



Seasonal predictions initialised by assimilating sea surface temperature observations with the EnKF

Yiguo Wang¹ · François Counillon¹ · Noel Keenlyside^{1,2} · Lea Svendsen² · Stephanie Gleixner³ · Madlen Kimmritz¹ · Panxi Dai⁴ · Yongqi Gao^{1,5}

Received: 6 December 2018 / Accepted: 11 July 2019 / Published online: 23 July 2019
© The Author(s) 2019

Abstract

This study demonstrates that assimilating SST with an advanced data assimilation method yields prediction skill level with the best state-of-the-art systems. We employ the Norwegian Climate Prediction Model (NorCPM)—a fully-coupled forecasting system—to assimilate SST observations with the ensemble Kalman filter. Predictions of NorCPM are compared to predictions from the North American Multimodel Ensemble (NMME) project. The global prediction skill of NorCPM at 6- and 12-month lead times is higher than the averaged skill of the NMME. A new metric is introduced for ranking model skill. According to the metric, NorCPM is one of the most skilful systems among the NMME in predicting SST in most regions. Confronting the skill to a large historical ensemble without assimilation, shows that the skill is largely derived from the initialisation rather than from the external forcing. NorCPM achieves good skill in predicting El Niño–Southern Oscillation (ENSO) up to 12 months ahead and achieves skill over land via teleconnections. However, NorCPM has a more pronounced reduction in skill in May than the NMME systems. An analysis of ENSO dynamics indicates that the skill reduction is mainly caused by model deficiencies in representing the thermocline feedback in February and March. We also show that NorCPM has skill in predicting sea ice extent at the Arctic entrance adjacent to the north Atlantic; this skill is highly related to the initialisation of upper ocean heat content.

Keywords Seasonal prediction · Advanced data assimilation · EnKF · SST · NorCPM · ENSO · Sea ice extent

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s00382-019-04897-9>) contains supplementary material, which is available to authorized users.

✉ Yiguo Wang
yiguo.wang@nersc.no

- ¹ Nansen Environmental and Remote Sensing Center, Bjerknes Centre for Climate Research, Bergen, Norway
- ² Geophysical Institute, University of Bergen and Bjerknes Centre for Climate Research, Bergen, Norway
- ³ Potsdam Institute for Climate Impact Research, Potsdam, Germany
- ⁴ Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing, China
- ⁵ Nansen-Zhu International Research Center, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

1 Introduction

In recent decades, it has been proven that tailored seasonal climate forecasts are more beneficial than climatology for decision-making in many sectors of society, e.g. energy, agriculture, transport, insurance and water resource management (Soares and Dessai 2015). For example, Torralba et al. (2017) recently demonstrated that skilful seasonal predictions of near-surface air temperature and wind speed in winter improve the predictability of wind energy demand and supply; Gunda et al. (2017) demonstrated that using seasonal forecasts increases the agricultural income. To fulfil societal needs of seasonal climate forecasts, many meteorology centres (e.g., the European Centre for Medium-Range Weather Forecasts, ECMWF) and climate prediction centres (e.g., the Climate Prediction Center of the National Centers for Environmental Prediction, NCEP) have been providing seasonal forecasts operationally for many years; and many others are developing operational systems.

Seasonal climate predictions (Doblas-Reyes et al. 2013) are performed by statistical models (van den Dool 2006),

dynamical models (Ji et al. 1996) or a combination of both (Krishnamurti et al. 1999). Dynamical models, such as ocean–atmosphere coupled general circulation models (CGCMs), are more suitable than statistical models to deal with unprecedented climate signals and chaotic behaviour in the climate system (Barnston et al. 1999). Most climate forecasting systems (e.g. systems shown in Weisheimer et al. 2009; Kirtman et al. 2014) deliver ensemble forecasts, where the forecast and its uncertainty are provided by the statistics of the ensemble, including the mean, median, variance and range. The ensemble forecasts are mainly based on a single dynamical model and generated by perturbing initial conditions and forcing, and through stochastic perturbations of model physics. Multimodel or perturbing model physics approaches were proposed to enhance the prediction skill by reducing the errors caused by initialisation and model physics (Palmer et al. 2004). A successful example is the North American Multi-Model Ensemble (NMME, Kirtman et al. 2014) operational forecasting system that consists of more than 13 CGCMs from US and Canadian modelling centres with ensemble members in individual CGCMs ranging from 6 to 24. A combination of statistical and dynamical models is an alternative for the seasonal forecasts tailored to certain specific needs. For instance, Gleixner et al. (2017) showed that a statistical–dynamical model beats a multimodel ensemble of 11 CGCMs for the seasonal prediction of the Kiremt rainfall. Analogs based on dynamical models has also introduced competitive means to perform seasonal predictions (Ding et al. 2018).

Initial conditions fed to CGCMs largely dominate dynamical climate predictions at short time scales (Kirtman et al. 2013; Doblas-Reyes et al. 2013). Balmaseda and Anderson (2009) evaluated three commonly used initialisation strategies (including atmosphere or/and ocean initialisations) in the ECMWF System3 for seasonal climate predictions. They demonstrated that ocean initialisation has a significant impact on seasonal predictions of CGCMs and recommended to initialise CGCMs with both ocean and atmosphere data. Most current ocean initialisations used for seasonal forecasting systems (Balmaseda et al. 2009) assimilate sea surface temperature (SST) observations, since SST plays a key role in influencing atmospheric circulations (Shukla 1998). They also assimilate many other oceanic observation types, e.g. subsurface temperature and salinity and altimeter data that have been demonstrated to improve seasonal predictions (Balmaseda and Anderson 2009).

As reviewed in Balmaseda et al. (2009) and Penny et al. (2017), most operational centres are evolving towards the use of sophisticated initialisation schemes in which observations in different components are best used with advanced data assimilation (DA) methods such as the ensemble Kalman filter (EnKF, Evensen 2003), the 4-dimensional variational DA (4DVAR, Dimet and Olivier

1986) and hybrid methods (Hamill and Snyder 2000; Carrassi et al. 2018). However, some previous studies (e.g., Luo et al. 2005; Zhu et al. 2017) demonstrated that a simpler initialisation scheme (i.e., nudging SST towards observations) already achieves skilful seasonal predictions. The mechanism for this scheme is that nudging SST towards observations reproduces well the variability of temperature in the mixed layer and the variability of subsurface temperature through air–sea interaction (Luo et al. 2005; Keenlyside et al. 2005; Kumar et al. 2014; Kumar and Zhu 2018). The method is most effective in the tropical Pacific where air–sea interaction is strongest, and can be problematic in high–latitudes and lead to spurious impacts on ocean thermohaline circulation in some models (Dunstone and Smith 2010). Nevertheless, the SST nudging scheme can reduce errors caused by inconsistencies between separately initialised components of CGCMs (i.e., initialisation shock). Luo et al. (2005) used the SST nudging initialisation scheme in the Scale Interaction Experiment-Frontier Research Center for Global Change (SINTEX-F) and achieved skilful predictions of the El Niño–Southern Oscillation (ENSO) up to 12 months. Zhu et al. (2017) even showed that such simple initialisation scheme could achieve a comparable seasonal SST prediction skill to the averaged skill of the NMME models that use more data (e.g., oceanic subsurface data and atmospheric data) and/or more advanced DA methods for initialisation.

The main aim of this study is to assess whether assimilating SST observations with an advanced DA method can be as or even more competitive than current state of the systems that assimilate more data. To do so, we will work with the Norwegian Climate Prediction Model (NorCPM, Counillon et al. 2014) which aims to provide long-term reanalyses and seasonal-to-decadal climate predictions. NorCPM uses the EnKF, which is an advanced and flow-dependent DA method. Counillon et al. (2016) showed that upper ocean heat content in the equatorial and north Pacific, the north Atlantic subpolar gyre region and the Nordic Seas can be well constrained by assimilating SST anomalies (SSTAs) with the EnKF. Since seasonal predictions largely depend on upper ocean heat content (Balmaseda and Anderson 2009; Balmaseda et al. 2009), we expect that assimilation of SSTA with the EnKF achieves skilful seasonal predictions. In addition, this paper is the first demonstration of the seasonal prediction skill of NorCPM in a real-experiment framework. Note that although Counillon et al. (2014) studied the seasonal prediction of NorCPM, their study was carried out in a twin-experiment (also known as a perfect model) framework in which the ‘truth’ was an independently run simulations of the same model. Pseudo observations were generated from

the ‘truth’ and used to initialise the model and assess the (perfect model) prediction experiments.

This paper is organised as follows. The model, experimental design and data used for validation are described in Sect. 2. The global prediction skill of NorCPM is presented and compared to the NMME systems in Sect. 3. The ENSO prediction skill is investigated in Sect. 4. The prediction skill of regional Arctic sea ice extent (SIE) is assessed in Sect. 5.

2 Model, experimental design and data

2.1 Norwegian climate prediction model

NorCPM (Counillon et al. 2014) is a climate prediction system developed for seasonal-to-decadal climate predictions and long-term reanalyses. It combines the Norwegian Earth system model (NorESM, Bentsen et al. 2013) and the EnKF (Evensen 2003). NorCPM is unique in the sense that it uses a global isopycnal ocean model (i.e., MICOM, Bentsen et al. 2013) and an advanced error flow-dependent DA method (i.e., the EnKF).

NorESM (Bentsen et al. 2013) is a global fully-coupled model for climate simulations. It is based on the Community Earth System Model version 1.0.3 (CESM1, Verstein et al. 2012), a successor to the Community Climate System Model version 4 (CCSM4, Gent et al. 2011). In NorESM, the ocean component is an updated version (Bentsen et al. 2013) of the isopycnal coordinate ocean model MICOM (Bleck et al. 1992); the sea ice component is the Los Alamos sea ice model (CICE4, Gent et al. 2011; Holland et al. 2012); the atmosphere component is a version of the Community Atmosphere Model (CAM4-Oslo, Kirkevåg et al. 2013); the land component is the Community Land Model (CLM4, Oleson et al. 2010; Lawrence et al. 2011); the version 7 coupler (CPL7, Craig et al. 2012) is used.

In this study, we employ the version of NorESM that is included in the Coupled Model Intercomparison Project Phase 5 (CMIP5, Taylor et al. 2012). MICOM has a horizontal resolution of approximately 1°, 51 isopycnal layers and 2 additional layers for representing the bulk mixed layer; CICE4 is on the same grid as MICOM; CAM4 has a horizontal resolution of 1.9° at latitude and 2.5° at longitude and 26 vertical levels in a hybrid sigma-pressure coordinate; CLM4 shares the same horizontal grid as CAM4. We briefly introduce the model performance of NorESM with respect to the ENSO (Sect. 4) and the Arctic sea ice (Sect. 5) in the following paragraphs.

The version of NorESM used in this paper features positive SST biases in the eastern boundary upwelling regions in the tropics that are common in CMIP5 models (Richter 2015) and positive precipitation biases related to the double Intertropical Convergence Zone (ITCZ, Bentsen et al. 2013).

The meridional width of the tropical Pacific cold tongue is also too narrow (Bentsen et al. 2013). Despite these common shortcomings, NorESM is evaluated as one of better models for simulating ENSO (Bellenger et al. 2014). NorESM reproduces realistic SST variability in the Nino3 region (5°S–5°N, 150°W–90°W) and the Nino4 region (5°S–5°N, 160°E–150°W). The power spectrum, seasonality and spatial structure of ENSO in NorESM are also comparable to observations (Bellenger et al. 2014). In addition, NorESM simulates well ENSO teleconnection patterns, such as inter-annual variations of the relationship between ENSO and the Asian summer monsoon (Sperber et al. 2013).

The version of NorESM used in this paper reproduces a fairly realistic geographic distribution of the Arctic sea ice. However, the summer Arctic SIE is too large, since the summer melting is too slow. It is likely linked to the too thick sea ice in NorESM, in particular in the polar oceans adjacent to the Eurasian continent. The general thick Arctic sea ice in NorESM is due to too little summer melt of snow (i.e., too little surface melt of the ice). Moreover, the winter melting in NorESM lags the observed melting. For further details see Bentsen et al. (2013).

2.2 Experimental design

We base the prediction skill assessment of the system on hindcasts (i.e., retrospective predictions), similar to many previous studies (e.g., Luo et al. 2015; Zhu et al. 2017). Seasonal hindcasts start on the 15th of January, April, July and October each year during 1985–2010. Totally, there are 104 hindcasts (26 years and 4 season starts per year). Each hindcast consists of 9 realisations (ensemble members) and is 13 months long. The hindcasts are forced by CMIP5 historical forcings (Taylor et al. 2012) before 2005 and the representative Concentration Pathway 8.5 (RCP8.5, van Vuuren et al. 2011) forcings after 2005. The CMIP5 historical forcings for 1850–2005 are based on observational variations in solar radiation (Lean et al. 2005; Wang et al. 2005), volcanic sulphate aerosol concentration (Ammann et al. 2003), GHG concentration (Lamarque et al. 2010), aerosol emission (Lamarque et al. 2010), and land-use (Hurt et al. 2009). Initial conditions are taken from the first 9 out of the 30 ensemble members¹ of a reanalysis covering the period of 1980–2010. In this reanalysis we use the same model and DA settings to Counillon et al. (2016). We assimilate monthly SSTA data from Hadley Centre Sea Ice and Sea Surface Temperature dataset version 2.1 (HadISST2.1.0.0, Kennedy et al., personal communication; Rayner et al., personal communication) with the EnKF into the ocean component of NorCPM at each assimilation step; the other

¹ All members are equally likely in the ensemble simulations.

components are dynamically adjusted during the system integration after the assimilation step. Note that we perform an anomaly assimilation with climatology defined for the period 1980–2010. The initial conditions of the reanalysis are taken from a 30-member historical simulation ensemble of NorESM that was integrated from 1850 to 2010 using CMIP5 historical forcings (Taylor et al. 2012). For further details see Counillon et al. (2016).

2.3 Data

In order to evaluate the prediction skill of NorCPM in a large framework, we compare the NorCPM hindcasts to the NMME hindcasts (Kirtman et al. 2014). The NMME is a multi-model seasonal forecasting system, that consists of several coupled climate models with different setups from US and Canadian modelling centres. For further details see Kirtman et al. (2014) or <https://www.earthsystemgrid.org/search.html?Project=NMME>. In this study, we select 13 NMME systems which provide SST hindcasts from 1985 to 2010 (Table 1). All NMME hindcasts start on the first day of each month and have lead times up to 8–12 months. Note that the NorCPM hindcasts start 15 days earlier than the NMME hindcasts (e.g., our hindcasts starting on the 15th of April are compared to the hindcasts of NMME starting on the 1st of May). The ensemble size ranges from 6 to 24 among the NMME models. Here we use the first 9 ensemble members of each NMME model (except for CCMS3 that only provides 6 ensemble members) to have a comparable ensemble size to NorCPM. The NMME hindcast data are provided as monthly means with a horizontal resolution of $1^\circ \times 1^\circ$.

For the validation of the hindcasts, we take SST data from the National Oceanic and Atmospheric Administration (NOAA) Optimum Interpolation SST (OISST) version 2 with a 1° resolution (Reynolds et al. 2002) which is not the SST dataset used to initialise the hindcasts of NorCPM. Note that the depth of SST is different in OISST, NorCPM and NMME systems. SST data in OISST determined from in situ observations (e.g., ships and buoys) and satellite data (e.g., Advanced Very High Resolution Radiometer) are the temperature at depths from a micro to several meters (Reynolds et al. 2002). The depth of SST in a model depends on vertical resolution (Table 1). For example, SST in the ocean component of CanCM3 and CanCM4 is the temperature at the first vertical level from the ocean surface to 10 m (Merryfield et al. 2013). SST in NorCPM is the temperature in the top layer (from 1 to 10 m) of the 2 layers for representing the bulk mixed layer and its depth varies in time (Bentsen et al. 2013). Although the difference of the depth of SST in observations and models may influence results, we do not take it into account in this paper. Monthly sea surface height (SSH) anomaly data (sea level anomalies) are taken from the

Global ARMOR3D L4 Reprocessed dataset (<http://marine.copernicus.eu>, available from 1993 to present) produced by Ssalto/Duacs and distributed by Aviso with support from the Centre national d'études spatiales (www.aviso.oceanobs.com/duacs/). The product is gridded to a resolution of 1° in order to have a comparable resolution to NorCPM. Precipitation observations are the combined monthly precipitation dataset of the Global Precipitation Climatology Project (GPCP) version 2.2 (<https://climatedataguide.ucar.edu/climate-data/gpcp-monthly-global-precipitation-climatology-project>). They are available on a $2.5^\circ \times 2.5^\circ$ grid from 1979 to present. Other atmospheric data are derived from the NCEP reanalysis data (Kalnay et al. 1996) provided by the NOAA Physical Sciences Division, Boulder, Colorado, USA, from their Web site at <https://www.esrl.noaa.gov/psd/>. Oceanic subsurface data is taken from the EN4 objective analysis (EN4.1.1, Good et al. 2013). Sea ice concentration data is taken from HadISST2.1.0.0 (Torralba et al. 2017). Note that SST data in the regions covered by sea ice are not assimilated; the regions are defined by the sea ice data provided by HadISST2.1.0.0. In the following sections, the anomaly calculations are based on the reference period 1985–2010.

3 Global prediction

In this section, we assess the skill of NorCPM in predicting global SST, air temperature at 2 m (T2m) and precipitation and compare its performance to that of the NMME (Sect. 2.3). Skill is measured by anomaly correlation coefficient (ACC) and the bias-free root mean square error (RMSE; i.e., computed with anomalies) between hindcasts and observations.

Figures 1 and 2 show the global ACC and RMSE maps of T2m at 6- and 12-month lead times and precipitation at 3-month lead time. NorCPM and NMME have skill in predicting T2m at 6- and 12-month lead times with higher ACC and lower RMSE over the oceans and in the tropics. The skill of NorCPM is higher than the skill averaged over the NMME systems. In terms of precipitation, both NorCPM and NMME have skill at 3-month lead time over the tropical Pacific. Table 2 shows the global average of ACCs and RMSEs for T2m at 6- and 12-month lead times and precipitation at 3-month lead time. Note that we compute ACC for the individual NMME systems and the average of the ACCs of the NMME systems is shown in Table 2. For skill at 12-month lead time, we only use the 8 NMME systems providing data. NorCPM has higher ACCs and lower RMSEs than the average of ACCs and RMSEs of the NMME for T2m and precipitation on the global scale.

In the following, we focus on assessing the skill in predicting SST. Since SST is very sensitive to the air–sea

Table 1 Brief description of NorCPM and 13 NMME models used in this paper

System	General information						Atmosphere	
	Institution	Hindcast period	Ensemble size	Lead times (months)	Reference	Model	Resolution	
NorCPM	Bjerknes Centre for Climate Research	1985–2010	9	1–12	Coumillon et al. (2016)	CAM4-Oslo	2.5° × 1.9°, 26 levels	
CanCM3	Canadian Meteorological Centre (CMC)	1981–2010	10	1–12	Merryfield et al. (2013)	CanAM3	T63, 31 levels	
CanCM4	CMC	1981–2010	10	1–12	Merryfield et al. (2013)	CanAM4	T63, 35 levels	
CCSM3	National Center for Atmospheric Research (NCAR)	1982–2017	6	1–12	Kirtman and Min (2009), Gent et al. (2011), Paolino et al. (2012), Infanti and Kirtman (2016)	CAM3	T85, 26 levels	
CCSM4	NCAR	1982–2017	10	1–12	Kirtman and Min (2009), Gent et al. (2011), Paolino et al. (2012), Infanti and Kirtman (2016)	CAM4	0.9° × 1.25°, 26 levels	
CM2p1	Geophysical Fluid Dynamics Laboratory (GFDL)	1982–2012	10	1–12	Delworth et al. (2006), Zhang et al. (2007), personal communication	AM2.0	2° × 2.5°, 26 levels	
CM2p1-aer04	GFDL	1982–2017	10	1–12	Delworth et al. (2006), Zhang et al. (2007), personal communication	AM2.0	2° × 2.5°, 26 levels	
CM2p5-FLOR-A06	GFDL	1980–2017	12	1–12	Vecchi et al. (2014), Zhang et al. (2007), personal communication	AM2.5	0.5° × 0.5°, 32 levels	
CM2p5-FLOR-B01	GFDL	1980–2017	12	1–12	Vecchi et al. (2014), Zhang et al. (2007), personal communication	AM2.5	0.5° × 0.5°, 32 levels	
ECHAM4p5-Anomaly	International Research Institute for Climate and Society (IRI)	1982–2012	12	1–8	DeWitt (2005)	ECHAM4	T42, 19 levels	
ECHAM4p5-Direct	IRI	1982–2012	12	1–8	DeWitt (2005)	ECHAM4	T42, 19 levels	
GMAO	National Aeronautics and Space Administration (NASA)	1981–2012	10	1–9	Vernieres et al. (2012), Borovikov et al. (2017), personal communication	Fortuna-1	1° × 1.25°, 72 levels	

Table 1 (continued)

System	General information				Atmosphere		
	Institution	Hindcast period	Ensemble size	Lead times (months)	Reference	Model	Resolution
GMAO-062012	NASA	1981–2017	12	1–9	Vernieres et al. (2012), Borovikov et al. (2017), personal communication	Fortuna-2.5	1° × 1.25°, 72 levels
CFSv2	National Centers for Environmental Prediction (NCEP)	1982–2010	24	1–10	Saha et al. (2006, 2014)	GFS	T126, 64 levels
System	Atmosphere			Ocean			
	Radiative forcing			Model	Resolution	Initialisation (DA method)	
NorCPM	CMIP5 historical forcing until 2005 and RCP8.5 after 2005	No atmospheric DA	MICOM	1° × 1°, 53 levels		Assimilation of SST data (EnKF)	
CanCM3	Equivalent CO2 forcing representing effects of all GHGs, aerosol compensation: 40% reduction on GHGs forcing	T, u, v and q assimilation	CanOM4	1.41° × 0.94°, 40 levels		SST nudging, offline temperature assimilation (2DVAR) and salinity adjustment	
CanCM4	CMIP5 historical forcing until 2005 and RCP4.5 after 2005	T, u, v and q assimilation	CanOM4	1.41° × 0.94°, 40 levels		SST nudging, offline temperature assimilation (2DVAR) and salinity adjustment	
CCSM3	CO2 of 355 ppmv	Atm-only runs forced by observed SST	POP	1° × 1°, 40 levels		Global Ocean Data Assimilation System (GODAS, 3DVAR)	
CCSM4	CO2 and other greenhouse gases from IPCC AR3	Climate Forecast System Reanalysis (CFSR)	POP2	1.11° × 0.27° – 0.54°, 60 levels		CFSR (3DVAR)	
CM2p1	1990's level until 2005 and RCP4.5 after 2005	Atm-land-only runs forced by SST	OM3.1	1° × 1/3 – 1°, 50 levels		Assimilation of temperature and salinity profiles (EnKF)	
CM2p1-aer04	1990's level until 2005 and RCP4.5 with 2004's aerosol after 2005	Atm-land-only runs forced by SST	OM3.1	1° × 1/3 – 1°, 50 levels		Assimilation of temperature and salinity profiles (EnKF)	
CM2p5-FLOR-A06	1990's level until 2005 and RCP4.5 after 2005	Atm-land-only runs forced by SST	OM3.1	1° × 1/3 – 1°, 50 levels		Assimilation of temperature and salinity profiles (EnKF)	
CM2p5-FLOR-B01	1990's level until 2005 and RCP4.5 after 2005	Atm-land-only runs forced by SST	OM3.1 ocean	1° × 1/3 – 1°, 50 levels		Assimilation of temperature and salinity profiles (EnKF)	
ECHAM4p5-Anom-ally	Neglected anthropogenic forcing	Atm-only forced by SST	MOM3	1.5° × 0.5° – 1.5°, 25 levels		GODAS (3DVAR)	
ECHAM4p5-Direct	Neglected anthropogenic forcing	Atm-only forced by SST	MOM3	1.5° × 0.5° – 1.5°, 25 levels		GODAS (3DVAR)	

Table 1 (continued)

System	Atmosphere		Ocean	
	Radiative forcing	Initialisation	Model	Resolution
GMAO	CMIP5 historical forcing until 2005 and RCP4.5 after 2005	Modern-Era Retrospective Analysis for Research and Applications (MERRA)	MOM4	0.5° × 0.25 – 0.5°, 40 levels
GMAO-062012	CMIP5 historical forcing until 2005 and RCP4.5 after 2005	MERRA	MOM4	0.5° × 0.25 – 0.5°, 40 levels
CFSv2	Climatological seasonal cycle superimposed the observed estimate at the initial time	CFSR	MOM4	0.5° × 0.25 – 0.5°, 40 levels
				Initialisation (DA method)
				Assimilation of SST, temperature and salinity profiles (ensemble optimal interpolation, EnOI)
				Assimilation of SST, temperature and salinity profiles (EnOI)
				CFSR (3DVAR)

interaction, it is suitable for monitoring the surface ocean state changes (Shukla 1998; Deser et al. 2010). Furthermore, SST has been very well observed by satellites since 1982 and SST datasets are commonly used to assess forecasts. We also expect the best predictive skill for SST, as this quantity is used to initialise our hindcasts (Sect. 2.2).

Figure 3 shows ACCs between SSTAs from OISST and NorCPM and the averaged ACCs between SSTAs from OISST and the individual NMME systems at 6- and 12-month lead times. Similar to the NMME systems, NorCPM has high ACCs in the tropical Pacific. Furthermore, NorCPM is skilful (ACCs > 0.6) in the tropical western Atlantic and the Iceland Basin. Overall, the ACCs of NorCPM are higher than the ACCs averaged over the NMME systems in most regions.

It is of great interest to disentangle the part of prediction skill related to natural internal variability from the part driven by the external forcing (Sect. 2.2). To do so, we split the ACC of NorCPM into two parts as follows:

$$ACC(\mathbf{x}_1, \mathbf{y}) = \underbrace{\sqrt{\frac{\text{var}(\mathbf{x}_2)}{\text{var}(\mathbf{x}_1)}} ACC(\mathbf{x}_2, \mathbf{y})}_{\text{Part 1}} + \underbrace{\sqrt{\frac{\text{var}(\mathbf{I})}{\text{var}(\mathbf{x}_1)}} ACC(\mathbf{I}, \mathbf{y})}_{\text{Part 2}} \tag{1}$$

where var(·) is the variance over time, \mathbf{y} is the observed SST, \mathbf{x}_1 is SST taken from the hindcasts initialised by the NorCPM reanalysis (Sect. 2.2), \mathbf{x}_2 is SST taken from the uninitialised hindcasts and $\mathbf{I} = \mathbf{x}_1 - \mathbf{x}_2$ (i.e., the part added by DA). The uninitialised hindcasts are taken from the 30-member free running NorCPM model that is a continuous run from 1980 to 2010 without the DA (named the NorCPM free run). This is equivalent to running hindcasts for each start date from the NorCPM free run. Note that \mathbf{x}_1 is based on the ensemble of 9 members while \mathbf{x}_2 is based on the ensemble of 30 members. A large ensemble is required to average out the internal variability in each ensemble member of the free run to robustly reveal the externally forced variability in the ensemble mean (Solomon et al. 2011). While a large ensemble can be desirable to isolate the initialised component (Scaife and Smith 2018), we have restricted our initialised hindcast ensemble size to 9 because of computational limitations. Part 1 can be considered as the skill due to external forcing and Part 2 can be considered as the added skill due to DA. Note that it is not possible to fully distinguish between skill due to external forcing and DA (Solomon et al. 2011), since the DA also corrects the contribution of external forcing.

Figure 4 presents the parts of the decomposition of the NorCPM ACCs at 6-month and 12-month lead times (top panels in Fig. 3). We find that the prediction skill of NorCPM in the tropical Pacific related to the high prediction skill of ENSO (Luo et al. 2015) is due to the DA. The prediction skill in the tropical western Atlantic and the Iceland

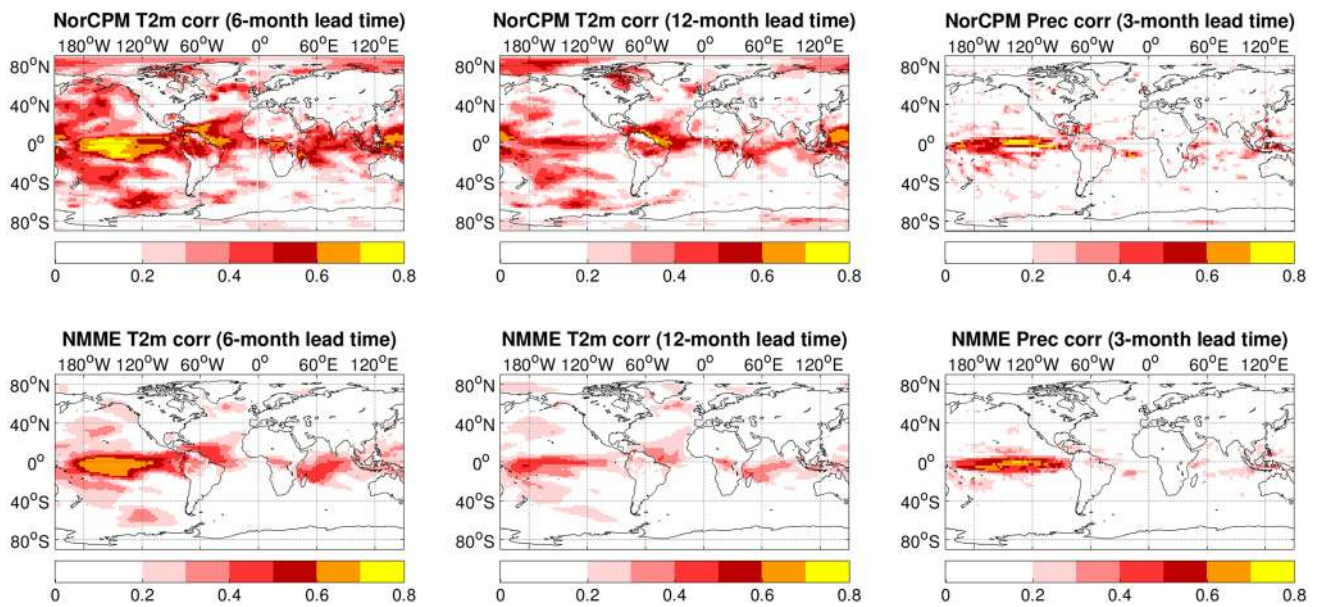


Fig. 1 Top (bottom) panels present the ACCs of NorCPM (the average of the ACCs of the NMME systems) for T2m at 6- and 12-month lead times and precipitation at 3-month lead time

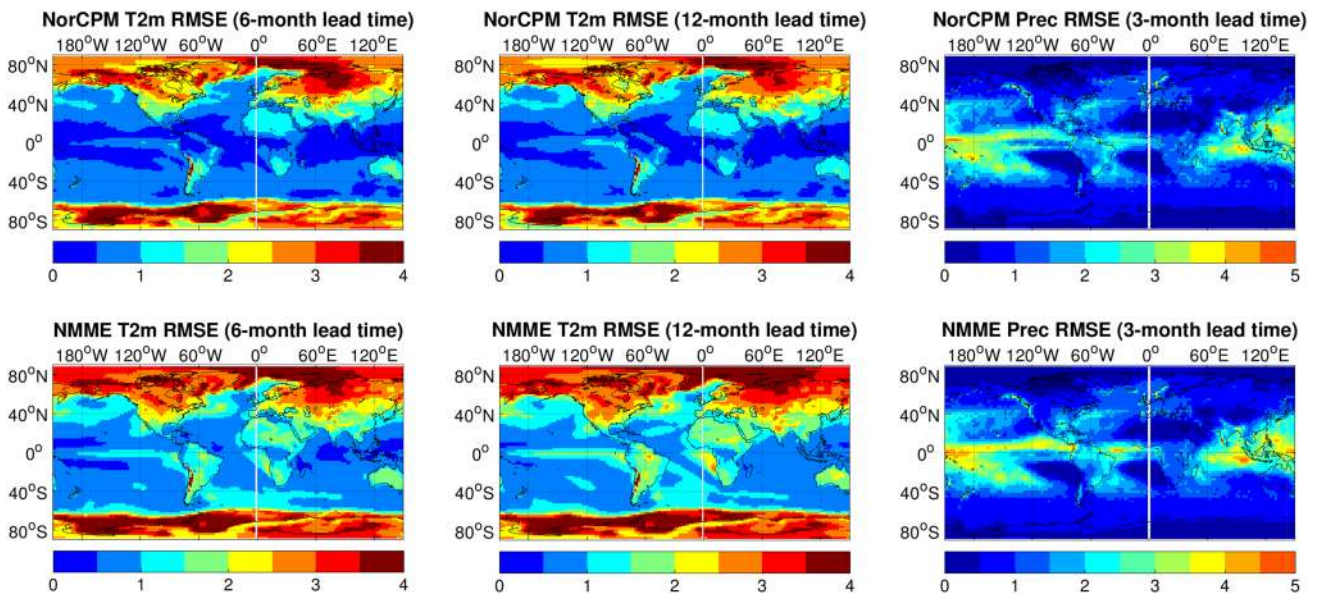


Fig. 2 Top (bottom) panels present the RMSEs of NorCPM (the average of the RMSEs of the NMME systems) for T2m at 6- and 12-month lead times ($^{\circ}\text{C}$) and precipitation at 3-month lead time (mm/day)

Basin is related to both external forcing (left panels in Fig. 4) and the DA (right panels in Fig. 4). Note that most NMME systems include external forcing as well (Table 1). However, NorCPM is based on a climate model, and as such is developed to well capture the climate response to external forcing (Bentsen et al. 2013).

Figure 5 shows RMSEs of NorCPM SSTAs and the averaged RMSEs of SSTAs of the NMME systems at 6- and

12-month lead times. Like the NMME, NorCPM has higher RMSEs in the east-central equatorial Pacific, the North Pacific, the Gulf Stream and the Kuroshio Stream than in other regions, because of the larger (observed and predicted) SST variability in these regions. Nevertheless, the RMSEs in NorCPM are generally lower than the RMSEs averaged over the 13 NMME systems (Table 2) in particular in the Gulf and Kuroshio Streams.

Table 2 Global average of ACCs and RMSEs for SST (°C), T2m (°C) and precipitation (mm/day)

Metric	Variable	Lead time	NorCPM	NMME
ACC	SST	6-month	0.34	0.26
	SST	12-month	0.23	0.19
	T2m	6-month	0.23	0.14
	T2m	12-month	0.18	0.11
	Precipitation	3-month	0.07	0.06
RMSE	SST	6-month	0.52	0.59
	SST	12-month	0.56	0.59
	T2m	6-month	1.52	1.82
	T2m	12-month	1.52	1.93
	Precipitation	3-month	1.08	1.16

The above results show that NorCPM has a skill of global SST predictions higher than the averaged skill of the NMME. In the following, we compare NorCPM to individual NMME systems. To do so, we define four bins that relate to the ACC or RMSE. For the ACC, the bins are distinguished by a significance test (Fisher z-transformation) at a 5% significance level. For the RMSE, we use a threshold of 0.1 °C. These four skill bins are described as follows:

- *SIMILAR* all models have a similar prediction skill (no significant difference of ACCs or less than 0.1 °C difference of RMSEs);
- *BEST* NorCPM leads to a prediction skill higher than or equal to the highest skill of the NMME systems;

- *INTERMEDIATE* NorCPM leads to a prediction skill lower than the highest skill but higher than the lowest skill of the NMME systems;
- *WORST* NorCPM leads to a prediction skill lower than or equal to the lowest skill of the NMME systems.

Figure 6 shows the ranks of NorCPM among the 13 NMME systems with respect to ACC and RMSE. The proportions of the four skill bins are presented in Table 3. The results indicate that NorCPM belongs to the *SIMILAR*/*BEST* skill bins in most oceanic regions. In the tropical western Atlantic and in the region that extends from the Iceland Basin to the Barents Sea, NorCPM belongs to the *BEST* bin for both ACC and RMSE up to 12-month lead time. This analysis reveals that NorCPM is among the better systems of the NMME for predicting global SST at 6- and 12-month lead times. This confirms that skilful seasonal predictions using CGCM can be achieved by initialisation approaches using only SST observations as shown in previous studies (Luo et al. 2005; Keenlyside et al. 2005; Zhu et al. 2017).

Our better skill can be because of better data, better initialisation scheme, or a better model. Use of better data is an unlikely reason, because we use similar SST products to the NMME systems that mostly use more ocean data (Table 1). Without a detailed comparison it is difficult to be conclusive regarding differences in model performance. However, we use a climate model not specifically developed for seasonal prediction, while the NMME systems have been mostly developed for this task. Hence, it is unlikely at least compared to the NMME models that our good performance is because of a superior model. Thus, it is most likely our

Fig. 3 Top left (right) panel represents the ACCs between observed SSTA and SSTA predicted in NorCPM at 6-month (12-month) lead time. Bottom left (right) panel represents the average of the ACCs between observed SSTA and SSTA predicted in the NMME systems at 6-month (12-month) lead time. ACCs are masked as white colour in the regions covered by sea ice

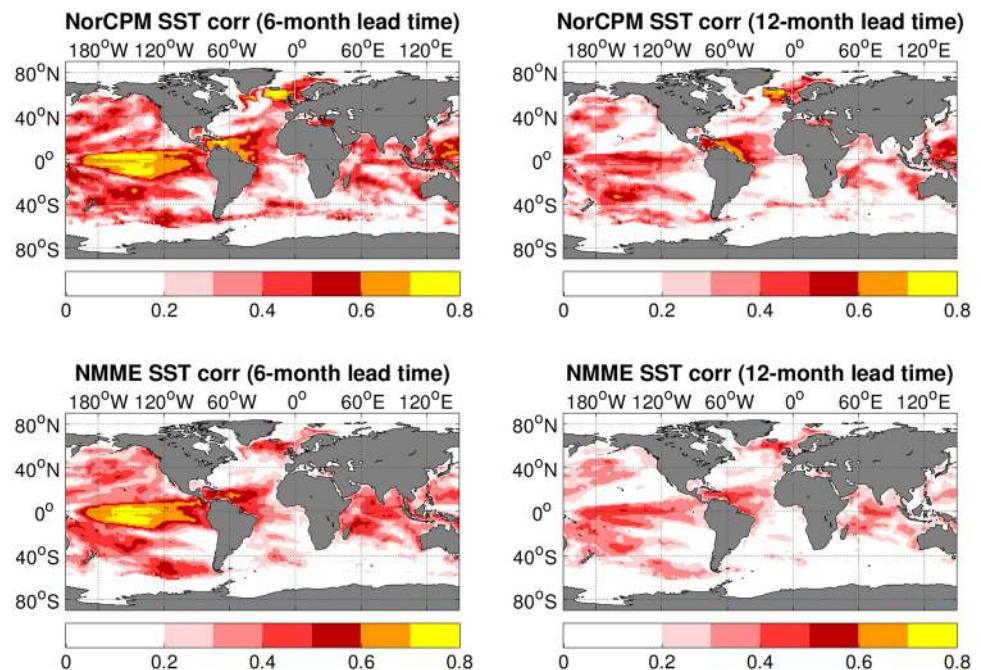


Fig. 4 Parts in the decomposition of the NorCPM ACCs at 6-month (top panels) and 12-month (bottom panels) lead times. Please refer to Eq. (1)

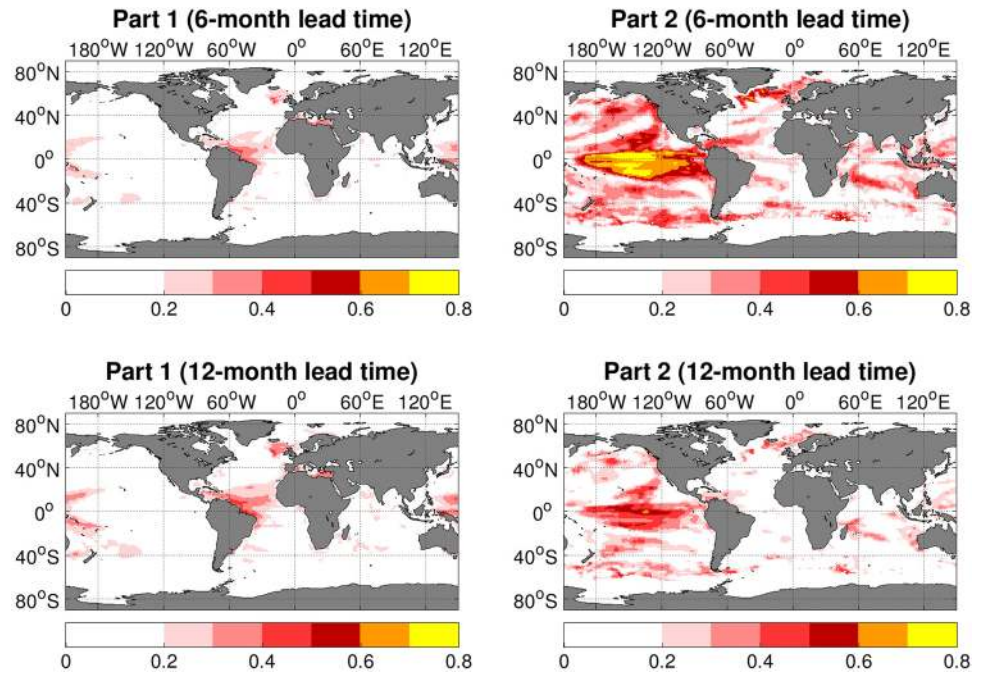
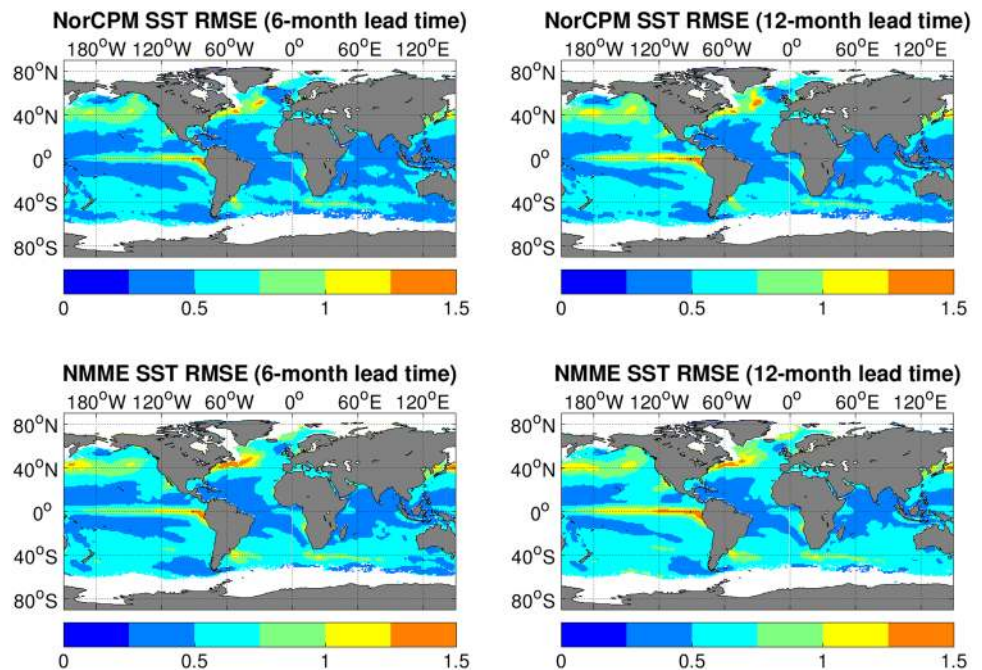


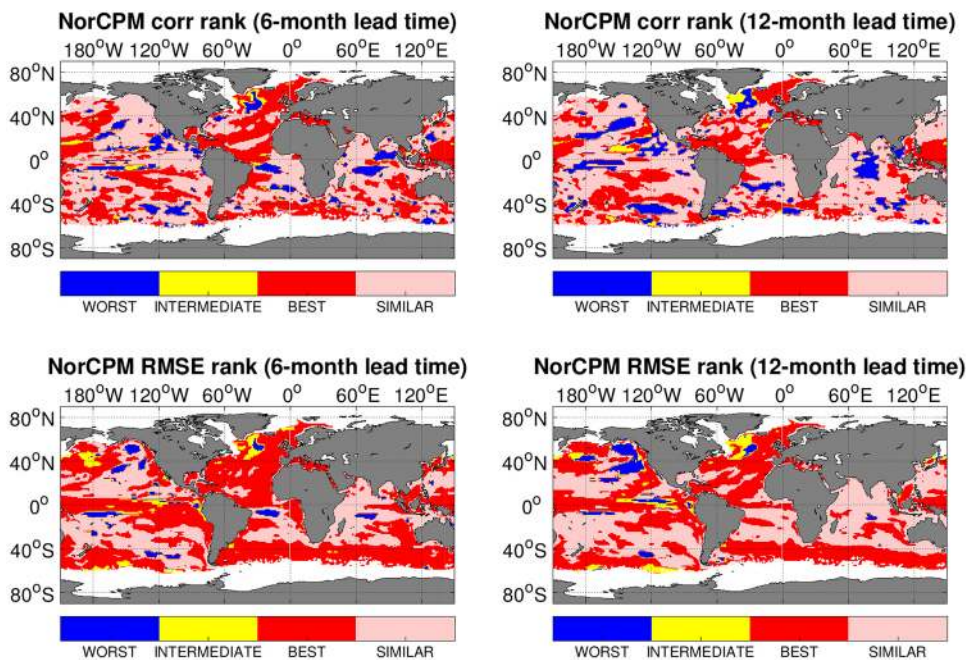
Fig. 5 As in Fig. 3, but for the RMSE ($^{\circ}\text{C}$) of SSTA



initialisation scheme is the main reason for the on average better skill of our system compared to the NMME. The advanced DA method we used (i.e., the EnKF) is a multivariate and flow-dependent DA method (Evensen 2003). When assimilating SST observations, it can effectively propagate information to other ocean state variables than SST (e.g., mixed layer depth and upper ocean heat/salt contents, Counillon et al. 2016) making a better use of SST data to estimate

initial conditions for hindcasts. Furthermore, most NMME systems perform their hindcasts in a full field initialisation framework from uncoupled GCMs and reanalyses produced by the other models (Table 1). This will yield initialisation shock and model drift that will degrade prediction skill. This is not a major issue in our hindcasts, as we use an anomaly initialisation and use the same coupled model to generate initial conditions and perform hindcasts.

Fig. 6 Ranks of NorCPM among the NMME systems for the ACC and RMSE at 6-month or 12-month lead time. Please refer to Sect. 3 for the definition of the four skill bins: SIMILAR, BEST, INTERMEDIATE and WORST



It is worth noting that a low-resolution CFSv2 with SST nudging scheme (Zhu et al. 2017) achieved a skill comparable to the averaged skill of the NMME while the skill of NorCPM is better than the averaged skill of the NMME. The SST nudging and multivariate EnKF schemes work well in regions where two-way ocean atmosphere interaction is strong, such as the tropical Pacific. We expect our scheme can perform better than SST nudging for the following reasons: firstly, the multivariate EnKF scheme can work in regions where there is a strong anti-correlation between surface and subsurface variability and air-sea coupling is weak (e.g., the extra-tropics). Second, the SST nudging scheme is not necessarily able to maintain a consistent relation between temperature and salinity. This can lead to spurious behaviour through impacts on buoyancy (Dunstone and Smith 2010). This can also be important in the tropical Pacific warm pool (Maes et al. 2005) and for ENSO prediction (Hackert et al. 2011). The EnKF captures these physical relations because it makes use of covariances (Counillon et al. 2016). Thus, in general we expect that a nudging scheme would not perform as well as our advanced assimilation scheme, but we have not performed any additional experiments to confirm this.

In the following sections, we will focus on the prediction skill of ENSO in the tropical Pacific and sea ice at the entrance of the Arctic adjacent to the north Atlantic. The tropical Pacific is selected because it is a region where prediction systems perform well and the Arctic where seasonal predictions are less developed. In the ENSO region, we will assess the reason why NorCPM is skilful for ENSO predictions. At the entrance of the Arctic where NorCPM seems

to be one of the better models, we will assess the reasons for the skill of NorCPM in predicting sea ice.

4 ENSO prediction

Nowadays, climate prediction systems can typically provide skilful ENSO prediction 6–9 months ahead (e.g., Zheng et al. 2006; Zhu et al. 2017). One notable exception is the SINTEX-F model, which can successfully predict ENSO up to 2 years ahead (Luo et al. 2015). We consider the Nino3.4 SSTA index (SSTA in the region, 5°S–5°N and 120°W–170°W) for ENSO predictions. The green lines in Fig. 7 show the time series of Nino3.4 SSTA predicted 6 and 12 months ahead by NorCPM. As reported in Jin et al. (2008), the stronger ENSO events are more predictable than weaker or neutral ones. El Niño and La Niña events are well predicted 6 months ahead by NorCPM in terms of their timings and amplitudes. The ensemble spread of

Table 3 The proportions of the four bins related to ACC and RMSE at 6-month and 12-month lead times in Fig. 6

Metric	Lead time	SIMILAR (%)	BEST (%)	INTER-MEDIATE (%)	WORST (%)
ACC	6-month	51	37	2	10
	12-month	57	29	2	12
RMSE	6-month	27	67	3	3
	12-month	59	35	2	4

the ENSO predictions are much larger at 12 months lead time, since forecast uncertainties grow due to the chaotic behaviour of the system. It is found that the ENSO events predicted 12 months ahead have weaker amplitudes than the observations and the phases of some weak ENSO events are shifted. Nevertheless, the observations generally fall within the range of the NorCPM predictions (green shading in Fig. 7) for most ENSO events, indicating that our system is quite reliable. The strongest El Niño of 1997/98 is well predicted 12 months ahead albeit its amplitude is too weak.

We assess the skill of the NorCPM reanalysis (Sect. 2.2)—from which our hindcasts are initialised—in representing the variability of the depth of the 20 °C isotherm (Z20), because realistic fluctuations of thermocline depth (equivalent to warm volume) are critical for ENSO prediction (Smith et al. 1995; Meinen and McPhaden 2000; McPhaden 2003; Zhu et al. 2015a). Figure 8 presents monthly variation of ACCs between Z20 anomalies (averaged over 5°S–5°N) derived from the EN4 dataset (EN4.1.1) and the NorCPM reanalysis over the period of 1985–2010 in the equatorial Pacific. It is found that assimilation of SSTA in NorCPM constrains well Z20 in the equatorial Pacific, apart from August and September over 180°–160°W and February to June over the 160°–80°W. The poor skill in the central equatorial Pacific in August and September is connected to an unrealistic thermocline feedback in our model. The thermocline feedback describes the influence of subsurface temperature variability on SST, mainly through the mean vertical advection of subsurface temperature anomalies (Jin and An 1999). We use the correlation between Z20 and SST anomalies to assess the strength of this feedback. Figure 9 shows the seasonality of SST-Z20 relationship in the observations, a free run of NorCPM (no DA, 30 members) and the reanalysis of NorCPM (30 members). Compared to the free run (top right panel in Fig. 9), the thermocline feedback over 180°–160°W is improved in the reanalysis but is still opposite to that in observations between August and September. The decrease in ACC in the boreal late winter and spring across the eastern equatorial Pacific (Fig. 8) is consistent with the findings of Zhu et al. (2015b). During that period, the thermocline feedback is very weak (Fig. 9). As a consequence of small covariance between SST and subsurface, DA does not effectively propagate the surface information into the ocean interior.

Figures 10 and 11 presents the ACCs and RMSEs between observed and predicted Nino3.4 SSTAs as a function of hindcast lead time for all four season starts. The results indicate that the ACCs and RMSEs of NorCPM are within the range of that of the NMME models. This is consistent with Fig. 6. NorCPM is one of the models that perform best in terms of ACC and RMSE for July start and shows relatively lower RMSEs for all four season starts (Fig. 11). In addition, the ENSO prediction skill in

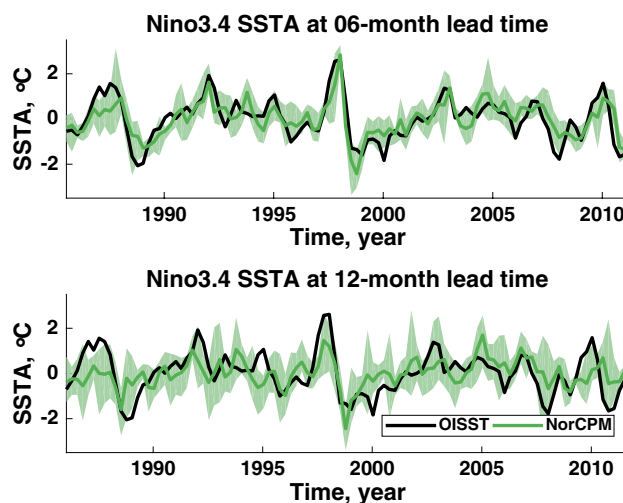


Fig. 7 Time series of Nino3.4 SSTA. The black lines represent observations (OISST) and the green lines represent the NorCPM hindcasts (ensemble means) with the ensemble envelope defined by the minimum and maximum ensemble members (green shading)

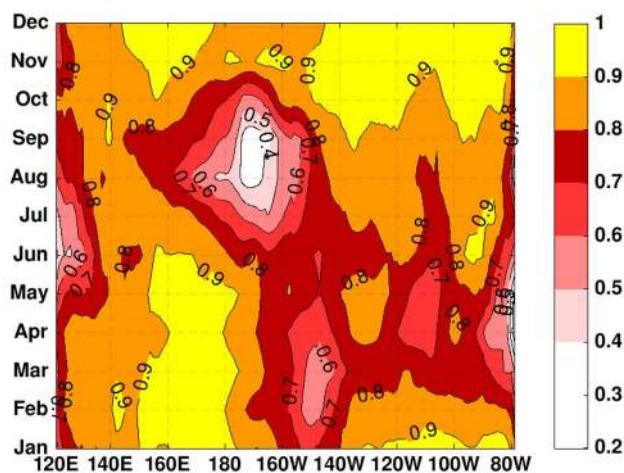
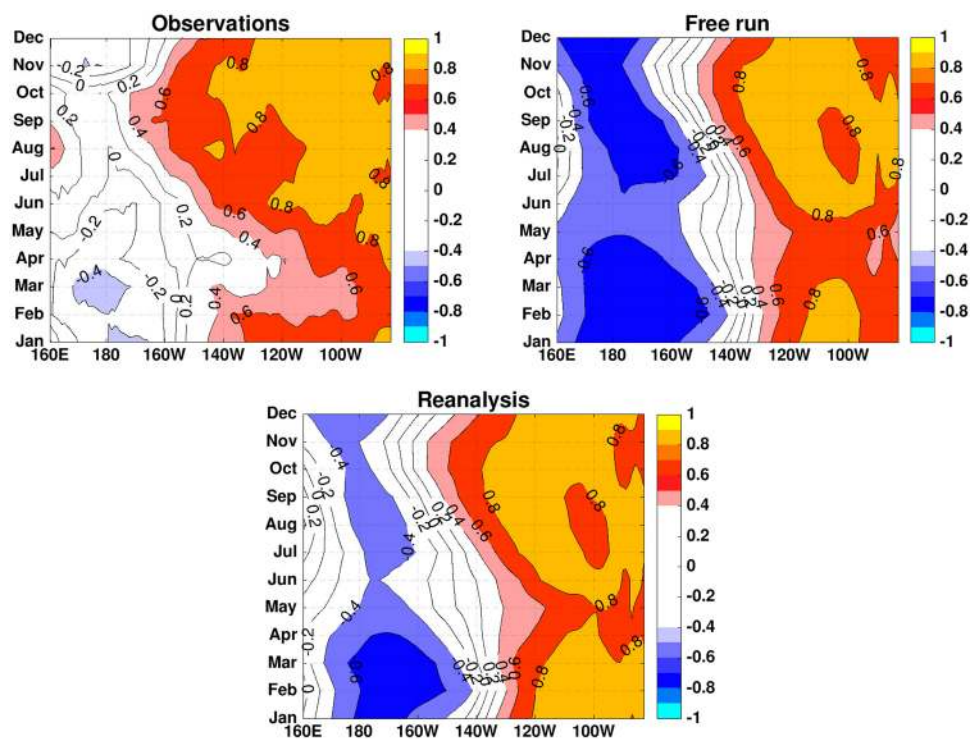


Fig. 8 Seasonality of ACCs between Z20 anomalies (averaged over 5°S–5°N) derived from the EN4.1.1 dataset and the NorCPM reanalysis over the period of 1985–2010 in the equatorial Pacific

all models is season-dependent (Luo et al. 2015; Zhu et al. 2015b, 2017). However, NorCPM has a more pronounced dip in ACC in May (i.e., a more pronounced spring predictability barrier) than the NMME systems (Fig. 10). NorCPM is even out of range of the NMME systems in May, but its ACC reemerges into the ACC range of the NMME systems beyond the boreal spring.

As the May dip of the ENSO predictability occurs for all season starts (Fig. 10), it is likely due to a deficiency in our model in the representing ENSO dynamics rather than in the initial conditions. However, it is unlikely that the Bjerknes feedback during the boreal spring is responsible

Fig. 9 Seasonality of ACCs between Z20 and SST in observations (EN4 and OISST), a free run of NorCPM and the reanalysis of NorCPM



for the May dip in SST prediction skill, as the relation is reasonably captured from February to May (Fig. 12). The thermocline feedback is a more likely factor. In the free run the thermocline feedback is of opposite sign in the western and eastern parts of the Nino3.4 region (170° – 120° W), while in observations the feedback is generally positive but strongest in the east (Fig. 9). The discrepancies between observations and the model are largest from February to May, as the model overestimates both the negative correlation in the western part and the positive correlation in the eastern part of the Nino3.4 region. During the rest of the year, the thermocline feedback strengthens in the east and the negative correlations weaken in the west, and as a result the model and observations agree better (Fig. 8). The poorly simulated relation between Nino3.4 SST and Z20 anomalies during boreal spring can explain the discrepancy between predicted and observed SST anomalies during these months. As discussed above, it can also explain the poorer skill in constraining the subsurface temperature in the central and eastern Pacific from February to May in the reanalysis seen in Fig. 8. A further contributing factor is the sharp reduction in skill in predicting thermocline depth anomalies in February and March in the Nino3.4 region for all four season starts (Fig. 13), as it can take 1–2 months for the SST in this region to respond to the subsurface changes (Zelle et al. 2004). The reason for this reduction skill is not clear but it could be potentially related to the poorly simulated Bjerknes feedback during the second half of the year (Fig. 12), as the warm water volume will adjust with 6–9 months delay to

the covarying off-equatorial wind stress (Jin 1997). In summary, the May dip in predicting SST in the Nino3.4 region is likely caused by the poorly predicted Z20 anomalies in early boreal spring together with an incorrect influence of these Z20 anomalies on SST.

The cause of the model errors leading to the May dip in skill are difficult to conclusively identify because of the complexity of the coupled system. However, we speculate that there is a close relation between these errors and those in the model climatology. During March to May our model shows an excessive rain band south of the equator that extends across the entire Pacific, while there is a dry bias in the equatorial western Pacific as common to many climate models (Fig. i in supplementary material). As a result, the seasonal weakening of the equatorial trade winds is not properly simulated by our model and nor is the development of the seasonal warming of the equatorial eastern Pacific during boreal spring. The observed equatorial seasonal cycle is connected with the southern hemisphere seasonal cycle because the ITCZ remains north of the equator (Chang and Philander 1994; Xie 1994); this connection is weakened by the double ITCZ bias in boreal spring in our model. We can infer from the poor seasonal cycle in equatorial zonal winds and SST (Fig. ii in supplementary material) that the seasonal cycle of equatorial upwelling is also underestimated. (We don't plot the seasonal cycle of vertical velocity as it was not output by the model.) This can explain why the simulated between SST and Z20 anomalies remain strongly related across

Fig. 10 ACCs of Nino3.4 SST as a function of hindcast lead time. The ACCs are computed by different season start hindcasts during 1985–2010 against OISST. The black solid curve is the NorCPM hindcast. The coloured solid curves are the hindcasts of the NMME models. The black dashed line is the persistence forecast

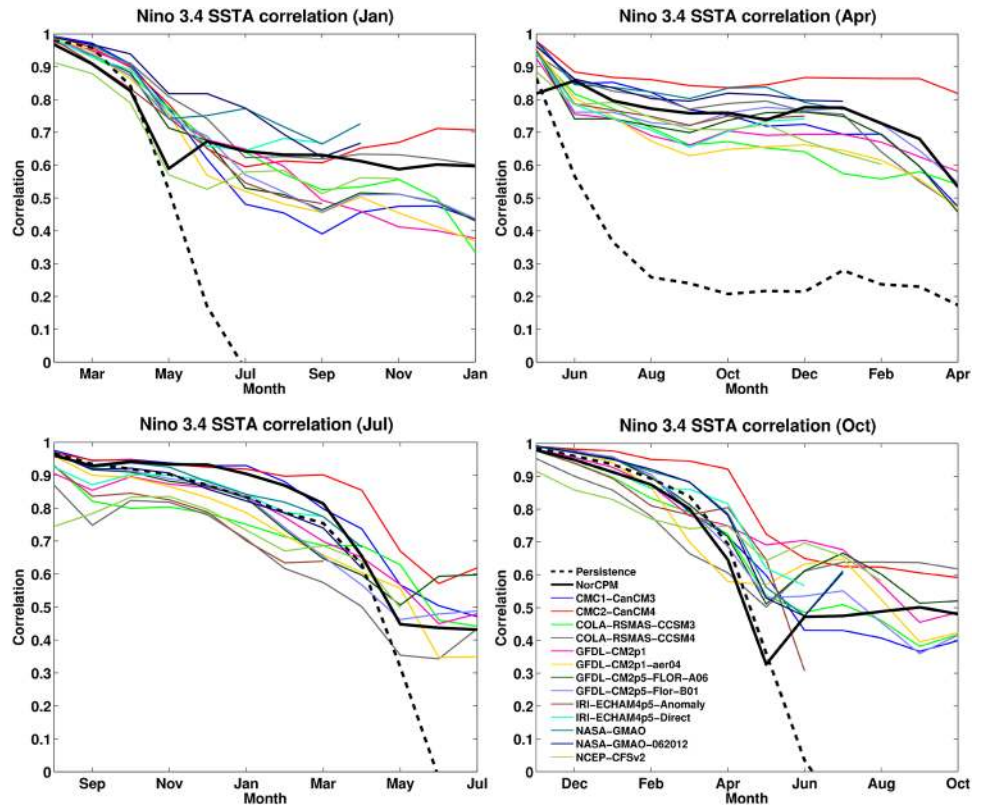
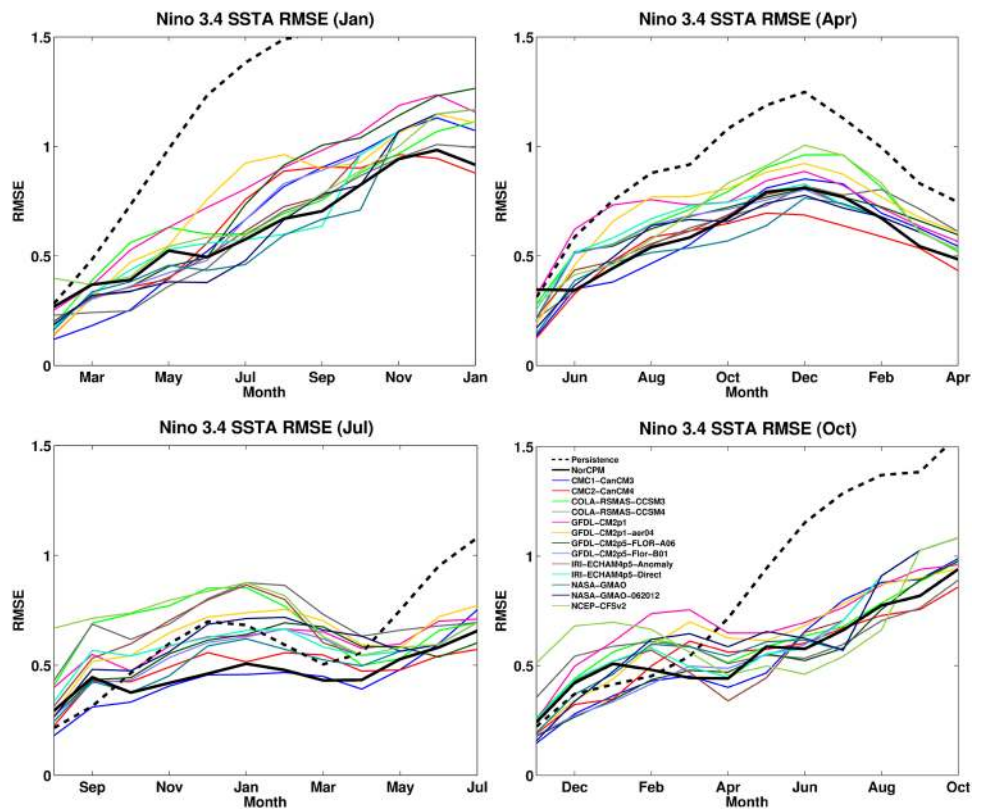


Fig. 11 As in Fig. 10, but for the RMSE (°C) of Nino3.4 SSTA



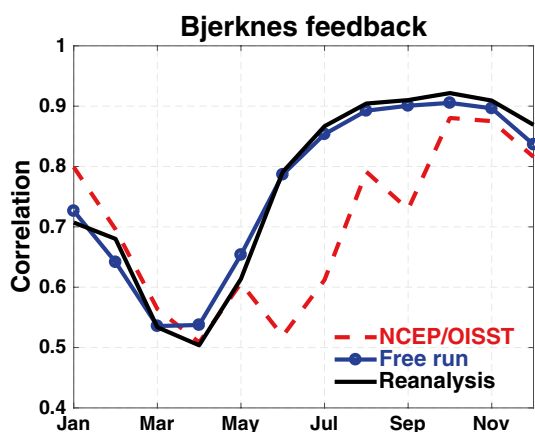


Fig. 12 Bjerknes feedback (correlations between surface zonal wind stress anomalies and the SST gradient) in NCEP/OISST, a free run of NorCPM and the reanalysis of NorCPM. These correlations are computed over the period of 1985–2010

boreal spring, when observations show a weakening of the relation (Fig. 9). We will further assess the impact of model climatology errors on seasonal prediction skill using an anomaly coupled version of the model (Toniazzo and Koseki 2018), but this work is beyond the scope of this current paper.

ENSO can influence other quantities than SST or regions beyond the tropical Pacific (ENSO teleconnections, Luo

et al. 2015). Here, we focus on April hindcasts for the boreal winter (December, January and February; DJF) in which strong ENSO teleconnections have been reported (Wallace and Gutzler 1981). Figure 14 shows the ACCs between observations and NorCPM hindcasts for SSH, precipitation, SLP and geopotential height at 500 hPa (Z500) in the boreal winter. The SSH prediction is skilful in the equatorial Pacific and eastern Indian ocean and the tropical Atlantic. The high prediction skill of precipitation in the equatorial Pacific, Indonesia and Amazon is associated with the good skill in predicting El Niño events. For SLP and Z500, higher ACCs are mostly found in the tropics. Some regions with significant ACCs are found in the north Pacific in good agreement with the teleconnections to ENSO reported by Alexander et al. (2002).

5 Regional Arctic SIE prediction

The Arctic sea ice forecast is of great importance for local communities and stakeholders (McGoodwin 2007; Liu and Kronbak 2010). We expect that a system predicting SST well has some skill in predicting the Arctic sea ice, because SST and sea ice are highly anticorrelated. The top panels of Fig. 15 present the seasonality of determination coefficient (R^2) between detrended SIE and heat content in the upper 300 m (HC300) from the reanalysis of NorCPM for

Fig. 13 ACCs of SST in the Nino3.4 region (black solid line) and Z20 in the Nino3.4 region (red dashed line) as a function of hindcast lead time

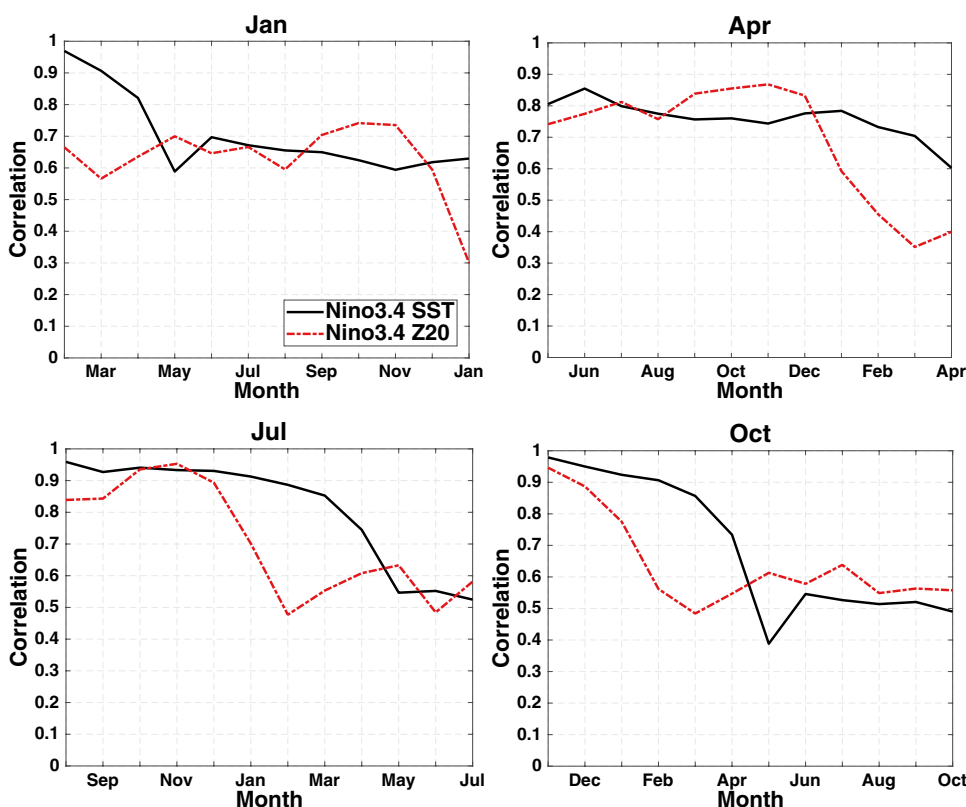
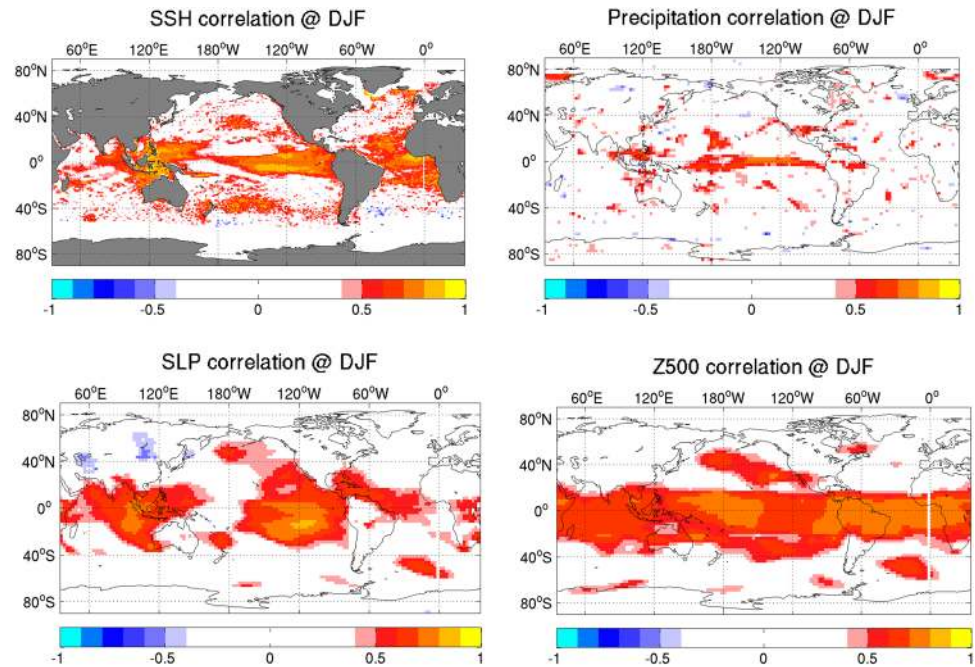


Fig. 14 ACCs for SSH, precipitation, SLP and Z500 in the boreal winter (DJF) between the validation datasets and the NorCPM hindcasts starting on the 15th of April during 1985–2010



the Greenland–Iceland–Norwegian (GIN) Seas, Barents Sea and Labrador Sea. The determination coefficient R^2 is computed over the period of 1985–2010. The SIE is computed for each individual ensemble member as the area sum of grid cells where the sea ice concentration exceeds 15%. Note that the 30-member ensemble mean of SIEs is used in these panels. The results indicate that the variability of SIE is highly related to HC300, in particular in the GIN Seas and in the boreal winter–spring in the Barents and Labrador Seas. This agrees with previous findings (Bitz et al. 2005; Onarheim et al. 2015; Årthun et al. 2017). As shown in Sect. 3, NorCPM is skilful in predicting the SST variability in the region that extends from the Iceland Basin to the Barents Sea. Thus we expect skill in predicting the sea ice variability in the Arctic regions adjacent to the north Atlantic.

The bottom panels of Fig. 15 show the correlation coefficients between detrended SIEs from the hindcasts and HadISST2.1.0.0 for the GIN Seas, Barents Sea and Labrador Sea. We perform the Student's t -test at the significance level of 5% to test statistical significance of the correlation coefficient. The correlation coefficients that are not significantly different from zero are marked by black dots (Fig. 15). The correlation coefficients in the other regions of the Arctic basin are not presented in this paper, since it is mostly not significant. Consistent with the findings of Koenigk and Mikolajewicz (2008), Day et al. (2014) and Bushuk et al. (2017), the highest SIE predictability is found in the basins adjacent to the north Atlantic. Overall, NorCPM provides skilful SIE predictions in the boreal winter and spring while the prediction skill is not significant in the rest of the year (bottom panels of Fig. 15). During the boreal winter and

spring in particular in the Barents Sea, warmer ocean condition yields larger heat loss to the atmosphere that reduces sea ice freezing and smaller sea ice cover (Årthun et al. 2012). In the rest of the year, both the initial state of sea ice and the atmospheric variability have large influence on the SIE variability (Bushuk et al. 2017). This is confirmed by the seasonality of the relation between SIE and HC300 (top panels in Fig. 15) as well.

The Arctic SIE prediction skill of NorCPM is region-dependent (Fig. 15). In the GIN Seas, NorCPM is skilful in the first three lead months from January and April. SIE predictions starting in July and October are not skilful, since SIE also depends on the volume of outflow of sea ice via the Fram Straits (Bitz et al. 2005). In the Barents Sea, NorCPM is skilful in the target months from January to June at lead times of from 1 to 8 months. This reflects that the heat transported by the Gulf Stream and North Atlantic Current, and through the Norwegian Seas into the Barents Sea plays a relevant role in determining the sea ice during the melting season in this region. In the Labrador Sea, the predictions initialised in January are skilful until June, those initialised in April are only skilful in the first 2 months and those initialised in October are skilful in the first three lead months.

6 Conclusions

This study clearly demonstrates that assimilating SST observations with an advanced DA method can be as or even more competitive than the current state of the systems that assimilate more data. We demonstrate this using the NorCPM, and

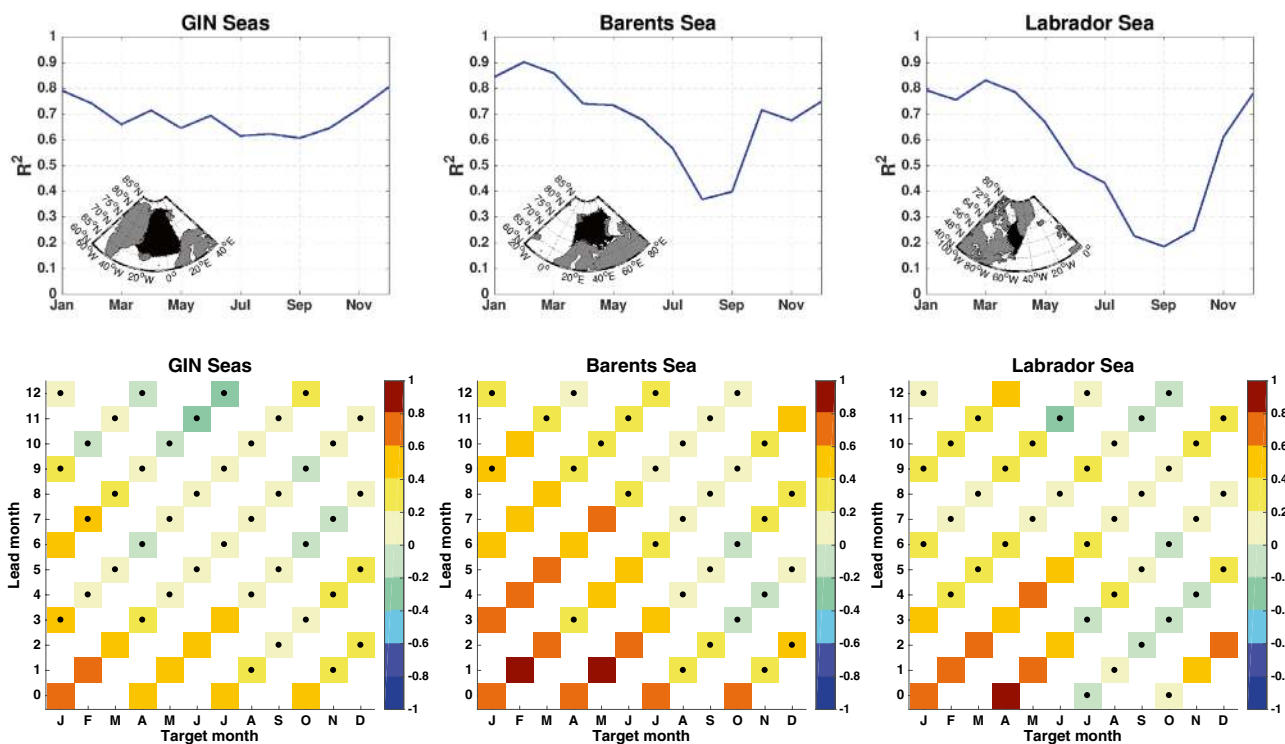


Fig. 15 Top panels present the seasonality of determination coefficient (R^2) between detrended SIE and HC300 from the reanalysis of NorCPM for the GIN Seas, the Barents Sea and the Labrador Sea (black areas). Bottom panels show the correlation coefficients

between detrended SIEs from the hindcasts of NorCPM and observations in these regions. Dots indicate the correlation coefficients that are not statistically significant at the 95% confidence level

thus this study introduces the first seasonal predictions performed with the NorCPM in a real-experiment framework.

The global SST prediction skill (ACC and RMSE) of NorCPM at 6- and 12-month lead times is generally higher than the averaged skill of 13 NMME systems, especially in the tropical western Atlantic and the region extending from the Iceland Basin to the Barents Sea. The skill in these regions is due to the combined effect of external forcing and DA. Like the NMME systems, NorCPM has the highest prediction skill in the ENSO region, which is found mainly due to DA. In addition, NorCPM is found to rank among the better models of the NMME in most regions.

NorCPM can predict well the ENSO events 6 months ahead. Although the ENSO events predicted 12 months ahead are less accurate with respect to their amplitudes and timings, the observations fall mostly within the prediction uncertainties of NorCPM. The high ENSO predictability is due to the fact that NorCPM skilfully constrains warm water volume by assimilation of SSTA. The ENSO prediction skill of NorCPM is season-dependent and NorCPM is one of the better systems for July starts. However, NorCPM has a spring predictability barrier more pronounced than that of the NMME systems. There is a pronounced skill drop in May independent of season starts but the model performance recovers in June. An analysis demonstrates that the

skill drop in May is likely linked to a weak and inconsistent thermocline feedback in NorCPM from February and March. Despite this limitation, NorCPM reproduces reasonably well ENSO teleconnection patterns in the boreal winter and provide skilful predictions beyond the tropical Pacific.

NorCPM shows some skill in predicting SIE in the GIN Seas, the Barents Sea and the Labrador Sea from January and April. In these regions, the SIE predictability in the boreal winter and spring is highly related to the initialisation of upper ocean heat content, which is consistent with the findings of Bushuk et al. (2017). In addition, the SIE variability is most predictable in the Barents Sea where NorCPM is skilful in the months between January and June and with up to 8 months lead time.

In the future, we are going to include oceanic subsurface data (such as Argo float data) into the initialisation scheme of NorCPM for seasonal-to-decadal predictions, since such data are crucial to constrain the ocean vertical structure (Wang et al. 2017). It will be interesting to demonstrate whether and at which time scales using more data in NorCPM leads to higher prediction skill. On the other hand, Bushuk et al. (2017) demonstrated that both subsurface ocean observations and sea ice thickness observations are necessary to predict Arctic SIE on seasonal time scale. Kimrritz et al. (2018) recently investigated the optimal setting

for assimilating sea ice concentration observations into Nor-CPM and showed a great potential of strongly ocean–sea ice coupled DA. We will, in addition to SST, assimilate sea ice observations into NorCPM, with the expectation of improving the seasonal prediction skill of Arctic sea ice.

Acknowledgements This study was co-funded by the Center for Climate Dynamics at the Bjerknes Center, the Norwegian Research Council under the EPOCASA (229774/E10) and SFE (270733) research projects, the NordForsk under the Nordic Centre of Excellence (ARCPATH, 76654), and the Trond Mohn Foundation under the project number BFS2018TMT01. NK, LS, and FC also acknowledge support from the ERC STERCP project (Grant Agreement No. 648982). This work received grants for computer time from the Norwegian Program for supercomputer (NOTUR2, NN9039K and NN9385K) and storage grants (NORSTORE, NS9039K and NS9207K). We acknowledge the agencies that support the NMME-Phase II system, and we thank the climate modelling groups (Environment Canada, NASA, NCAR, NOAA/GFDL, NOAA/NCEP, and University of Miami) for producing and making available their model output. NOAA/NCEP, NOAA/CTB, and NOAA/CPO jointly provided coordinating support and led development of the NMME-Phase II system.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Alexander MA, Bladé I, Newman M, Lanzante JR, Lau N-C, Scott JD (2002) The atmospheric bridge: the influence of ENSO teleconnections on air–sea interaction over the global oceans. *J Clim* 15:2205–2231. [https://doi.org/10.1175/1520-0442\(2002\)015<2205:TABTIO>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<2205:TABTIO>2.0.CO;2)
- Ammann CM, Meehl GA, Washington WM, Zender CS (2003) A monthly and latitudinally varying volcanic forcing dataset in simulations of 20th century climate. *Geophys Res Lett*. <https://doi.org/10.1029/2003GL016875>
- Årthun M, Eldevik T, Smedsrud LH, Skagseth O, Ingvaldsen RB (2012) Quantifying the influence of Atlantic heat on Barents Sea ice variability and retreat. *J Clim* 25:4736–4743. <https://doi.org/10.1175/JCLI-D-11-00466.1>
- Årthun M, Eldevik T, Viste E, Drange H, Furevik T, Johnson HL, Keenlyside NS (2017) Skillful prediction of northern climate provided by the ocean. *Nat Commun*. <https://doi.org/10.1038/ncomms15875>
- Balmaseda M, Anderson D (2009) Impact of initialization strategies and observations on seasonal forecast skill. *Geophys Res Lett*. <https://doi.org/10.1029/2008GL035561>
- Balmaseda MA, Alves OJ, Arribas A, Awaji T, Behringer DW, Ferry N, Fujii Y, Lee T, Rienecker M, Rosati T, Stammer D (2009) Ocean initialization for seasonal forecasts. *Oceanography*. <https://doi.org/10.5670/oceanog.2009.73>
- Barnston AG, Glantz MH, He Y (1999) Predictive skill of statistical and dynamical climate models in SST forecasts during the 1997–98 El Niño Episode and the 1998 La Niña onset. *Bull Am Meteorol Soc* 80:217–244. [https://doi.org/10.1175/1520-0477\(1999\)080<0217:PSOSAD>2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080<0217:PSOSAD>2.0.CO;2)
- Bellenger H, Guilyardi E, Leloup J, Lengaigne M, Vialard J (2014) ENSO representation in climate models: from CMIP3 to CMIP5. *Clim Dyn* 42:1999–2018. <https://doi.org/10.1007/s00382-013-1783-z>
- Bentsen M, Bethke I, Debernard JB, Iversen T, Kirkevåg A, Seland O, Drange H, Roelandt C, Seierstad IA, Hoose C, Kristjánsson JE (2013) The Norwegian Earth system model, NorESM1–part 1: description and basic evaluation of the physical climate. *Geosci Model Dev* 6:687–720. <https://doi.org/10.5194/gmd-6-687-2013>
- Bitz CM, Holland MM, Hunke EC, Moritz RE (2005) Maintenance of the sea-ice edge. *J Clim* 18:2903–2921. <https://doi.org/10.1175/JCLI3428.1>
- Bleck R, Rooth C, Hu D, Smith LT (1992) Salinity-driven thermocline transients in a wind- and thermohaline-forced isopycnic coordinate model of the North Atlantic. *J Phys Oceanogr* 22:1486–1505. [https://doi.org/10.1175/1520-0485\(1992\)022<1486:SDTTIA>2.0.CO;2](https://doi.org/10.1175/1520-0485(1992)022<1486:SDTTIA>2.0.CO;2)
- Borovikov A, Cullather R, Kovach R, Marshak J, Vernieres G, Vikhliayev Y, Zhao B, Li Z (2017) GEOS-5 seasonal forecast system. *Clim Dyn*. <https://doi.org/10.1007/s00382-017-3835-2>
- Bushuk M, Msadek R, Winton M, Vecchi GA, Gudgel R, Rosati A, Yang X (2017) Skillful regional prediction of Arctic sea ice on seasonal timescales. *Geophys Res Lett* 44:4953–4964. <https://doi.org/10.1002/2017GL073155>
- Carrasi A, Bocquet M, Bertino L, Evensen G (2018) Data assimilation in the geosciences: an overview of methods, issues, and perspectives. *Wiley Interdiscip Rev Clim Change* 9:e535. <https://doi.org/10.1002/wcc.535>
- Chang P, Philander SG (1994) A coupled ocean–atmosphere instability of relevance to the seasonal cycle. *J Atmos Sci* 51:3627–3648. [https://doi.org/10.1175/1520-0469\(1994\)051<3627:ACOIOR>2.0.CO;2](https://doi.org/10.1175/1520-0469(1994)051<3627:ACOIOR>2.0.CO;2)
- Counillon F, Bethke I, Keenlyside N, Bentsen M, Bertino L, Zheng F (2014) Seasonal-to-decadal predictions with the ensemble Kalman filter and the Norwegian Earth System Model: a twin experiment. *Tellus A* 66:1–21. <https://doi.org/10.3402/tellusa.v66.21074>
- Counillon F, Keenlyside N, Bethke I, Wang Y, Billeau S, Shen ML, Bentsen M (2016) Flow-dependent assimilation of sea surface temperature in isopycnal coordinates with the Norwegian Climate Prediction Model. *Tellus A* 68:1–17. <https://doi.org/10.3402/tellusa.v68.32437>
- Craig AP, Vertenstein M, Jacob R (2012) A new flexible coupler for earth system modeling developed for CCSM4 and CESM1. *Int J High Perform Comput Appl* 26:31–42. <https://doi.org/10.1177/1094342011428141>
- Day JJ, Tietsche S, Hawkins E (2014) Pan-Arctic and regional sea ice predictability: initialization month dependence. *J Clim* 27:4371–4390. <https://doi.org/10.1175/JCLI-D-13-00614.1>
- Delworth TL, Broccoli AJ, Rosati A, Stouffer RJ, Balaji V, Beesley JA, Cooke WF, Dixon KW, Dunne J, Dunne KA, Durachta JW, Findell KL, Ginoux P, Gnanadesikan A, Gordon CT, Griffies SM, Gudgel R, Harrison MJ, Held IM, Hemler RS, Horowitz LW, Klein SA, Knutson TR, Kushner PJ, Langenhorst AR, Lee H-C, Lin S-J, Lu J, Malyshev SL, Milly PCD, Ramaswamy V, Russell J, Schwarzkopf MD, Shevliakova E, Sirutis JJ, Spelman MJ, Stern WF, Winton M, Wittenberg AT, Wyman B, Zeng F, Zhang R (2006) GFDL’s CM2 global coupled climate models. Part I: Formulation and simulation characteristics. *J Clim* 19:643–674. <https://doi.org/10.1175/JCLI3629.1>
- Deser C, Alexander MA, Xie S-P, Phillips AS (2010) Sea surface temperature variability: patterns and mechanisms. *Ann Rev Mar Sci* 2:115–143. <https://doi.org/10.1146/annurev-marine-120408-151453>
- DeWitt DG (2005) Retrospective forecasts of interannual sea surface temperature anomalies from 1982 to present using a directly

- coupled atmosphere–ocean general circulation model. *Mon Weather Rev* 133:2972–2995. <https://doi.org/10.1175/MWR3016.1>
- Dimet FL, Olivier T (1986) Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus A* 38A:97–110. <https://doi.org/10.1111/j.1600-0870.1986.tb00459.x>
- Ding H, Newman M, Alexander MA, Wittenberg AT (2018) Skillful climate forecasts of the tropical Indo-Pacific Ocean using model-analogs. *J Clim* 31:5437–5459. <https://doi.org/10.1175/JCLI-D-17-0661.1>
- Doblas-Reyes FJ, Garcia-Serrano J, Lienert F, Biescas AP, Rodrigues LRL (2013) Seasonal climate predictability and forecasting: status and prospects. *Wiley Interdiscip Rev Clim Change* 4:245–268. <https://doi.org/10.1002/wcc.217>
- Dunstone NJ, Smith DM (2010) Impact of atmosphere and sub-surface ocean data on decadal climate prediction. *Geophys Res Lett*. <https://doi.org/10.1029/2009GL041609>
- Evensen G (2003) The ensemble Kalman filter: theoretical formulation and practical implementation. *Ocean Dyn* 53:343–367. <https://doi.org/10.1007/s10236-003-0036-9>
- Gent PR, Danabasoglu G, Donner LJ, Holland MM, Hunke EC, Jayne SR, Lawrence DM, Neale RB, Rasch PJ, Vertenstein M, Worley PH, Yang ZL, Zhang M (2011) The community climate system model version 4. *J Clim* 24:4973–4991. <https://doi.org/10.1175/2011JCLI4083.1>
- Gleixner S, Keenlyside NS, Demissie TD, Counillon F, Wang Y, Viste E (2017) Seasonal predictability of Kiremt rainfall in coupled general circulation models. *Environ Res Lett* 12:114016. <https://doi.org/10.1088/1748-9326/aa8cfa>
- Good SA, Martin MJ, Rayner NA (2013) EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *J Geophys Res Oceans* 118:6704–6716. <https://doi.org/10.1002/2013JC009067>
- Gunda T, Bazuin J T, Nay J, Yeung K L (2017) Impact of seasonal forecast use on agricultural income in a system with varying crop costs and returns: an empirically-grounded simulation. *Environ Res Lett* 12:034001. <http://stacks.iop.org/1748-9326/12/i=3/a=034001>
- Hackert E, Ballabrera-Poy J, Busalacchi AJ, Zhang R-H, Murtugudde R (2011) Impact of sea surface salinity assimilation on coupled forecasts in the tropical Pacific. *J Geophys Res Oceans*. <https://doi.org/10.1029/2010JC006708>
- Hamil TM, Snyder C (2000) A hybrid ensemble Kalman filter-3D variational analysis scheme. *Mon Weather Rev* 128:2905–2919. [https://doi.org/10.1175/1520-0493\(2000\)128<2905:AHEKFV>2.0.CO;2](https://doi.org/10.1175/1520-0493(2000)128<2905:AHEKFV>2.0.CO;2)
- Holland MM, Bailey DA, Briegleb BP, Light B, Hunke E (2012) Improved sea ice shortwave radiation physics in CCSM4: the impact of melt ponds and aerosols on arctic sea ice. *J Clim* 25:1413–1430. <https://doi.org/10.1175/JCLI-D-11-00078.1>
- Hurtt GC, Chini LP, Frolking S, Betts R, Feddema J, Fischer G, Goldewijk KK, Hibbard K, Janetos A, Jones C, Kinderman G, Kinoshita T, Riahi K, Shevliakova E, Smith S, Stehfest E, Thomson A, Thornton P, van Vuuren DP, Wang Y (2009) Harmonization of global land-use scenarios for the period 1500–2100 for IPCC-AR5. *iLEAPS Newslett* 7: 6–8
- Infanti JM, Kirtman BP (2016) Prediction and predictability of land and atmosphere initialized CCSM4 climate forecasts over North America. *J Geophys Res Atmos* 121:12690–12701. <https://doi.org/10.1002/2016JD024932>
- Ji M, Leetmaa A, Kousky VE (1996) Coupled model predictions of ENSO during the 1980s and the 1990s at the National Centers for Environmental Prediction. *J Clim* 9:3105–3120. [https://doi.org/10.1175/1520-0442\(1996\)009<3105:CMPOED>2.0.CO;2](https://doi.org/10.1175/1520-0442(1996)009<3105:CMPOED>2.0.CO;2)
- Jin F-F (1997) An equatorial ocean recharge paradigm for ENSO. Part I: Conceptual model. *J Atmos Sci* 54:811–829. [https://doi.org/10.1175/1520-0469\(1997\)054<0811:AEORPF>2.0.CO;2](https://doi.org/10.1175/1520-0469(1997)054<0811:AEORPF>2.0.CO;2)
- Jin F-F, An S-I (1999) Thermocline and zonal advective feedbacks within the equatorial ocean recharge oscillator model for ENSO. *Geophys Res Lett* 26:2989–2992. <https://doi.org/10.1029/1999GL002297>
- Jin EK, Kinter JL, Wang B, Park C-K, Kang I-S, Kirtman BP, Kug J-S, Kumar A, Luo J-J, Schemm J, Shukla J, Yamagata T (2008) Current status of ENSO prediction skill in coupled ocean-atmosphere models. *Clim Dyn* 31:647–664. <https://doi.org/10.1007/s00382-008-0397-3>
- Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Leetmaa A, Reynolds R, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Jenne R, Joseph D (1996) The NCEP/NCAR 40-year reanalysis project. *Bull Am Meteorol Soc* 77:437–471. [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2)
- Keenlyside N, Latif M, Botzet M, Jungclaus J, Schulzweida U (2005) A coupled method for initializing El Niño Southern Oscillation forecasts using sea surface temperature. *Tellus A* 57:340–356. <https://doi.org/10.1111/j.1600-0870.2005.00107.x>
- Kimmritz M, Counillon F, Bitz C, Massonnet F, Bethke I, Gao Y (2018) Optimising assimilation of sea ice concentration in an Earth system model with a multicategory sea ice model. *Tellus A Dyn Meteorol Oceanogr* 70:1435945. <https://doi.org/10.1080/16000870.2018.1435945>
- Kirkevåg A, Iversen T, Seland Ø, Hoose C, Kristjánsson JE, Struthers H, Ekman AML, Ghan S, Griesfeller J, Nilsson ED, Schulz M (2013) Aerosol-climate interactions in the Norwegian Earth System—NorESM1-M. *Geosci Model Dev* 6:207–244. <https://doi.org/10.5194/gmd-6-207-2013>
- Kirtman BP, Min D (2009) Multimodel ensemble ENSO prediction with CCSM and CFS. *Mon Weather Rev* 137:2908–2930. <https://doi.org/10.1175/2009MWR2672.1>
- Kirtman B, Power S, Adedoyin J, Boer G, Bojariu R, Camilloni I, Doblas-Reyes F, Fiore A, Kimoto M, Meehl G, Prather M, Sarr A, Schär C, Sutton R, van Oldenborgh G, Vecchi G, Wang H (2013) Near-term climate change: projections and predictability. In: Intergovernmental Panel on Climate Change (ed) *Climate change 2013: physical science basis*. Cambridge University Press, Cambridge, pp 953–1028. <https://doi.org/10.1017/CBO9781107415324.023>
- Kirtman BP, Min D, Infanti JM, Kinter JL, Paolino DA, Zhang Q, van den Dool H, Saha S, Mendez MP, Becker E, Peng P, Tripp P, Huang J, DeWitt DG, Tippet MK, Barnston AG, Li S, Rosati A, Schubert SD, Rienecker M, Suarez M, Li ZE, Marshak J, Lim Y-K, Tribbia J, Pegion K, Merryfield WJ, Denis B, Wood EF (2014) The North American multimodel ensemble: phase-1 seasonal-to-interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull Am Meteorol Soc* 95:585–601. <https://doi.org/10.1175/BAMS-D-12-00050.1>
- Koenigk T, Mikolajewicz U (2008) Seasonal to interannual climate predictability in mid and high northern latitudes in a global coupled model. *Clim Dyn* 32:783. <https://doi.org/10.1007/s00382-008-0419-1>
- Krishnamurti TN, Kishtawal CM, LaRow TE, Bacchiochi DR, Zhang Z, Williford CE, Gadgil S, Surendran S (1999) Improved weather and seasonal climate forecasts from multimodel superensemble. *Science* 285:1548–1550. <https://doi.org/10.1126/science.285.5433.1548>
- Kumar A, Zhu J (2018) Spatial variability in seasonal prediction skill of SSTs: inherent predictability or forecast errors? *J Clim* 31:613–621. <https://doi.org/10.1175/JCLI-D-17-0279.1>

- Kumar A, Wang H, Xue Y, Wang W (2014) How much of monthly subsurface temperature variability in the equatorial Pacific can be recovered by the specification of sea surface temperatures? *J Clim* 27:1559–1577. <https://doi.org/10.1175/JCLI-D-13-00258.1>
- Lamarque J-F, Bond TC, Eyring V, Granier C, Heil A, Klimont Z, Lee D, Liousse C, Mieville A, Owen B, Schultz MG, Shindell D, Smith SJ, Stehfest E, Van Aardenne J, Cooper OR, Kainuma M, Mahowald N, McConnell JR, Naik V, Riahi K, van Vuuren DP (2010) Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application. *Atmos Chem Phys* 10:7017–7039. <https://doi.org/10.5194/acp-10-7017-2010>
- Lawrence DM, Oleson KW, Flanner MG, Thornton PE, Swenson SC, Lawrence PJ, Zeng X, Yang Z-L, Levis S, Sakaguchi K, Bonan GB, Slater AG (2011) Parameterization improvements and functional and structural advances in version 4 of the community land model. *J Adv Model Earth Syst* 3:M03001. <https://doi.org/10.1029/2011MS000045>
- Lean J, Rottman G, Harder J, Kopp G (2005) *SORCE* contributions to new understanding of global change and solar variability. *Sol Phys* 230:27–53. <https://doi.org/10.1007/s11207-005-1527-2>
- Liu M, Kronbak J (2010) The potential economic viability of using the Northern Sea Route (NSR) as an alternative route between Asia and Europe. *J Transp Geogr* 18:434–444. <https://doi.org/10.1016/j.jtrangeo.2009.08.004>
- Luo J-J, Masson S, Behera S, Shingu S, Yamagata T (2005) Seasonal climate predictability in a coupled OAGCM using a different approach for ensemble forecasts. *J Clim* 18:4474–4497. <https://doi.org/10.1175/JCLI3526.1>
- Luo J-J, Yuan C, Sasaki W, Behera S K, Masumoto Y, Yamagata T, Lee J-Y, Masson S (2015) Chapter 3: Current status of intraseasonal-seasonal-to-interannual prediction of the Indo-Pacific climate, pp. 63–107. World Scientific, Singapore. https://doi.org/10.1142/9789814696623_0003
- Maes C, Picaut J, Belamari S (2005) Importance of the salinity barrier layer for the buildup of El Niño. *J Clim* 18:104–118. <https://doi.org/10.1175/JCLI-3214.1>
- McGoodwin JR (2017) Effects of climatic variability on three fishing economies in high-latitude regions: implications for fisheries policies. *Mar Policy* 31:40–55
- McPhaden MJ (2003) Tropical Pacific ocean heat content variations and ENSO persistence barriers. *Geophys Res Lett* 30:1995–1998. <https://doi.org/10.1029/2003GL016872>
- Meinen CS, McPhaden MJ (2000) Observations of warm water volume changes in the equatorial Pacific and their relationship to El Niño and La Niña. *J Clim* 13:3551–3559. [https://doi.org/10.1175/1520-0442\(2000\)013<3551:OOWWVC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<3551:OOWWVC>2.0.CO;2)
- Merryfield WJ, Lee W-S, Boer GJ, Kharin VV, Scinocca JF, Flato GM, Ajayamohan RS, Fyfe JC, Tang Y, Polavarapu S (2013) The Canadian seasonal to interannual prediction system. Part I: Models and initialization. *Mon Weather Rev* 141:2910–2945. <https://doi.org/10.1175/MWR-D-12-00216.1>
- Oleson KW, Lawrence DM, Bonan GB, Flanner MG, Kluzek E, Lawrence PJ, Levis S, Swenson SC, Thornton PE, Dai A, Decker M, Dickinson R, Feddema J, Heald CL, Hoffman F, Lamarque J-F, Mahowald N, Niu G-Y, Qian T, Randerson J, Running S, Sakaguchi K, Slater A, Stöckli R, Wang A, Yang Z-L, Zeng X, Zeng X (2010) Technical description of version 4.0 of the Community Land Model (CLM), Technical Report. NCAR/TN-478+STR. National Center for Atmospheric Research, Boulder
- Onarheim IH, Eldevik T, Årthun M, Ingvaldsen RB, Smedsrud LH (2015) Skillful prediction of Barents Sea ice cover. *Geophys Res Lett* 42:5364–5371. <https://doi.org/10.1002/2015GL064359>
- Palmer TN, Alessandri A, Andersen U, Cantelaube P, Davey M, Décluse P, Déqué M, Diez E, Doblas-Reyes FJ, Feddersen H, Graham R, Gualdi S, Guérémy J-F, Hagedorn R, Hoshen M, Keenlyside N, Latif M, Lazar A, Maisonnave E, Marletto V, Morse AP, Orfila B, Rogel P, Terres J-M, Thomson MC (2004) Development of a European multimodel ensemble system for seasonal-to-interannual prediction (demeter). *Bull Am Meteorol Soc* 85:853–872. <https://doi.org/10.1175/BAMS-85-6-853>
- Paolino DA, Kinter JL, Kirtman BP, Min D, Straus DM (2012) The impact of land surface and atmospheric initialization on seasonal forecasts with CCSM. *J Clim* 25:1007–1021. <https://doi.org/10.1175/2011JCLI3934.1>
- Penny S, Akella S, Alves O, Bishop C, Buehner M, Chevallier M, Counillon F, Draper C, Frolov S, Fujii Y, Karspeck A, Kumar A, Laloyaux P, Mahfouf J-F, Martin M, Peña M, de Rosnay P, Subramanian A, Tardif R, Wang Y, Wu X (2017) Coupled data assimilation for integrated earth system analysis and prediction: goals, challenges and recommendations, Technical Report. WWRP 2017-3, World Meteorological Organization (WMO). https://www.wmo.int/pages/prog/arep/wwrp/new/documents/Final_WWRP_2017_3_27_July.pdf. Accessed 18 July 2019
- Reynolds RW, Rayner NA, Smith TM, Stokes DC, Wang W (2002) An improved in situ and satellite SST analysis for climate. *J Clim* 15:1609–1625. [https://doi.org/10.1175/1520-0442\(2002\)015<1609:AIISAS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2)
- Richter I (2015) Climate model biases in the eastern tropical oceans: causes, impacts and ways forward. *Wiley Interdiscip Rev Clim Change* 6:345–358. <https://doi.org/10.1002/wcc.338>
- Saha S, Nadiga S, Thiaw C, Wang J, Wang W, Zhang Q, Van den Dool HM, Pan H-L, Moorthi S, Behringer D, Stokes D, Peña M, Lord S, White G, Ebisuzaki W, Peng P, Xie P (2006) The NCEP climate forecast system. *J Clim* 19:3483–3517. <https://doi.org/10.1175/JCLI3812.1>
- Saha S, Moorthi S, Wu X, Wang J, Nadiga S, Tripp P, Behringer D, Hou Y-T, Chuang H-y, Iredell M, Ek M, Meng J, Yang R, Mendez MP, van den Dool H, Zhang Q, Wang W, Chen M, Becker E (2014) The NCEP climate forecast system version 2. *J Clim* 27:2185–2208. <https://doi.org/10.1175/JCLI-D-12-00823.1>
- Scaife AA, Smith D (2018) A signal-to-noise paradox in climate science. *npj Clim Atmos Sci* 1:28. <https://doi.org/10.1038/s41612-018-0038-4>
- Shukla J (1998) Predictability in the midst of chaos: a scientific basis for climate forecasting. *Science* 282:728–731. <https://doi.org/10.1126/science.282.5389.728>
- Smith TM, Barnston AG, Ji M, Chelliah M (1995) The impact of pacific ocean subsurface data on operational prediction of tropical Pacific SST at the NCEP. *Weather Forecast* 10:708–714. [https://doi.org/10.1175/1520-0434\(1995\)010<0708:TIOPOS>2.0.CO;2](https://doi.org/10.1175/1520-0434(1995)010<0708:TIOPOS>2.0.CO;2)
- Soares MB, Dessai S (2015) Exploring the use of seasonal climate forecasts in Europe through expert elicitation. *Clim Risk Manag* 10:8–16. <https://doi.org/10.1016/j.crm.2015.07.001>
- Solomon A, Goddard L, Kumar A, Carton J, Deser C, Fukumori I, Greene AM, Hegerl G, Kirtman B, Kushnir Y, Newman M, Smith D, Vimont D, Delworth T, Meehl GA, Stockdale T (2011) Distinguishing the roles of natural and anthropogenically forced decadal climate variability. *Bull Am Meteorol Soc* 92:141–156. <https://doi.org/10.1175/2010BAMS2962.1>
- Sperber KR, Annamalai H, Kang I-S, Kitoh A, Moise A, Turner A, Wang B, Zhou T (2013) The Asian summer monsoon: an inter-comparison of CMIP5 vs. CMIP3 simulations of the late 20th century. *Clim Dyn* 41:2711–2744. <https://doi.org/10.1007/s00382-012-1607-6>
- Taylor KE, Stouffer RJ, Meehl GA (2012) An overview of CMIP5 and the experiment design. *Bull Am Meteorol Soc* 93:485–498. <https://doi.org/10.1175/BAMS-D-11-00094.1>
- Toniazzo T, Koseki S (2018) A methodology for anomaly coupling in climate simulation. *J Adv Model Earth Syst* 10:2061–2079. <https://doi.org/10.1029/2018MS001288>

- Torralba V, Doblas-Reyes FJ, MacLeod D, Christel I, Davis M (2017) Seasonal climate prediction: a new source of information for the management of wind energy resources. *J Appl Meteorol Climatol* 56:1231–1247. <https://doi.org/10.1175/JAMC-D-16-0204.1>
- van den Dool H (2006) Empirical methods in short-term climate prediction. Oxford University Press, Oxford
- van Vuuren DP, Edmonds J, Kainuma M, Riahi K, Thomson A, Hibbard K, Hurtt GC, Kram T, Krey V, Lamarque J-F, Masui T, Meinshausen M, Nakicenovic N, Smith SJ, Rose SK (2011) The representative concentration pathways: an overview. *Clim Change* 109:5. <https://doi.org/10.1007/s10584-011-0148-z>
- Vecchi GA, Delworth T, Gudgel R, Kapnick S, Rosati A, Wittenberg AT, Zeng F, Anderson W, Balaji V, Dixon K, Jia L, Kim H-S, Krishnamurthy L, Msadek R, Stern WF, Underwood SD, Villarini G, Yang X, Zhang S (2014) On the seasonal forecasting of regional tropical cyclone activity. *J Clim* 27:7994–8016. <https://doi.org/10.1175/JCLI-D-14-00158.1>
- Vernieres G, Rienecker MM, Kovach R, Keppenne CL (2012) The GEOS-iODAS: description and evaluation. NASA technical report, NASA/TM-2012-104606. <http://gmao.gsfc.nasa.gov/pubs/docs/Vernieres589.pdf>. Accessed 18 July 2019
- Vertenstein M, Craig T, Middleton A, Feddema D, Fischer C (2012) CESM1.0.3 user guide. <http://www.cesm.ucar.edu/models/cesm1.0/cesm/cesmdoc104/ug.pdf>. Accessed 23 Jan 2015
- Wallace JM, Gutzler DS (1981) Teleconnections in the geopotential height field during the northern hemisphere winter. *Mon Weather Rev* 109:784–812. [https://doi.org/10.1175/1520-0493\(1981\)109<0784:TITGHF>2.0.CO;2](https://doi.org/10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2)
- Wang Y-M, Lean JL, Sheeley NR Jr (2005) Modeling the Sun's magnetic field and irradiance since 1713. *Astrophys J* 625:522–538. <https://doi.org/10.1086/429689>
- Wang Y, Counillon F, Bethke I, Keenlyside N, Bocquet M, Shen M-I (2017) Optimising assimilation of hydrographic profiles into isopycnal ocean models with ensemble data assimilation. *Ocean Modell* 114:33–44. <https://doi.org/10.1016/j.ocemod.2017.04.007>
- Weisheimer A, Doblas-Reyes FJ, Palmer TN, Alessandri A, Arribas A, Déqué M, Keenlyside N, MacVean M, Navarra A, Rogel P (2009) ENSEMBLES: a new multi-model ensemble for seasonal-to-annual predictions-skill and progress beyond DEMETER in forecasting tropical Pacific SSTs. *Geophys Res Lett*. <https://doi.org/10.1029/2009GL040896>
- Xie S-P (1994) On the genesis of the equatorial annual cycle. *J Clim* 7:2008–2013. [https://doi.org/10.1175/1520-0442\(1994\)007<2008:OTGOTE>2.0.CO;2](https://doi.org/10.1175/1520-0442(1994)007<2008:OTGOTE>2.0.CO;2)
- Zelle H, Appeldoorn G, Burgers G, van Oldenborgh GJ (2004) The relationship between sea surface temperature and thermocline depth in the eastern equatorial Pacific. *J Phys Oceanogr* 34:643–655. <https://doi.org/10.1175/2523.1>
- Zhang S, Harrison MJ, Rosati A, Wittenberg A (2007) System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies. *Mon Weather Rev* 135:3541–3564. <https://doi.org/10.1175/MWR3466.1>
- Zheng F, Zhu J, Zhang R-H, Zhou G-Q (2006) Ensemble hindcasts of SST anomalies in the tropical Pacific using an intermediate coupled model. *Geophys Res Lett*. <https://doi.org/10.1029/2006GL026994>
- Zhu J, Kumar A, Huang B (2015a) The relationship between thermocline depth and SST anomalies in the eastern equatorial Pacific: seasonality and decadal variations. *Geophys Res Lett* 42:4507–4515. <https://doi.org/10.1002/2015GL064220>
- Zhu J, Kumar A, Wang H, Huang B (2015b) Sea surface temperature predictions in NCEP CFSv2 using a simple ocean initialization scheme. *Mon Weather Rev* 143:3176–3191. <https://doi.org/10.1175/MWR-D-14-00297.1>
- Zhu J, Kumar A, Lee H-C, Wang H (2017) Seasonal predictions using a simple ocean initialization scheme. *Clim Dyn* 49:3989–4007. <https://doi.org/10.1007/s00382-017-3556-6>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.