonal-mean TCC, so only day 1 is shown for reference. A noticeable feature from this figure is the difference in Météo-France-S2S and UKMO-S2S (Figures 2d and 2f) at day 1 of the prediction compared to the following days, which is mostly absent in the other models. In the case of UKMO-S2S, this singularity at the beginning of the prediction quickly vanishes at day 2 and little further adjustment occurs after that lead time. Likewise, Météo-France-S2S depicts a clear difference between day 1 and the following days in TCC over the Southern Hemisphere, but over the Northern Hemisphere (particularly north of 45°N) the sudden jump in zonal mean TCC from day 1 to day 2 is followed by a gradual trend back toward the day 1 pattern afterward. These differences arise from IC

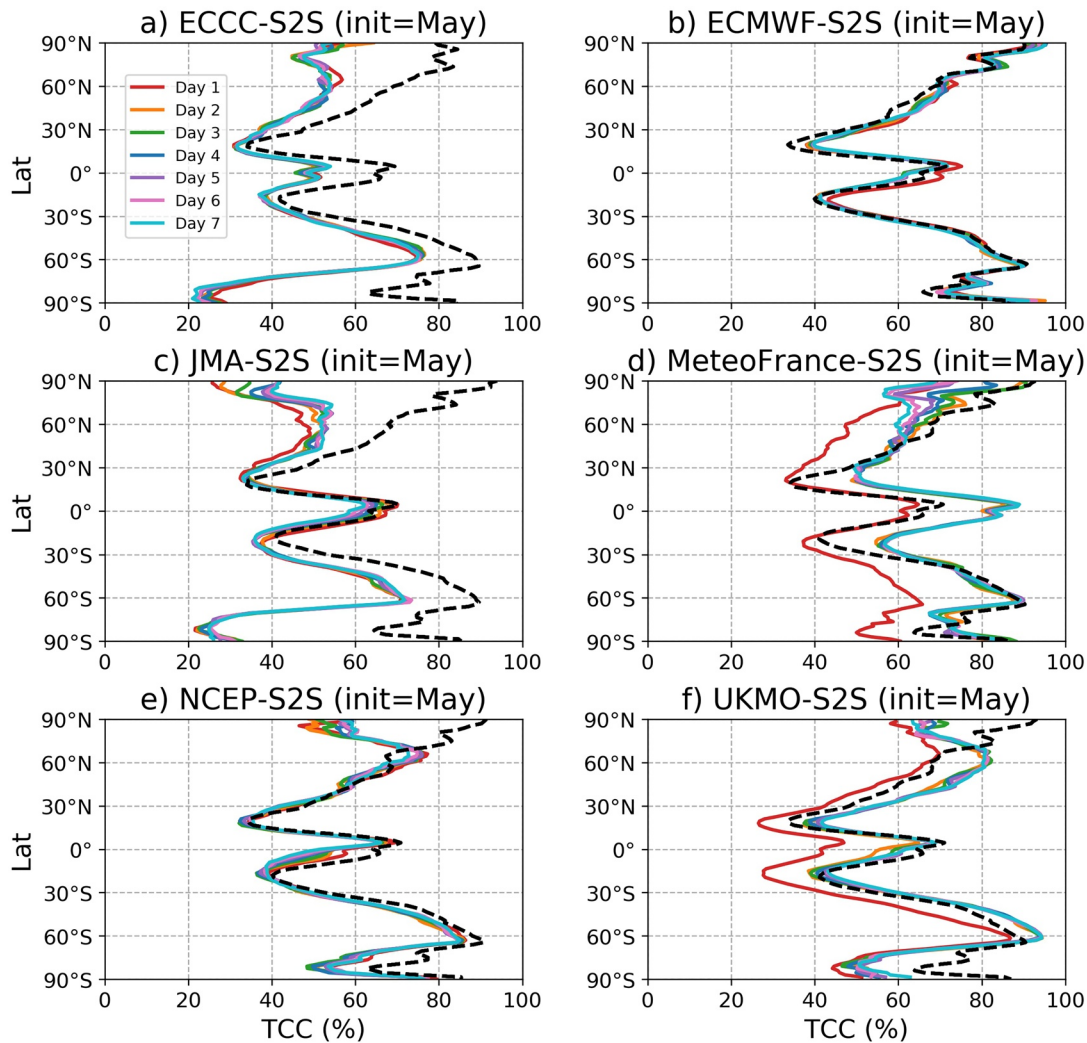


Figure 2. Zonal mean of total cloud cover (TCC) in sub-seasonal hindcasts initialized in May for each S2S prediction model on days 1–7 of the predictions (colors) and their respective ERA-Interim reference data sets for day 1 (dashed black line). Values in %.

being ERA-Interim in the case of UKMO-S2S and Météo-France-S2S rather than from their own analyses. In addition, ECCC-S2S and JMA-S2S show TCC that is much lower than the ERA-Interim at day 1 at high northern and southern latitudes, but with little evolution between day 1 and day 7, except at high northern latitudes in JMA-S2S. These models are both uncoupled and have physical formulations that are motivated primarily by accurately predicting atmospheric circulation out to two weeks or so rather than simulating a stable climate.

The zonal-mean patterns in ECMWF-S2S (Figure 2b) and NCEP-S2S (Figure 2e; excluding the North Pole) are very similar to those derived from the ERA-Interim, even though larger differences are found in the other forecast systems. Still, the initialization shock is largely gone by day 2 in all the models except for the mid- to high-latitudes of the Northern Hemisphere where some continued evolution is evident. Differences between successive days of the subseasonal predictions, that is, the DELTA metric depicted in Figure 1d, are shown in Figure 3. In line with the previous results, differences between days 1 and 2 in Météo-France-S2S and UKMO-S2S are much larger than in the other forecast systems, with both models exhibiting an increase of about 10%–20% in mean TCC (Figures 3d, 3f and 3g). The evolution of day $i+1$ minus day i differences in global-mean (Figure 3g) as well as 60°N–90°N averaged (Figure 3h) TCC largely ceases after day 2, once again highlighting the initial shock effect on the predictions. Regarding the models that do not display such a clear initialization shock, it is worth noting that there are indeed differences in TCC between the first and

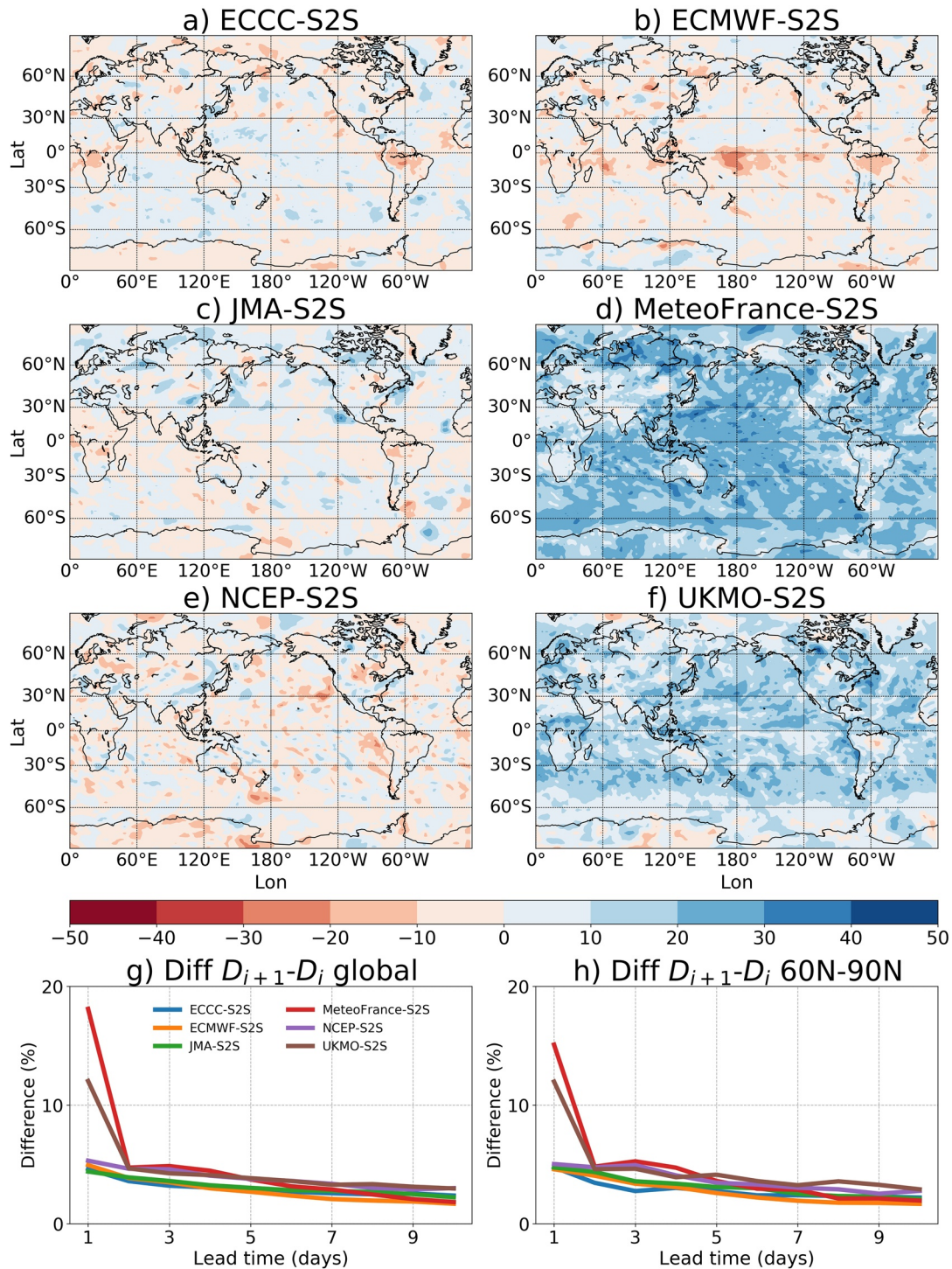


Figure 3. (a–f) Differences between days 1 and 2 (day 2 minus day 1) in subseasonal predictions of total cloud cover (TCC) initialized in May. (g–h) Evolution of the absolute difference in global mean TCC for each forecast system (colors) between two subsequent days, from day 1 to day 10, for (g) the global mean and (h) the region between 60°N and 90°N. Values in %.

second days of the predictions but such differences are not uniform over the globe (Figures 3a–3c, and 3e). ECMWF-S2S is the only model displaying a systematic underestimation of TCC over the Tropical Pacific in day 1 of the prediction (Figure 3b), followed by a drop toward a more stable condition that lasts for the rest of the forecast (Figure 2b).

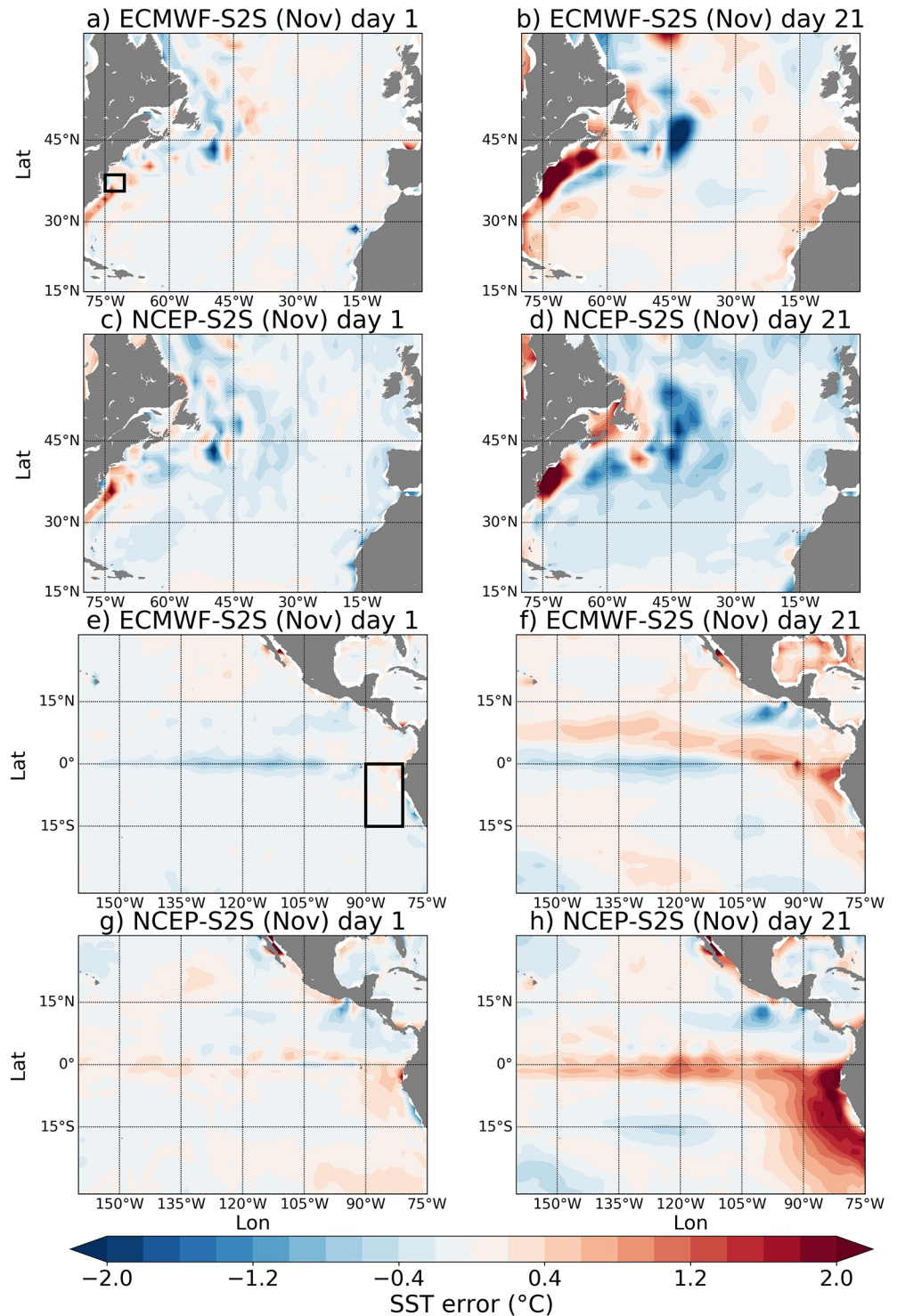


Figure 4. Differences in subseasonal predictions of sea surface temperature (SST) compared to corresponding reference data sets (OISSTv2) for days 1 and 21 (columns) of predictions initialized near Nov 1. Panels (a–d) highlight transient behavior in the North Atlantic and (e–h) in the central-eastern tropical Pacific. Boxes in panels (a and e) indicate the Mid-Atlantic Bight and southeast tropical Pacific regions considered in Figure 5.

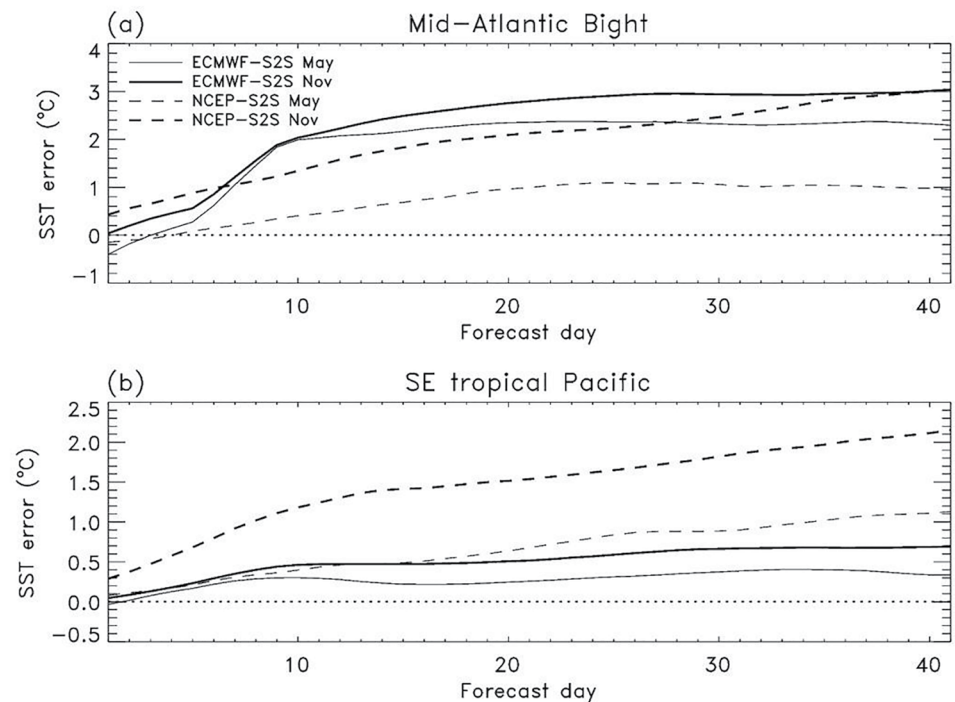


Figure 5. Development between hindcast days 1 and 41 of sea surface temperature (SST) biases averaged in the (a) Mid-Atlantic Bight and (b) central-east tropical Pacific regions indicated in Figure 4, in subseasonal hindcasts from the ECMWF- (solid) and NCEP- (dashed) S2S systems initialized near the beginning of May (thin) and November (thick). Differences are relative to the corresponding OISSTv2 reference data.

Initial transient evolution of SST in key regions where models tend to exhibit large SST biases is illustrated for the ECMWF- and NCEP-S2S models in Figure 4, where the Optimal Interpolation Sea Surface Temperature version 2 data set (OISSTv2; Reynolds et al., 2002) is employed as the observational reference to compute DIFF-OBS (Although only two models are considered here due to space limitations, the LRFTIP data set enables this analysis easily to be extended to the additional S2S models in Table 1). Biases on days 1 and 21 of hindcasts initialized near the start of November are shown for the North Atlantic region in panels a–d and for the central and eastern tropical Pacific in panels e–h. In all cases, biases on day 1 tend to be relatively small as should be expected when SST is initialized realistically, given that SST adjusts considerably less rapidly than TCC. However, even on day 1 the biases, though modest, bear some imprints of the more highly developed biases on day 21, for example, off the US east coast and in the tropical Pacific. Note that the cold SST bias in ECMWF-S2S over the equatorial Pacific, which is similar in May to that shown for November in Figure 4e could explain its negative drift in TCC between day 1 and day 2 for May initialization (Figure 3b). By day 21, particularly large biases have appeared off the northeastern US coast, where many models exhibit unrealistically warm ocean temperatures due to failing to simulate the separation of the Gulf Stream from the coast at Cape Hatteras (e.g., Ezer, 2016). The development of this bias as a function of forecast time in hindcasts initialized near the start of November and May is illustrated in Figure 5a, showing daily values of DIFF-OBS averaged in the Mid-Atlantic Bight region (box in Figure 4a), which is especially affected by the misrepresentation of the Gulf Stream separation. The warm SST bias in this region is seen to develop more rapidly in the ECMWF hindcasts, having substantially reached its eventual magnitude at $\approx 2^\circ\text{C}$ already by day 10 for hindcasts initialized in May. By contrast, the bias continues to develop more gradually in November-initialized NCEP hindcasts until at least day 41, when it has become comparable to the ECMWF bias at $\approx 3^\circ\text{C}$. The NCEP bias also shows greater seasonality, reaching only $\approx 1^\circ\text{C}$ by day 41 in hindcasts initialized in May, whereas the May-initialized ECMWF bias at day 41 approaches that from November initialization. In the southeast tropical Pacific region indicated in Figure 4e, the large warm bias in the November-initialized NCEP hindcasts develops steadily throughout the first 41 days, whereas the bias is considerably weaker in May-initialized NCEP and ECMWF in both seasons (Figure 5b). Pervasive warm SST biases in

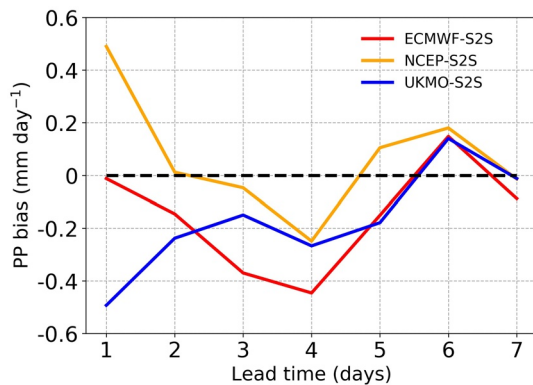


Figure 6. Precipitation biases for the North Atlantic region (70°W–10°E, 45°N–70°N) in subseasonal hindcasts initialized in November for three S2S prediction models on days 1–7. Units are mm day⁻¹.

this and other eastern tropical ocean regions have been ascribed to the underrepresentation of stratocumulus clouds and coastal upwelling in such regions (e.g., Richter, 2015).

Another example of transient behavior is demonstrated with daily PR over the North Atlantic (defined here as 70°W–10°E, 45°N–70°N) for subseasonal hindcasts initialized near the start of November. Figure 6 presents deviations from GPCPdd v1.3 daily observation data (a part of GPCP2.3 project; see Adler et al., 2018) for area-averaged PR in ECMWF-S2S, NCEP-S2S, and UKMO-S2S for the first 7 forecast days. The observed values for this region during this period vary between 3.7 and 4.4 mm day⁻¹, although it can be seen that the three models depict different behaviors. NCEP-S2S shows a positive bias at day 1 while UKMO-S2S shows the opposite behavior, and ECMWF-S2S yields no bias. The negative bias in UKMO-S2S at day 1 is consistent with the similar deficit of TCC in this forecast system, suggesting that the first-day biases in these two variables could be connected. Also worth noting is that the ECMWF-S2S system did not apply coupled atmosphere-ocean sea-ice data assimilation at the time of the LRFTIP data set collection, which could have an impact on

the initial shock in the atmospheric model (e.g., Browne et al., 2019). Beyond day 3 of the predictions, it is interesting to note that biases in the forecast systems tend to follow a similar evolution pattern.

3.3. Bias and Drift in Seasonal Predictions of PR and SST

Seasonal predictions included in the LRFTIP data set extend up to 12 months into the future and have been initialized at least two times per year, in May and in November. These months are close to the beginning of summer in the Northern and Southern hemispheres, respectively, and near the onset of the rainy season of monsoon circulations over several land areas of each hemisphere. Indeed, seasonal forecasts started in May are commonly used to predict the behavior of the Southeast Asia or the Indian monsoons, while the start date in November can provide useful information on the likely evolution of the South American monsoon or on the strength and location of the South Pacific Convergence Zone. The following analysis on the evolution of the biases in seasonal predictions of PR from 18 seasonal forecast systems in the CHFP archive is tackled by (at least) two target issues: the drift of the models toward their own climate and the season-dependent evolution of biases related to problems with the representation of the monsoon circulation and associated rainfall. Figure 7 displays the difference (prediction minus observational reference, i.e., DIFF-OBS as in Figure 1b) in seasonal hindcasts of PR for month 1 of the prediction initialized in May, while a similar analysis for month 3 of the prediction (July) is shown in Figure 8. The reference data set here is GPCPv2.3. The first outcome from the comparison of both figures is the growth of the differences between prediction months 1 and 3, which is more noticeable over the Northern Hemisphere, following the intraseasonal migration of the Intertropical Convergence Zone (ITCZ), suggesting that systematic problems exist regarding the simulation of its location and/or strength. Some models also show large differences over areas affected by monsoon rainfall such as JMA-MRI-CGCM2, which shows a strong negative bias over India in July (Figure 8l). Overall, the smallest mean absolute errors within the latitudinal belt 50°N–50°S at months 1 and 3 of the predictions are found for the two versions of the ECMWF and JMA models, with UKMO-DePreSys also performing well at month 3 (Figures 7g, 7h, 7k and 7l and Figures 8g, 8h, 8k 8l, and 8q). A similar analysis for predictions initialized in November

(Figures 9 and 10) shows that the largest differences are now displaced to the Southern Hemisphere, with most models attaining maxima in differences over the Maritime Continent, parts of the Indian Ocean, and over central and southeastern South America at month 3 of the predictions (i.e., January). There is a tendency for most models to overestimate PR over the tropical Indian Ocean and far western Pacific, whereas there is less consistency in the sign of the biases in the tropical Atlantic and central-eastern Pacific, in agreement with the analysis by Scaife et al. (2019) of PR forecasts in these regions from a similar set of models. The overall smallest differences are found for the same models as for May initialization, highlighting their capability to simulate the seasonal climatology of PR (i.e., the ITCZ). Open questions, which could be explored

Month 1 from May initialization

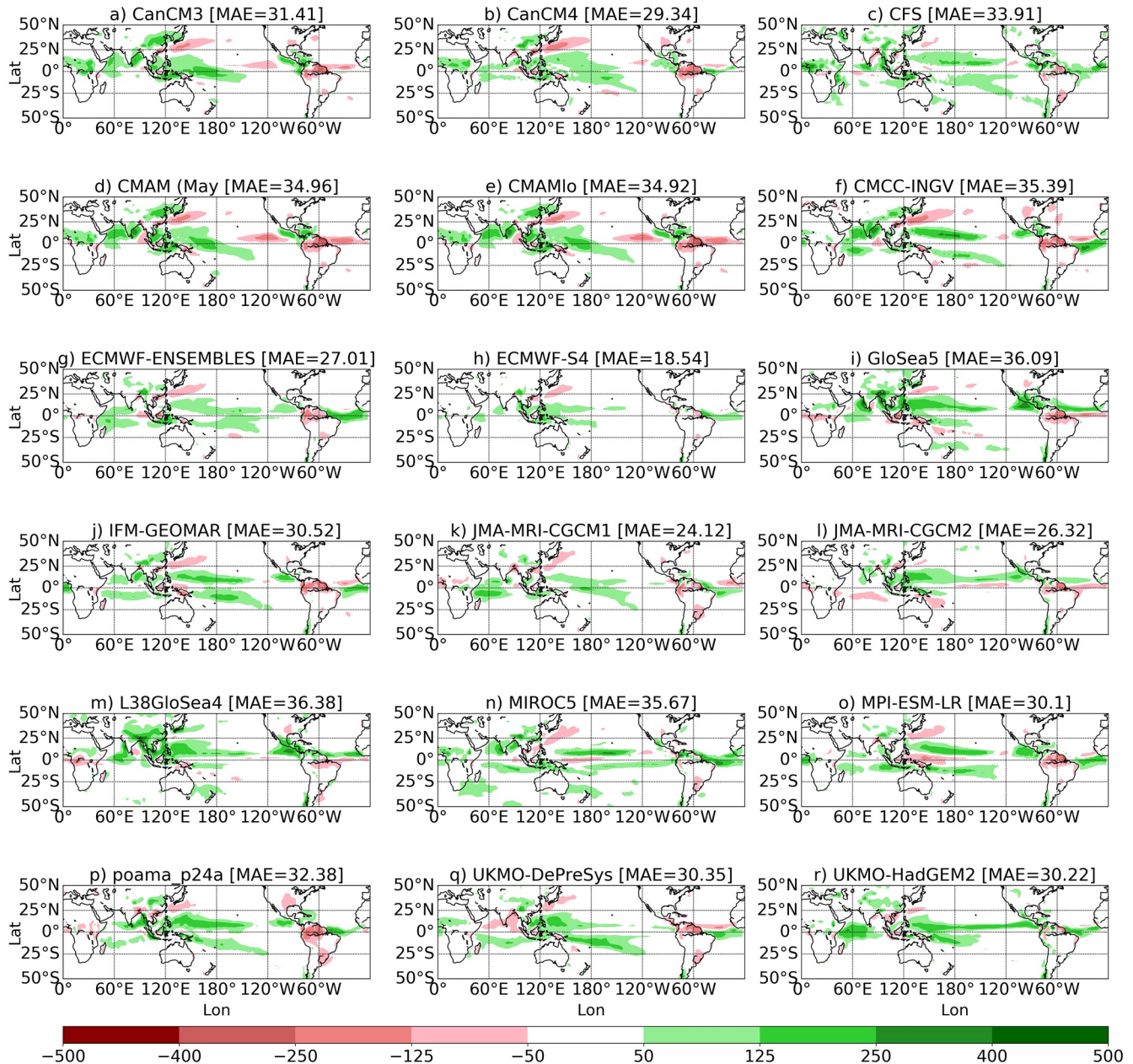


Figure 7. Differences in seasonal predictions of monthly mean precipitation compared to their corresponding reference data sets (GPCP2.3) for month 1 of predictions initialized in May. The areal-mean mean absolute error (MAE) is indicated in each subtitle. Units are mm month⁻¹.

in the future include whether the lead time-dependent difference in rainfall in the models is explained by the parameterization schemes used to simulate cloud microphysics and PR or by the differences in the dynamics.

Figure 11 shows the time evolution of the mean absolute error averaged over 50°N–50°S in PR from the different hindcast data sets for predictions initialized in May (Figures 11a) and November (Figures 11b). Beyond the point related to the smallest differences occurring in the ECMWF and JMA forecast systems as already discussed above (and which can also be seen here), most of the models depict a fairly stable pattern of differences from the first until the third prediction month. However, some of them display particular

Month 3 from May initialization

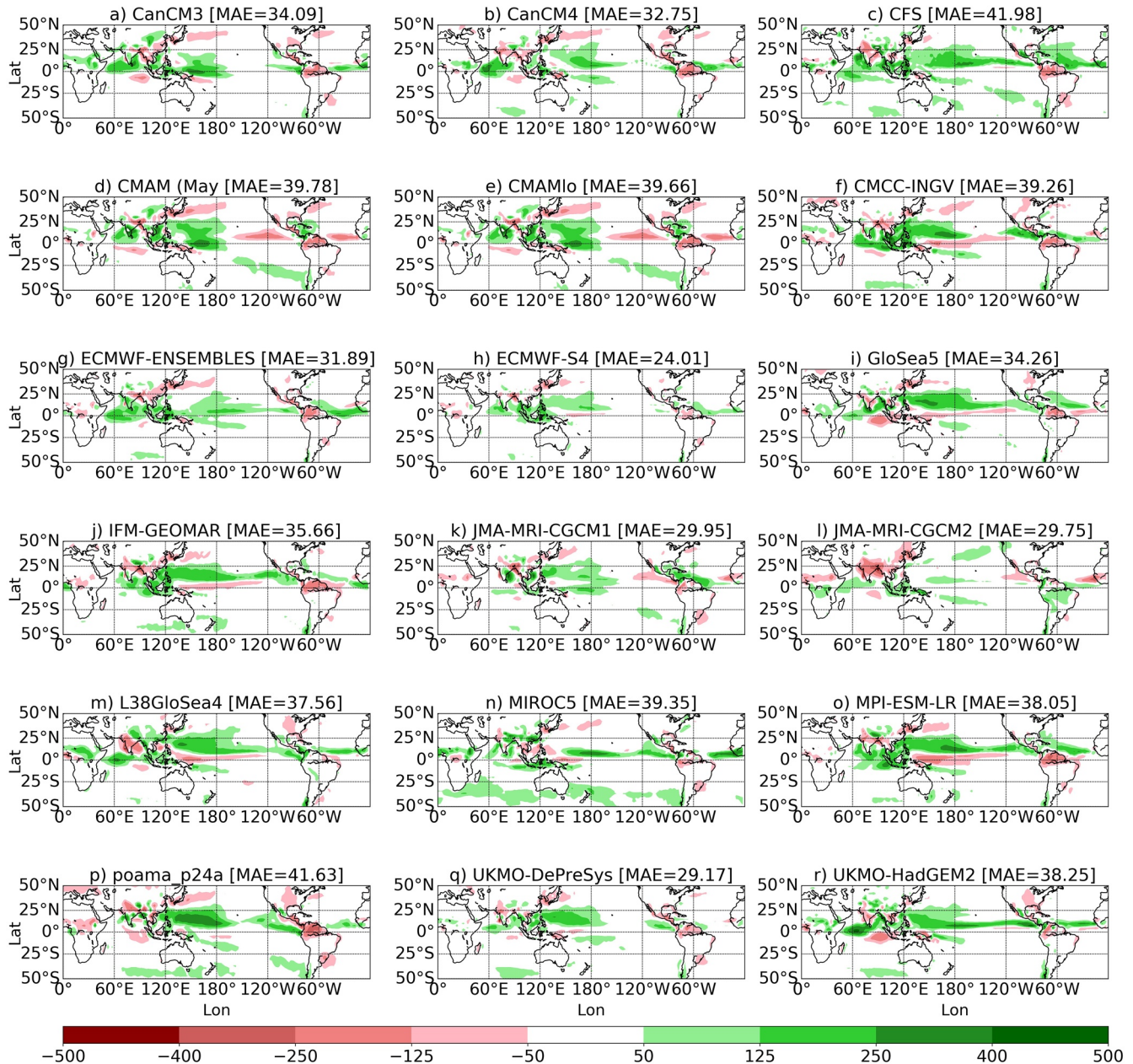


Figure 8. As in Figure 7, but for month 3 of the predictions.

patterns of variation: for instance, GloSea5 shows an increase in PR difference from month 1 to month 2, followed by a decrease in month 3, in hindcasts initialized in both May and November. Some others, such as L38GloSea4, show a maximum difference in month 1 of the November hindcasts, followed by declines in months 2 and 3. This could be related to large errors at the beginning of the simulation caused by initialization shock followed by a gradual drift toward smaller errors, but further analysis is needed to confirm that.

A similar analysis focused on SST predictions from 5 CHFP models is shown in Figures 12 and 13 for hindcasts initialized in May and November, respectively, and considering prediction months 1, 3, and 5. OISSTv2 is used as the reference data set in computing DIFF-OBS. As for PR, JMA-MRI-CGCM1 shows relatively small biases throughout the prediction, with little dependence on lead time. By contrast, other models are

Month 1 from November initialization

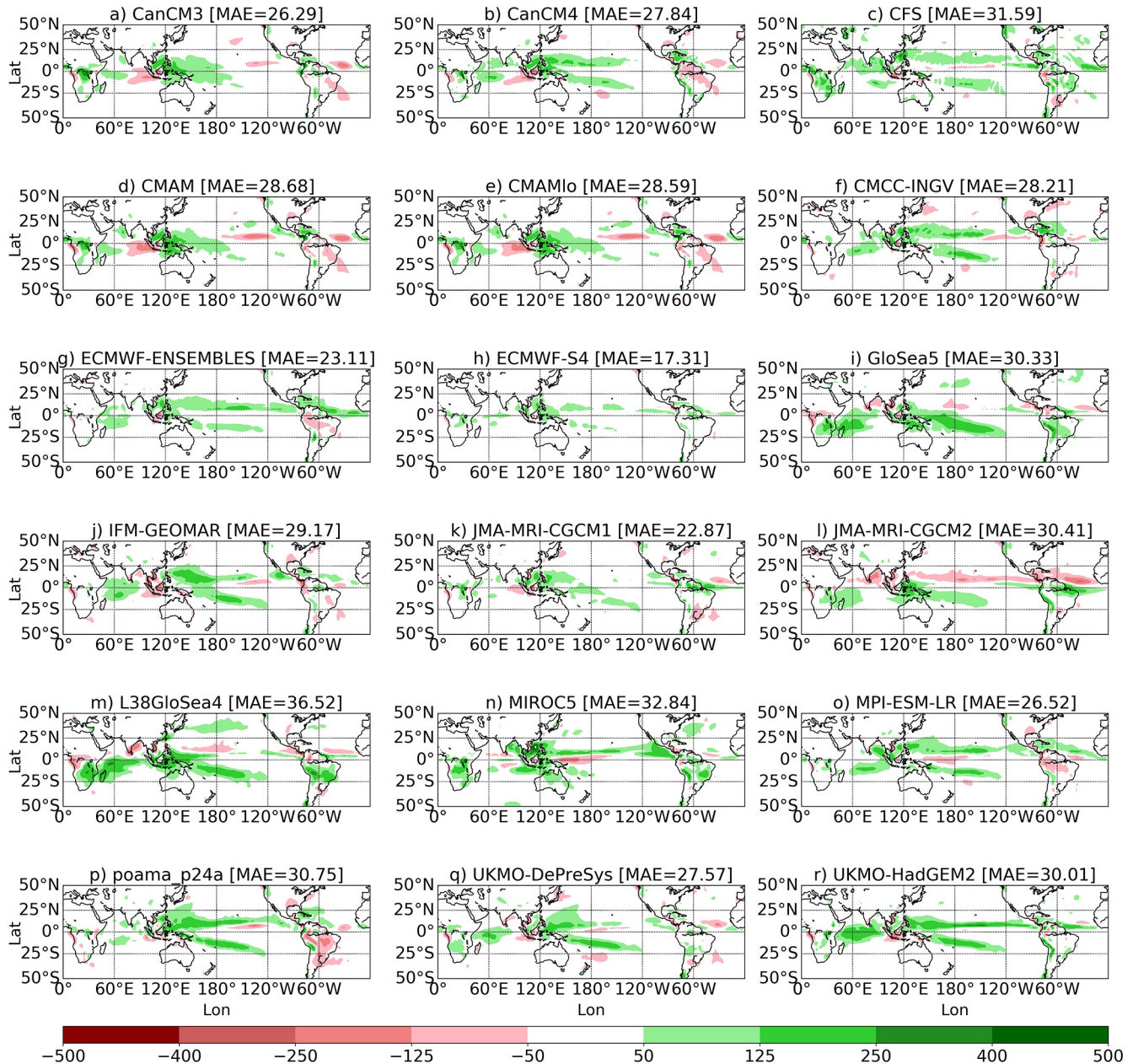


Figure 9. As in Figure 7 but for predictions initialized in November.

affected by substantial errors at each lead time (MIROC5), a decrease and reversal of sign of the biases after month 1 (JMA-MRI-CGCM2), or a notable increase in magnitude at longer lead times (CanCM3 and CanCM4). Comparisons between hindcasts initialized in May and in November show the seasonal dependency of biases in some forecast systems, as is the case of JMA-MRI-CGCM2. These results motivate the analysis of the physical mechanisms of these bias patterns in order to disentangle their causes.

Month 3 from November initialization

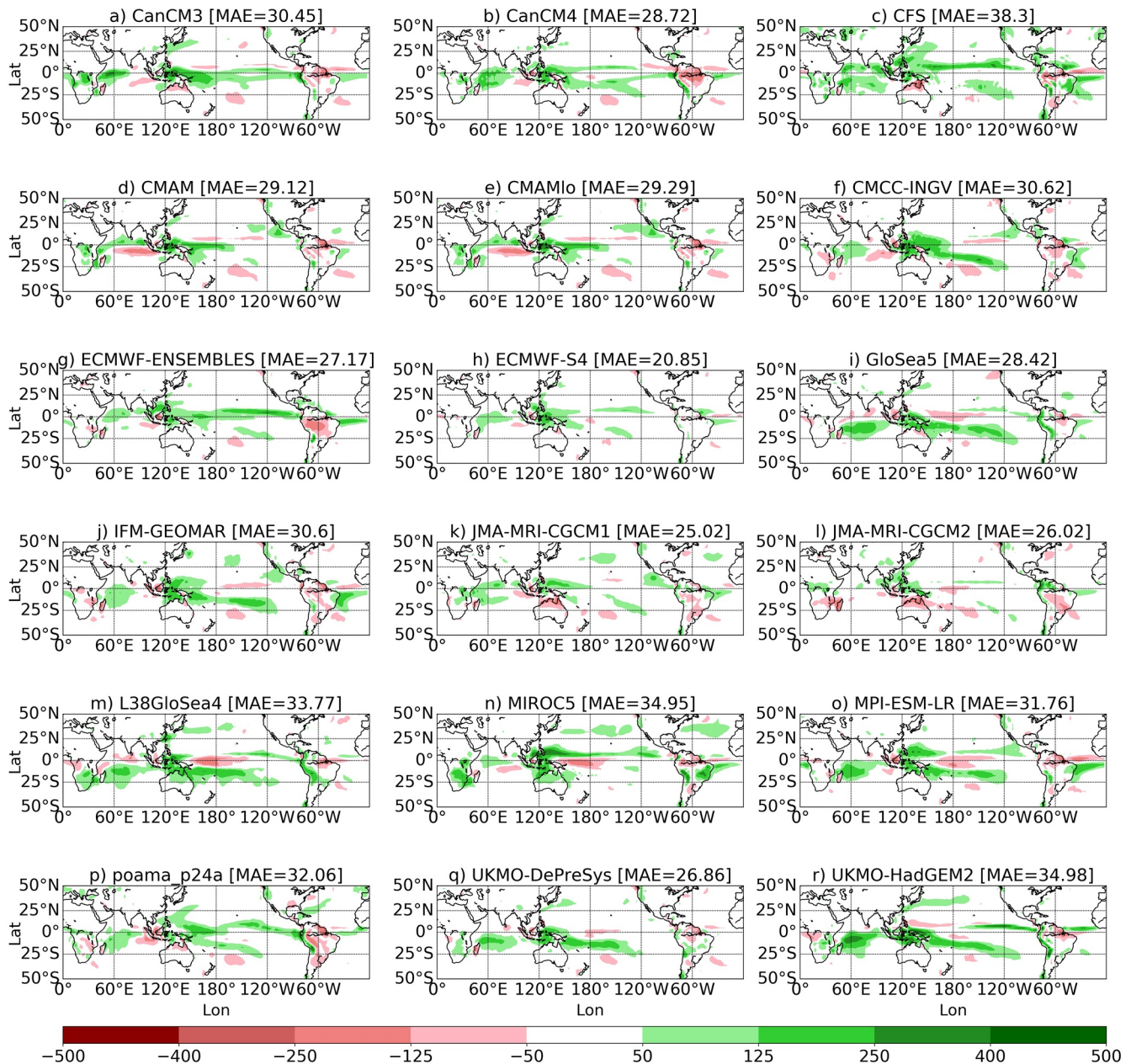


Figure 10. As in Figure 9 but for month 3 of the predictions.

3.4. Evolution of Transients in Decadal Prediction of SST

The LRFTIP data set also allows for multi-model intercomparison of transient behavior on multi-annual time scales. Figure 14 shows Hovmöller diagrams of the difference (hindcast minus historical, i.e., DIFF-HIST in Figure 1e) in monthly mean SST over the Equatorial Pacific (defined here as 120°E–90°W, 5°S–5°N) in various decadal forecast systems. In this case, the reference data set consists of CMIP historical (free-running) simulations for the same models, that is, without initialization. The differences thus represent departures from the free-running climatology of the models, which one may expect the hindcasts to evolve toward over the 10-year range of the forecasts. The Hovmöller diagram becomes useful for the type of analysis pursued in this article, since it can synthesize information on the fast response of the model at

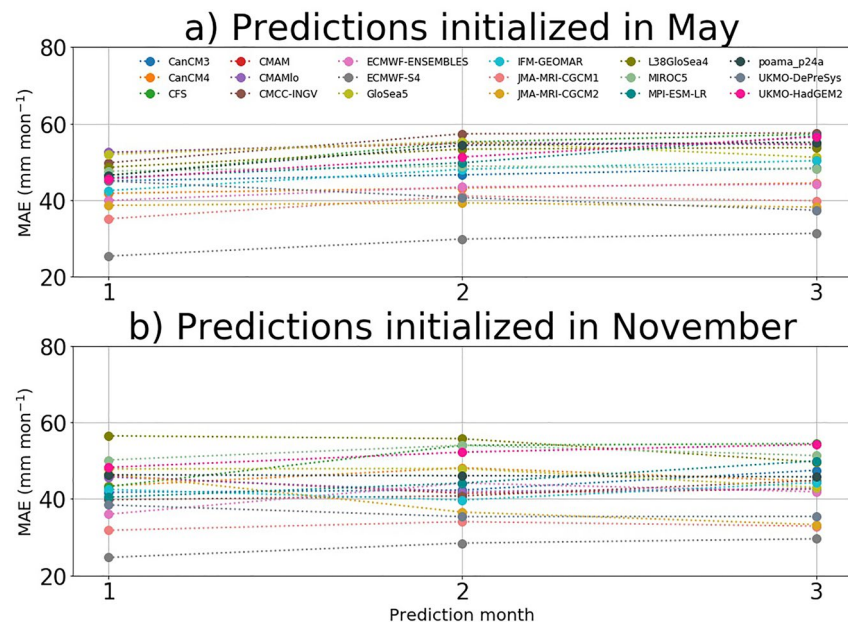


Figure 11. Evolution of mean absolute error (MAE) of seasonal predictions of monthly mean precipitation against the GPCPv2.3 reference data set, averaged over the 50°N–50°S region from month 1 to month 3 of the predictions. Units are mm month^{−1}.

the beginning of the prediction (first few months) as well as any zonal propagation and more gradually emerging transients.

Among the models considered, the most persistent difference occurs in BCC-CSM1.1 (Figure 14a), which exhibits positive (warm) differences relative to the free-running climatology between 180°W and 90°W from immediately after initialization until the end of the forecast range. The cause of this persistent difference is not obvious, considering that the BCC-CSM1.1 decadal predictions and historical simulations were subject to the same external forcing (Xin et al., 2019). In the cases of CanCM4 and CCSM4 (Figures 14b–14d), decadal hindcasts for CMIP5 were performed using two different initialization methods, labeled i1 and i2 (Karspeck et al., 2015; Merryfield et al., 2013). Although specifics of the initialization methods differ, for both models method i1 assimilates subsurface ocean data, whereas i2 applies surface ocean forcings only with no subsurface data constraints. With method i1 both models start from a condition very close to the reference data set (i.e., small differences in the first months of the predictions) but then evolve toward a temporary cold difference from year 2 to 4 in CanCM4_i1 and from year 1 to 3 in CCSM4_i1 (Figures 14b and 14d). The onset of both patterns apparently results from the emergence of cold biases that are present in sub-surface ocean temperatures at the start of the prediction (not shown). With method i2, CanCM4_i2 exhibits general warm differences (Figure 14c) somewhat like BCC-CSM1.1, whereas CCSM4_i2 shows warm differences in years 1 and 2 and cold differences in years 3 and 4 (Figure 14e), which have been attributed to spurious trends in the NCEP/NCAR reanalysis surface winds used to initialize the ocean component (Teng et al., 2017). CNRM-CM5 (Figure 14f) is affected by positive differences in the first year of the simulation that fade away around forecast year 3. Regarding the other models, EC-Earth and HadCM3 (Figures 14f and 14h) both show hints of zonal propagation of the errors from the coasts of South America westward in the first three years following initialization. The Hovmöller diagram suggests that these errors take around 12 months to cross the basin. After the third year, the differences in these models largely subside, although HadCM3 shows signs of warm differences emerging again east of 180° after year 8.

Having hindcast climatologies from a particular model initialized in multiple ways present opportunities for exploring the physical nature of transient behavior such as that for CanCM4 (Figures 14b and 14c), as an example. Figure 15 shows CanCM4 i1–i2 differences in SST and depth of the 20°C isotherm (which provides an indication of the depth of the equatorial Pacific thermocline). CanCM4 hindcasts using method i1 (which assimilates subsurface ocean temperatures) are slightly warmer than those using method i2 (no

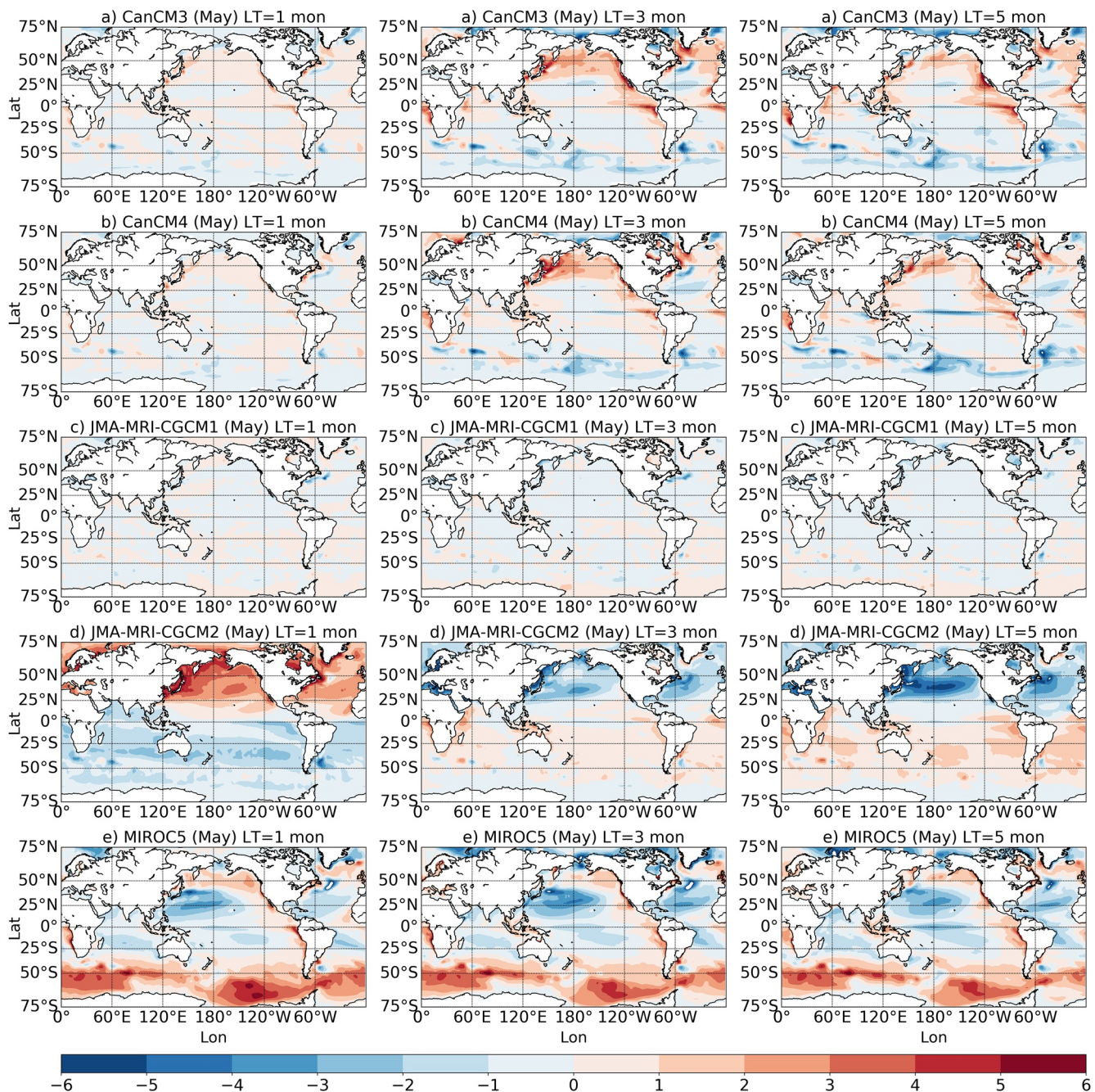


Figure 12. Differences in seasonal predictions of sea surface temperatures compared to their corresponding reference data sets (OISSTv2) for month 1, 3, and 5 (columns) of predictions initialized in May. Units are °C.

subsurface assimilation) in the first year (Figure 15a). However, at the end of the second year, a strong cold difference suddenly emerges (seen less distinctly in Figure 14b), which is followed by similar but gradually diminishing behavior in the next four years. Clues to the origins of these differences in transient behavior are provided by a comparable plot of differences in the 20°C-isotherm depth (Figure 15b). Late in the second year, a sudden upward, eastward propagating displacement of the thermocline appears, followed by similar but progressively weakening displacements in the next four years. These anomalous thermocline displacements resemble equatorial upwelling Kelvin waves, which propagate eastward across the Pacific in two to three months and are associated with cooling in the eastern equatorial Pacific and the onset of La Niña episodes (Sarachik & Cane, 2010). Determining the origin of these waves, which can be excited by

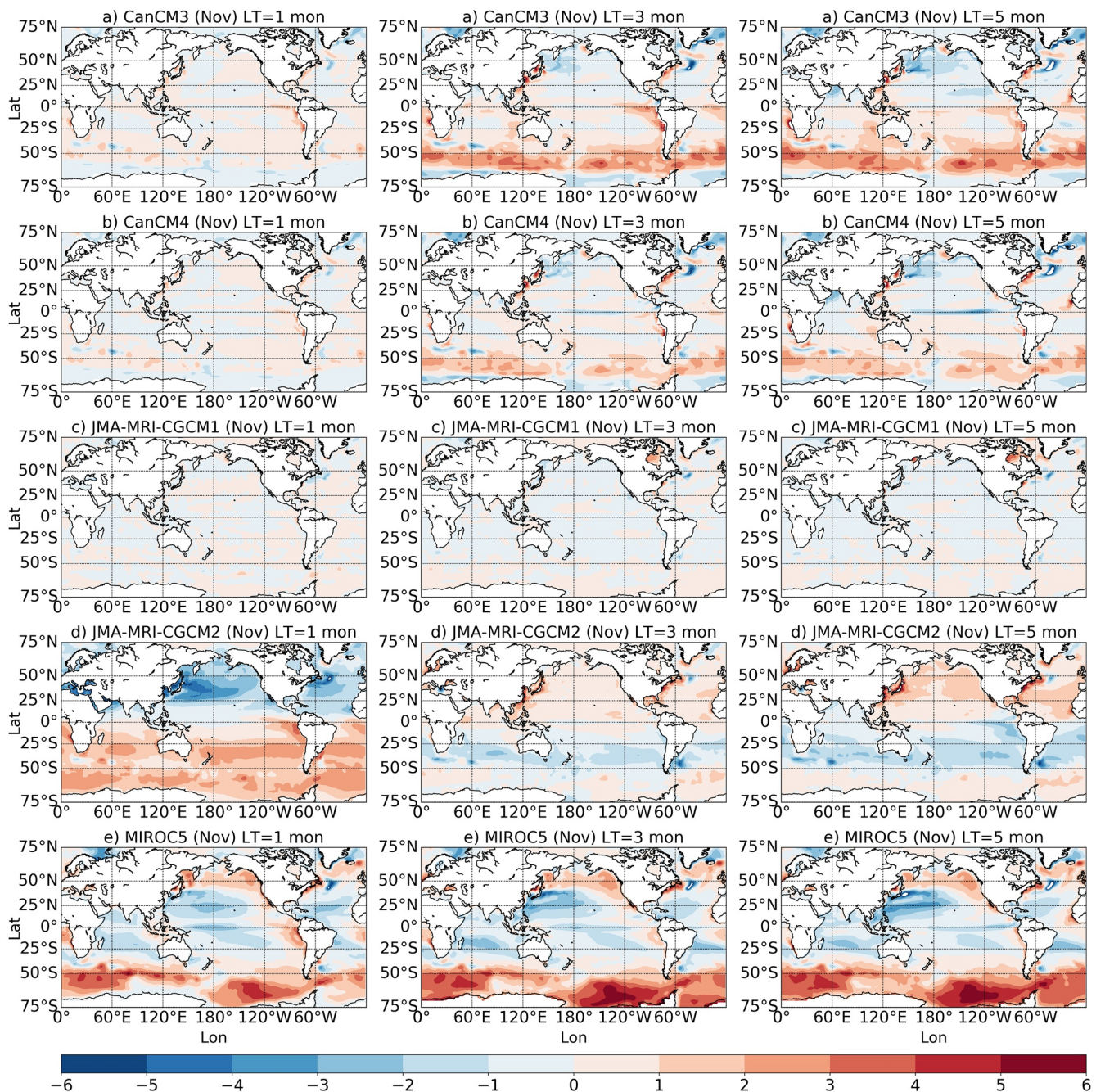


Figure 13. As in Figure 12 but for hindcasts initialized in November.

either convergent surface wind stress or ocean dynamical imbalances as might exist in the initial conditions, requires further analysis.

4. Concluding Remarks and Ways Forward

The development of a data set comprising hindcast climatologies derived from subseasonal, seasonal, and decadal predictions through the LRFTIP project provides an opportunity for the climate prediction research community to systematically characterize and better understand the physical mechanisms associated with transient behavior in initialized forecast systems. Although drawn from existing hindcast sources, the

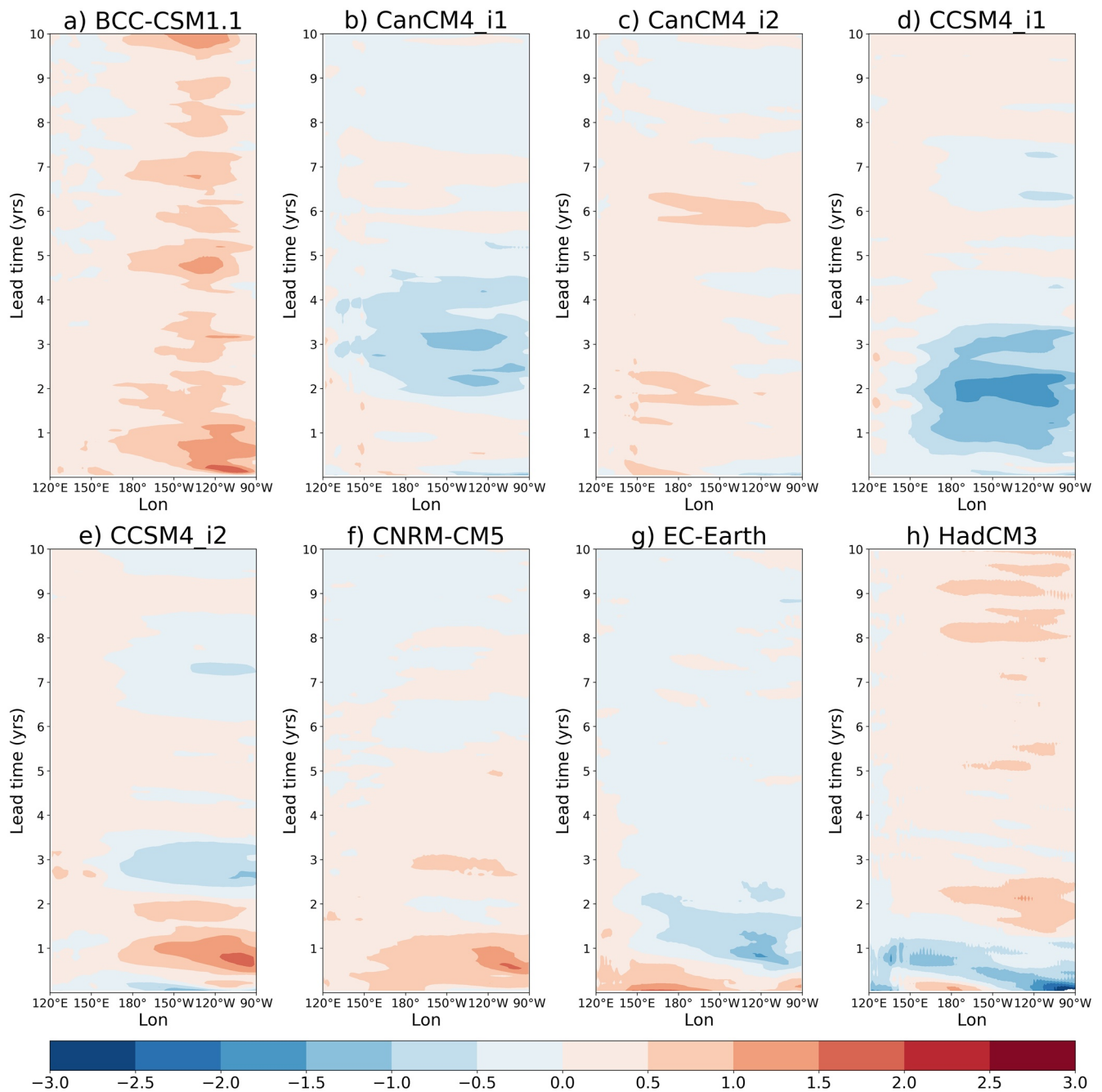


Figure 14. Hovmöller diagrams (longitude-lead time) of differences in decadal predictions of SST over the Equatorial Pacific (5°S–5°N, 120°E–90°W) compared to their corresponding historical runs. Units are °C and numbers in the y-axis are displayed in years.

LRFTIP data set removes the necessity of accessing all ensemble members for all hindcast years in order to compute the lead time-dependent climatology for a given variable, and thus reduces associated data volumes by orders of magnitude. In addition to providing precomputed hindcast climatologies for several variables from many models for time scales ranging from daily to decadal, the data set in most instances provides one or more reference data sets consisting of similarly gridded observations, initial conditions, and/or uninitialized historical simulations subject to the same temporal averaging as the hindcasts enabling the rapid calculation of metrics such as those illustrated in Figure 1.

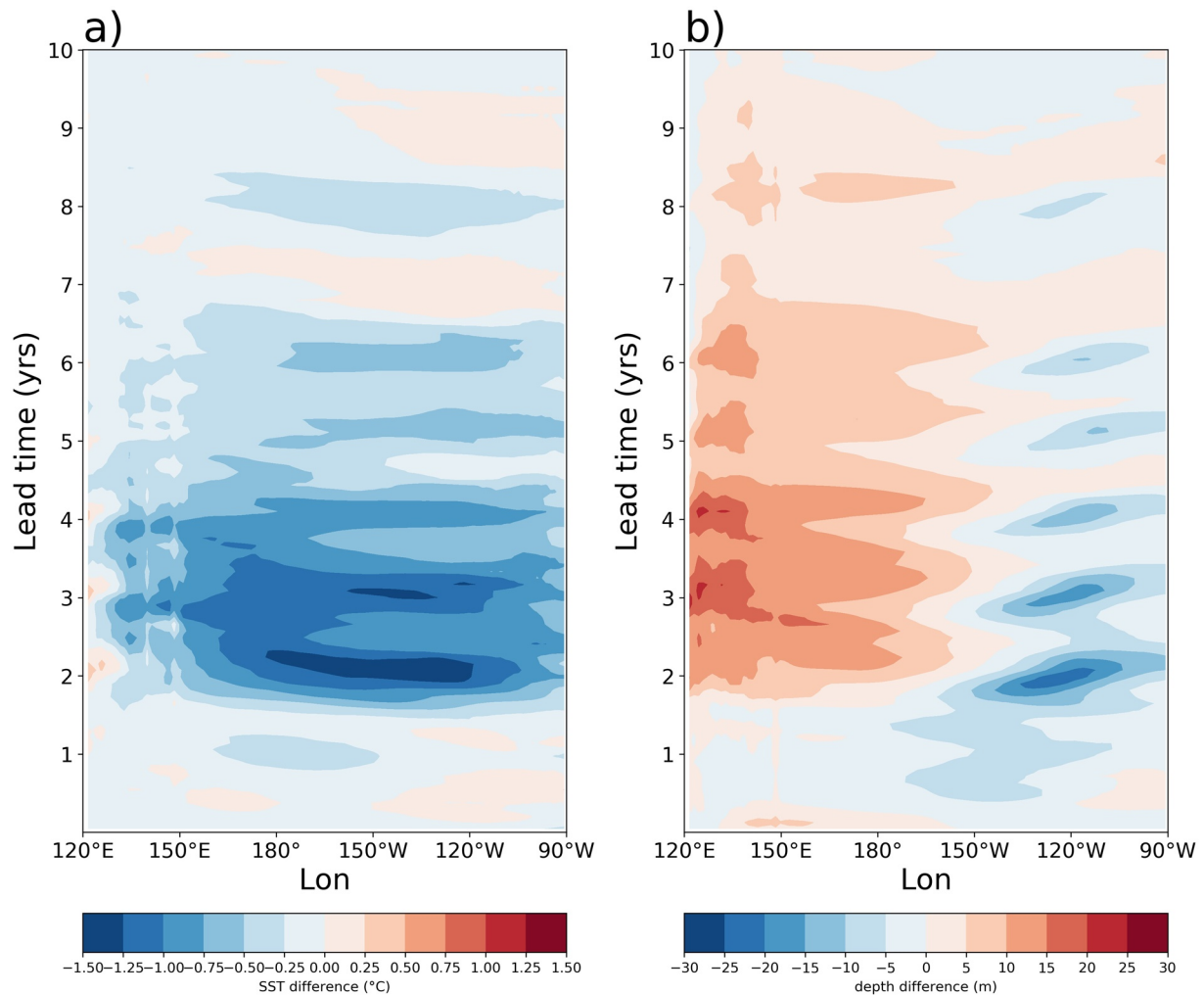


Figure 15. As in Figure 14, but showing i1-i2 differences between CanCM4 decadal predictions, where i1 and i2 refer to two different initialization methods in the CMIP5 archive, for (a) sea surface temperature (SST) and (b) depth of the 20°C isotherm, over the Equatorial Pacific.

This article aimed to describe and show some examples of the usage of the LRFTIP data set as an illustration of the types of analyses enabled by it. Any major progress in this field could prove extremely useful for both model developers (through providing guidance for improving model parameterizations and numerics) and to developers of climate services responsible for providing trustworthy climate prediction information across time scales.

In the context of open data policies and following examples such as the S2S, CHFP, and CMIP data sets, the LRFTIP data set aims to support community climate prediction and predictability studies. The usage of this data set is, therefore, strongly encouraged and indeed sought, as are feedbacks to and interactions with the WGSIP group.

Data Availability Statement

The data used in this study is freely available at the following link: <http://crd-data-donnees-rdc.ec.gc.ca/CCCMA/products/LRFTIP/> and is permanently registered as doi:10.18164/5dc6d3ae-8eb4-4292-b181-6a96104e63c5.

Acknowledgments

The authors would like to thank the two anonymous reviewers, whose suggestions helped improve and clarify several aspects of the manuscript, as well as Editor Dr. Stephen Griffies. In addition, Hai Lin and Yuhei Takaya provided helpful comments about the behavior of the ECCO-S2S and JMA-S2S models, respectively. The assistance of Marina Trubina in constructing S2S hindcast climatologies for the LRFTIP data set is also kindly acknowledged.

References

- Adler, R. F., Sapiiano, M., Huffman, G. J., Wang, J., Gu, G., Bolvin, D., et al. (2018). The Global Precipitation Climatology Project (GPCP) Monthly Analysis (new version 2.3) and a review of 2017 Global Precipitation. *Atmosphere*, 9, 138. <https://doi.org/10.3390/atmos9040138>
- Arribas, A., Glover, M., Maidens, A., Peterson, K., Gordon, M., MacLachlan, C., et al. (2011). The GloSea4 ensemble prediction system for seasonal forecasting. *Monthly Weather Review*, 139, 1891–1910. <https://doi.org/10.1175/2010MWR3615.1>
- Ballabrera-Poy, J., Kalnay, E., & Yang, S.-C. (2009). Data assimilation in a system with two scales—Combining two initialization techniques. *Tellus*, 61A, 539–549. <https://doi.org/10.1111/j.1600-0870.2009.00400.x>
- Balmaseda, M., & Anderson, D. (2009). Impact of initialization strategies and observations on seasonal forecast skill. *Geophysical Research Letters*, 36, L01701. <https://doi.org/10.1029/2008GL035561>
- Batté, L., & Doblas-Reyes, F. (2015). Stochastic atmospheric perturbations in the EC-Earth3 global coupled model: Impact of SPPT on seasonal forecast quality. *Climate Dynamics*, 45, 3419–3439. <https://doi.org/10.1007/s00382-015-2548-7>
- Bhargava, K., Kalnay, E., Carton, J., & Yang, F. (2018). Estimation of systematic errors in the GFS using analysis increments. *Journal of Geophysical Research: Atmosphere*, 123, 1626–1637. <https://doi.org/10.1002/2017JD027453>
- Bilbao, R., Wild, S., Ortega, P., Acosta-Navarro, J., Arsouze-Bretonnière, T. P.-A., Bretonnière, P.-A., et al. (2021). Assessment of a full-field initialized decadal climate prediction system with the CMIP6 version of EC-Earth. *Earth System Dynamics*, 12, 173–196. <https://doi.org/10.5194/esd-12-173-2021>
- Boer, G., Smith, D., Cassou, C., Doblas-Reyes, F., Danabasoglu, G., Kirtman, B., et al. (2016). The Decadal Climate Prediction Project (DCPP) contribution to CMIP6. *Geoscientific Model Development*, 9, 3751–3777. <https://doi.org/10.5194/gmd-9-3751-2016>
- Brient, F., Roehrig, R., & Voldoire, A. (2019). Evaluating marine stratocumulus clouds in the CNRM-CM6-1 model using short-term hindcasts. *Journal of Advances in Modeling Earth Systems*, 11, 127–148. <https://doi.org/10.1029/2018MS001461>
- Browne, P., de Rosnay, P., Zuo, H., Bennett, A., & Dawson, A. (2019). Weakly coupled ocean-atmosphere data assimilation in the ECMWF NWP System. *Remote Sensing*, 11, 234. <https://doi.org/10.3390/rs11030234>
- Cassou, C., Kushnir, Y., Hawkins, E., Pirani, A., Kucharski, F., Kang, I.-S., & Caltabiano, N. (2018). Decadal climate variability and predictability: Challenges and opportunities. *Bulletin of the American Meteorological Society*, 99, 479–490. <https://doi.org/10.1175/BAMS-D-16-0286.1>
- Chang, Y., Schubert, S., Koster, R., Molod, A., & Wang, H. (2019). Tendency bias correction in coupled and uncoupled global climate models with a focus on impacts over North America. *Journal of Climate*, 32, 639–661. <https://doi.org/10.1175/JCLI-D-18-0598.1>
- Chikamoto, Y., Timmermann, A., Widlansky, M., Zhang, S., & Balmaseda, M. (2019). A drift-free decadal climate prediction system for the Community Earth System Model. *Journal of Climate*, 32, 5967–5995. <https://doi.org/10.1175/JCLI-D-18-0788.1>
- Choudhury, D., Sen Gupta, A., Sharma, A., Mehrotra, R., & Sivakumar, B. (2017). An assessment of drift correction alternatives for CMIP5 decadal predictions. *Journal of Geophysical Research: Atmosphere*, 122, 10282–10. <https://doi.org/10.1002/2017JD026900>
- Cotter, C. (2013). Data assimilation on the exponentially accurate slow manifold. *Philosophical Transactions of the Royal Society A*, 371, 20120300. <https://doi.org/10.1098/rsta.2012.0300>
- Crawford, W., Frolov, S., McLay, J., Reynolds, C. A., Barton, N., Ruston, B., & Bishop, C. H. (2020). Using analysis corrections to address model error in atmospheric forecasts. *Monthly Weather Review*, 148, 3729–3745. <https://doi.org/10.1175/MWR-D-20-0008.1>
- Cruz-Garcia, R., Ortega, P., Guemas, V., Acosta Navarro, J. C., Massonnet, F., & Doblas-Reyes, F. J. (2021). An anatomy of Arctic sea ice forecast biases in the seasonal prediction system with EC-Earth. *Climate Dynamics*, 56, 1799–1813. <https://doi.org/10.1007/s00382-020-05560-4>
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., et al. (2011). The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137, 553–597. <https://doi.org/10.1002/qj.828>
- Doblas-Reyes, F. J., Balmaseda, M., Weisheimer, A., & Palmer, T. (2011). Decadal climate prediction with the European Centre for medium-range weather forecasts coupled forecast system: Impact of ocean observations. *Journal of Geophysical Research*, 116, 1984–2012. <https://doi.org/10.1029/2010JD015394>
- Doblas-Reyes, F. J., Weisheimer, A., Palmer, T., Murphy, J. M., & Smith, D. (2010). Forecast quality assessment of the ENSEMBLES seasonal-to-decadal Stream 2 hindcasts. *ECMWF Technical Memorandum*, 621, 45. <https://doi.org/10.21957/10x9tmf12>
- Exarchou, E., Prodhomme, C., Brodeau, L., Guemas, V., & Doblas-Reyes, F. J. (2018). Origin of the warm eastern tropical Atlantic SST bias in a climate model. *Climate Dynamics*, 51, 1819–1840. <https://doi.org/10.1007/s00382-017-3984-3>
- Ezer, T. (2016). Revisiting the problem of the Gulf Stream separation: On the representation of topography in ocean models with different types of vertical grids. *Ocean Modelling*, 104, 15–27. <https://doi.org/10.1016/j.ocemod.2016.05.008>
- Flato, G., Marotzke, J., Abiodun, B., Braconnot, P., Chou, S. C., Collins, W., et al. (2013). Evaluation of climate models. In *Climate change 2013: The physical science basis* (pp. 741–866). Cambridge University Press. <https://doi.org/10.1017/CBO9781107415324.020>
- García-Serrano, J., & Doblas-Reyes, F. J. (2012). On the assessment of near-surface global temperature and North Atlantic multi-decadal variability in the ENSEMBLES decadal hindcast. *Climate Dynamics*, 39, 2025–2040. <https://doi.org/10.1007/s00382-012-1413-1>
- Hawkins, E., Dong, B., Robson, J., Sutton, R., & Smith, D. (2014). The interpretation and use of biases in decadal climate predictions. *Journal of Climate*, 27, 2931–2947. <https://doi.org/10.1175/JCLI-D-13-00473.1>
- Hazeleger, W., Guemas, V., Wouters, B., Corti, S., Andreu-Burillo, I., Doblas-Reyes, F. J., et al. (2013). Multiyear climate predictions using two initialization strategies. *Geophysical Research Letters*, 40, 1794–1798. <https://doi.org/10.1002/grl.50355>
- He, Y., Wang, B., Huang, W., Xu, S., Wang, Y., Liu, L., et al. (2020). A new DRP-4DVar-based coupled data assimilation system for decadal predictions using a fast online localization technique. *Climate Dynamics*, 54, 3541–3559. <https://doi.org/10.1007/s00382-020-05190-w>
- He, Y., Wang, B., Liu, M., Yu, Y., Liu, J., Li, R., et al. (2017). Reduction of initial shock in decadal predictions using a new initialization strategy. *Geophysical Research Letters*, 44, 8538–8547. <https://doi.org/10.1002/2017GL074028>
- Hermanson, L., Ren, H.-L., Vellinga, M., Dunstone, N., Hyder, P., Ineson, S., et al. (2018). Different types of drifts in two seasonal forecast systems and their dependence on ENSO. *Climate Dynamics*, 51, 1411–1426. <https://doi.org/10.1007/s00382-017-3962-9>
- Huang, B., Zhu, J., Marx, L., Wu, X., Kumar, A., Hu, Z., et al. (2015). Climate drift of AMOC, North Atlantic salinity and Arctic sea ice in CFSv2 decadal predictions. *Climate Dynamics*, 44, 559–583. <https://doi.org/10.1007/s00382-017-2395-y>
- Karspeck, A., Yeager, S., Danabasoglu, S., & Teng, H. (2015). An evaluation of experimental decadal predictions using CCSM4. *Climate Dynamics*, 44, 907–923. <https://doi.org/10.1007/s00382-014-2212-7>
- Keane, R., Williams, K., Stirling, A., Martin, G., Birch, C., & Parker, D. (2019). Fast biases in monsoon rainfall over southern and central India in the Met Office Unified Model. *Journal of Climate*, 32, 6385–6402. <https://doi.org/10.1175/JCLI-D-18-0650.1>

- Kharin, V. V., Boer, G. J., Merryfield, W. J., Scinocca, J. F., & Lee, W.-S. (2012). Statistical adjustment of decadal predictions in a changing climate. *Geophysical Research Letters*, 39, L19705. <https://doi.org/10.1029/2012GL052647>
- Kharin, V. V., & Scinocca, J. F. (2012). The impact of model fidelity on seasonal predictive skill. *Geophysical Research Letters*, 39, L18803. <https://doi.org/10.1029/2012GL052815>
- Kim, H.-M., Webster, P. J., & Curry, J. A. (2012). Evaluation of short-term climate change prediction in multi-model CMIP5 decadal hindcasts. *Geophysical Research Letters*, 39, L10701. <https://doi.org/10.1029/2012GL051644>
- Kim, S., Jeong, H.-I., & Jin, F.-F. (2017). Mean Bias in Seasonal Forecast Model and ENSO Prediction Error. *Scientific Reports*, 7, 6029. <https://doi.org/10.1038/s41598-017-05221-3>
- Ma, H.-Y., Cheska Siongco, A., Klein, S. A., Xie, S., Karspeck, A. R., Raeder, K., et al. (2020). On the correspondence between seasonal forecast biases and long-term climate biases in sea surface temperature. *Journal of Climate*, 34, 427–446. <https://doi.org/10.1175/JCLI-D-20-0338.1>
- Ma, H.-Y., Xie, S., Klein, S. A., Williams, K. D., Boyle, J. S., Bony, S., et al. (2014). On the correspondence between mean forecast errors and climate errors in CMIP5 models. *Journal of Climate*, 27, 1781–1798. <https://doi.org/10.1175/JCLI-D-13-00474.1>
- Magnusson, L., Alonso-Balmaseda, M., Corti, S., Molteni, F., & Stockdale, T. (2013). Evaluation of forecast strategies for seasonal and decadal forecasts in presence of systematic model errors. *Climate Dynamics*, 41, 2393–2409. <https://doi.org/10.1007/s00382-012-1599-2>
- Manzanas, R. (2020). Assessment of model drifts in seasonal forecasting: Sensitivity to ensemble size and implications for bias correction. *Journal of Advances in Modeling Earth Systems*, 12, e2019MS001751. <https://doi.org/10.1029/2017MS001751>
- Meehl, G. A., Goddard, L., Boer, G., Burgman, R., Branstator, G., Cassou, C., et al. (2014). Decadal climate prediction: An update from the trenches. *Bulletin of the American Meteorological Society*, 95, 243–267. <https://doi.org/10.1175/BAMS-D-12-00241.1>
- Meehl, G. A., RichterTengCapotondi, J. H. A., Cobb, K., Doblas-Reyes, F., Cobb, K., Doblas-Reyes, F., et al. (2021). Initialized Earth system prediction from subseasonal to decadal timescales. *Nature Reviews Earth and Environment*, 2, 340–357. <https://doi.org/10.1038/s43017-021-00155-x>
- Merryfield, W. J., Baehr, J., Batté, L., Becker, E. J., ButlerCoelho, A. H. C. A. S., Coelho, C. A. S., et al. (2020). Current and emerging developments in subseasonal to decadal prediction. *Bulletin of the American Meteorological Society*, 101, E869–E896. <https://doi.org/10.1175/BAMS-D-19-0037.1>
- Merryfield, W. J., Lee, W.-S., Boer, G. J., Kharin, V. V., Scinocca, J. F., Flato, G. M., et al. (2013). The Canadian seasonal to interannual prediction system. Part I: Models and initialization. *Monthly Weather Review*, 141, 2910–2945. <https://doi.org/10.1175/MWR-D-12-00216.1>
- Mulholland, D., Laloyaux, P., Haines, K., & Balmaseda, M. (2015). Origin and impact of initialization shocks in coupled atmosphere-ocean forecasts. *Monthly Weather Review*, 143, 4631–4644. <https://doi.org/10.1175/MWR-D-15-0076.1>
- Mulholland, D. P., Haines, K., & Balmaseda, M. (2016). Improving seasonal forecasting through tropical ocean bias corrections. *Quarterly Journal of the Royal Meteorological Society*, 142, 2797–2807. <https://doi.org/10.1002/qj.2869>
- Polkova, I., Köhl, A., & Stammer, D. (2014). Impact of initialization procedures on the predictive skill of a coupled ocean-atmosphere model. *Climate Dynamics*, 42, 3151–3169. <https://doi.org/10.1007/s00382-013-1969-4>
- Polkova, I., Köhl, A., & Stammer, D. (2019). Climate-mode initialization for decadal climate predictions. *Climate Dynamics*, 53, 7097–7111. <https://doi.org/10.1007/s00382-019-04975-y>
- Puri, K. (1981). Local geostrophic wind correction in the assimilation of height data and its relationships to the slow manifold. *Monthly Weather Review*, 109, 52–55. [https://doi.org/10.1175/1520-0493\(1981\)109%3C0052:LGWCT%3E2.0.CO;2](https://doi.org/10.1175/1520-0493(1981)109%3C0052:LGWCT%3E2.0.CO;2)
- Reynolds, R., Rayner, N., Smith, T., Stokes, D., & Wang, W. (2002). An improved in situ and satellite SST analysis for climate. *Journal of Climate*, 15, 1609–1625. https://doi.org/10.1175/1520-0442_2002_015_1609_aiaas_2.0.co_2
- Richter, I. (2015). Climate model biases in the eastern tropical oceans: Causes, impacts and ways forward. *WIREs: Climate Change*, 6, 345–358. <https://doi.org/10.1002/wcc.338>
- Robertson, A. W., Kumar, A., Peña, M., & Vitart, F. (2015). Improving and promoting subseasonal to seasonal prediction. *Bulletin of the American Meteorological Society*, 96, ES49–ES53. <https://doi.org/10.1175/BAMS-D-14-00139.1>
- Rodwell, M., & Palmer, T. (2007). Using numerical weather prediction to assess climate models. *Quarterly Journal of the Royal Meteorological Society*, 133, 129–146. <https://doi.org/10.1002/qj.23>
- Saha, S., Moorthi, S., Wu, X., Wang, J., Nadiga, S., Tripp, P., et al. (2014). The NCEP Climate Forecast System version 2. *Journal of Climate*, 27, 2185–2208. <https://doi.org/10.1175/JCLI-D-12-00823.1>
- Sánchez-Gómez, E., Cassou, C., Ruprich-Robert, Y., Fernandez, E., & Terray, L. (2016). Drift dynamics in a coupled model initialized for decadal forecasts. *Climate Dynamics*, 46, 1819–1840. <https://doi.org/10.1007/s00382-015-2678-y>
- Sarachik, E. S., & Cane, M. A. (2010). *The El Niño–Southern oscillation phenomenon* (p. 384). Cambridge University Press. <https://doi.org/10.1017/CBO9780511817496>
- Scaife, A. A., Ferranti, L., Alves, O., Athanasiadis, P., Baehr, J., Dequé, M., et al. (2019). Tropical rainfall predictions from multiple seasonal forecast systems. *International Journal of Climatology*, 39, 974–988. <https://doi.org/10.1002/joc.5855>
- Schubert, S., Chang, Y., Wang, H., Koster, R., & Molod, A. (2019). A systematic approach to assessing the sources and global impacts of errors in climate models. *Journal of Climate*, 32, 8301–8321. <https://doi.org/10.1175/JCLI-D-19-0189.1>
- Shonk, J. K. P., Demissie, T. D., & Toniazzo, T. (2019). A double ITCZ phenomenology of wind errors in the equatorial Atlantic in seasonal forecasts with ECMWF models. *Atmospheric Chemistry and Physics*, 19, 11383–11399. <https://doi.org/10.5194/acp-19-11383-2019>
- Shonk, J. K. P., Guilyardi, E., Toniazzo, T., Woolnough, S. J., & Stockdale, T. (2018). Identifying causes of western Pacific ITCZ drift in ECMWF System 4 hindcasts. *Climate Dynamics*, 50, 939–954. <https://doi.org/10.1007/s00382-017-3650-9>
- Shukla, R. P., Huang, B., Marx, L., Kinter, J. L., & Shin, C.-S. (2018). Predictability and prediction of Indian summer monsoon by CFSv2: Implication of the initial shock effect. *Climate Dynamics*, 50, 159–178. <https://doi.org/10.1007/s00382-017-3594-0>
- Siongco, A. C., Ma, H., Klein, S. A., Xie, S., Karspeck, A. R., Raeder, K., & Anderson, J. L. (2020). A hindcast approach to diagnosing the equatorial Pacific cold tongue SST bias in CESM1. *Journal of Climate*, 33, 1437–1453. <https://doi.org/10.1175/JCLI-D-19-0513.1>
- Smith, D., Eade, R., & Pohlmann, H. (2013). A comparison of full-field and anomaly initialization for seasonal to decadal climate prediction. *Climate Dynamics*, 41, 3325–3338. <https://doi.org/10.1007/s00382-013-1683-2>
- Teng, H., Meehl, G., Branstator, G., Yeager, S., & Karspeck, A. (2017). Initialization shock in CCSM4 decadal prediction experiments. In *CLIVAR Exchanges*, No. 72 (Vol. 25, pp. 41–46). International CLIVAR Project Office. <https://doi.org/10.22498/pages.25.1.41>
- Tompkins, A., Ortiz de Zárate, M. I., Saurral, R. I., Vera, C. S., Saulo, A. C., Merryfield, W., et al. (2017). The climate-system historical forecast project: Providing open access to seasonal forecast ensembles from centers around the globe. *Bulletin of the American Meteorological Society*, 98, 2293–2301. <https://doi.org/10.1175/BAMS-D-16-0209.1>
- Toniazzo, T., & Woolnough, S. (2014). Development of warm SST errors in the southern tropical Atlantic in CMIP5 decadal hindcasts. *Climate Dynamics*, 43, 2889–2913. <https://doi.org/10.1007/s00382-013-1691-2>

- Vannière, B., Guilyardi, E., Madec, G., Doblas-Reyes, F. J., & Woolnough, S. (2013). Using seasonal hindcasts to understand the origin of the equatorial cold tongue bias in CGCMs and its impact on ENSO. *Climate Dynamics*, 40, 963–981. <https://doi.org/10.1007/s00382-012-1429-6>
- Vitart, F., Robertson, A., & Anderson, D. (2012). Subseasonal to seasonal prediction project: Bridging the gap between weather and climate. *Bulletin of the World Meteorological Organization*, 61, 23–28.
- Voldoire, A., Exarchou, E., Sánchez-Gómez, E., Demissie, T., Deppenmeier, A.-L., Frauen, C., et al. (2019). Role of wind stress in driving SST biases in the Tropical Atlantic. *Climate Dynamics*, 53, 3481–3504. <https://doi.org/10.1007/s00382-019-04717-0>
- Volpi, D., Guemas, V., & Doblas-Reyes, F. J. (2017). Comparison of full field and anomaly initialisation for decadal climate prediction: Towards an optimal consistency between the ocean and sea-ice anomaly initialisation state. *Climate Dynamics*, 49, 1181–1195. <https://doi.org/10.1007/s00382-016-3373-3>
- Wang, C., Zhang, L., Lee, S.-K., Wu, L., & Mechoso, C. (2015). A global perspective on CMIP5 climate model biases. *Nature Climate Change*, 4, 201–205. <https://doi.org/10.1038/nclimate2118>
- Williams, K. D., Bodas-Salcedo, A., Déqué, M., Fermepin, S., Medeiros, B., Watanabe, M., et al. (2013). The Transpose-AMIP II experiment and its application to the understanding of Southern Ocean cloud biases in climate models. *Journal of Climate*, 26, 3258–3274. <https://doi.org/10.1175/JCLI-D-12-00429.1>
- Woelfle, M., Yu, S., Bretherton, C., & Pritchard, M. (2018). Sensitivity of coupled tropical Pacific model biases to convective parameterization in CESM1. *Journal of Advances in Modeling Earth Systems*, 10, 126–144. <https://doi.org/10.1002/2017MS001176>
- Xin, X., Wei, M., Li, Q., Zhou, W., Luo, Y., & Zhao, Z. (2019). Decadal prediction skill of BCC-CSM1.1 with different initialization strategies. *Journal of the Meteorological Society of Japan*, 97, 733–744. <https://doi.org/10.2151/jmsj.2019-043>
- Yeager, S. G., Danabasoglu, G., Rosenbloom, N. A., Stand, W., Bates, S. C., Meehl, G., et al. (2018). Predicting near-term changes in the earth system: A large ensemble of initialized decadal prediction simulations using the community earth system model. *Bulletin of the American Meteorological Society*, 99, 1867–1886. <https://doi.org/10.1175/BAMS-D-17-0098.1>
- Zadra, A., Williams, K., Frasson, A., Rixen, M., Adames, A., Berner, J., et al. (2018). Systematic errors in weather and climate models: Nature, origins, and ways forward. *Bulletin of the American Meteorological Society*, 99, ES67–ES70. <https://doi.org/10.1175/BAMS-D-17-0287.1>