The Canadian Seasonal to Interannual Prediction System Version 2 (CanSIPsv2)

HAI LIN,^a WILLIAM J. MERRYFIELD,^b RYAN MUNCASTER,^a GREGORY C. SMITH,^a
MARKO MARKOVIC,^c FRÉDÉRIC DUPONT,^c FRANÇOIS ROY,^a JEAN-FRANÇOIS LEMIEUX,^a
ARLAN DIRKSON,^d VIATCHESLAV V. KHARIN,^b WOO-SUNG LEE,^b MARTIN CHARRON,^a AND
AMIN ERFANI^c

^a *Recherche en Prévision Numérique, Environment and Climate Change Canada, Dorval, Québec, Canada*

^b *Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada, Victoria, British Columbia, Canada*

^c *Canadian Centre for Meteorological and Environmental Prediction, Environment and Climate Change Canada, Dorval, Québec, Canada*

^d *Université du Québec à Montréal, Montréal, Québec, Canada*

(Manuscript received 30 December 2019, in final form 4 May 2020)

ABSTRACT

The second version of the Canadian Seasonal to Interannual Prediction System (CanSIPsv2) was implemented operationally at Environment and Climate Change Canada (ECCC) in July 2019. Like its predecessors, CanSIPsv2 applies a multimodel ensemble approach with two coupled atmosphere–ocean models, CanCM4i and GEM-NEMO. While CanCM4i is a climate model, which is upgraded from CanCM4 of the previous CanSIPsv1 with improved sea ice initialization, GEM-NEMO is a newly developed numerical weather prediction (NWP)-based global atmosphere–ocean coupled model. In this paper, CanSIPsv2 is introduced, and its performance is assessed based on the reforecast of 30 years from 1981 to 2010, with 10 ensemble members of 12-month integrations for each model. Ensemble seasonal forecast skill of 2-m air temperature, 500-hPa geopotential height, precipitation rate, sea surface temperature, and sea ice concentration is assessed. Verification is also performed for the Niño-3.4, the Pacific–North American pattern (PNA), the North Atlantic Oscillation (NAO), and the Madden–Julian oscillation (MJO) indices. It is found that CanSIPsv2 outperforms the previous CanSIPsv1 system in many aspects. Atmospheric teleconnections associated with the El Niño–Southern Oscillation (ENSO) are reasonably well captured by the two CanSIPsv2 models, and a large part of the seasonal forecast skill in boreal winter can be attributed to the ENSO impact. The two models are also able to simulate the Northern Hemisphere teleconnection associated with the tropical MJO, which likely provides another source of skill on the subseasonal to seasonal time scale.

1. Introduction

The Canadian Centre for Meteorological and Environmental Prediction (CCMEP) of Environment and Climate Change Canada (ECCC) was among the first to use a multimodel ensemble (MME) approach in operational seasonal predictions. The first generation of the ECCC dynamical seasonal prediction system was implemented in September 1995, based on the Historical Forecasting Project (HFP; Derome et al. 2001). Ensemble seasonal forecasts were produced using two atmospheric

models, the Atmospheric General Circulation Model version 2 (AGCM2) of the Canadian Centre for Climate Modelling and Analysis (CCCma), and a reduced-resolution version of the global spectral model (SEF) developed at Recherche en Prévision Numérique (RPN; Ritchie 1991). The former is a climate model while the latter is a numerical weather prediction (NWP) model. Both are atmospheric-only models and uncoupled with the ocean, thus the seasonal forecast system was a two-tier system. Each model was integrated for three months with six ensemble members with specified persistent sea surface temperature (SST) and sea ice anomalies. Forecasts for the standard seasons DJF, MAM, JJA, and SON were issued four times per year on 1 December, 1 March, 1 June, and 1 September. SEF was replaced

Denotes content that is immediately available upon publication as open access.

Corresponding author: Hai Lin, hai.lin@canada.ca

DOI: 10.1175/WAF-D-19-0259.1

© 2020 American Meteorological Society. For information regarding reuse of this content and general copyright information, consult the [AMS Copyright Policy](https://www.ametsoc.org/PUBSReuseLicenses) (www.ametsoc.org/PUBSReuseLicenses).

by the Global Environmental Multiscale (GEM) model (Côté et al. 1998) in March 2004, which was developed at RPN and has served in its successive versions as an operational NWP model at CCMEP.

The CCMEP's dynamical seasonal forecast system was upgraded in October 2007 to a four-model system using AGCM2, SEF, GEM, and CCCma's Atmospheric General Circulation Model, version 3 (AGCM3) (Scinocca et al. 2008), with 40 ensemble members, 10 from each model. Seasonal forecasts were issued both at 0-month lead (months 1–3) and 1-month lead (months 2–4) on the first day of every month, and a first-month forecast for surface temperature was issued on the 1st and 15th of every month. Hindcasts for bias correction and skill assessment were provided by the second Historical Forecasting Project that applied the same four models (HFP2; Kharin et al. 2009).

A major upgrade of CCMEP's seasonal prediction system took place in December 2011 when the two-tier system was replaced by the Canadian Seasonal to Interannual Prediction System (CanSIPS version 1, hereafter CanSIPsv1), a global coupled one-tier system. CanSIPsv1 used two atmosphere–ocean coupled climate models, CCCma Coupled Climate Model, versions 3 and 4 (CanCM3 and CanCM4), with a total of 20 ensemble members, 10 from each model. From the beginning of each month, 10 members of ensemble forecasts for each model were produced for 12 months from initial conditions containing small perturbations at the same time. Official products included seasonal temperature and precipitation forecasts of lead time from 0 to 9 months. Details on the CanSIPsv1 system are described in Merryfield et al. (2013). Output in the first month was used to produce the monthly temperature anomaly forecast over Canada until July 2015 when the monthly forecast was replaced by that from the Global Ensemble Prediction System (GEPS) based monthly forecasting system (Lin et al. 2016).

The second version of CanSIPS (CanSIPsv2) is also based on a multimodel approach. CanCM4 is upgraded to CanCM4i with an improved sea ice initialization procedure. The major change from CanSIPsv1 to CanSIPsv2 is that CanCM3 is replaced by a very different numerical platform based on the NWP GEM model (Côté et al. 1998; Girard et al. 2014), the Nucleus for European Modelling of the Ocean (NEMO; <http://www.nemo-ocean.eu>) ocean model and the Community Ice Code (CICE; Hunke and Lipscomb 2010) sea ice model. The latter coupled model is referred to as GEM-NEMO in this manuscript. Development of the GEM-NEMO seasonal forecast model took place mainly at RPN, exploiting the coupled model and modeling expertise available there (e.g., Smith et al. 2018), which also benefits from the contribution of the Canadian Operational

Network of Coupled Environmental Prediction Systems (CONCEPTS; Smith et al. 2013) project.

Besides CCMEP, dynamical seasonal predictions are routinely produced in many other operational centers using atmosphere–ocean coupled model prediction systems. Examples include the Climate Forecast System, version 2, of the National Centers for Environmental Prediction (NCEP CFSv2; Saha et al. 2014), the Met Office's Global Seasonal Forecasting System, version 5 (GloSea5; MacLachlan et al. 2015), and the fifth-generation seasonal forecast system of the European Centre for Medium-Range Weather Forecasts (ECMWF SEAS5; Johnson et al. 2019). CanSIPS has been an integral part of several international activities related to multimodel ensemble seasonal forecasts, including the North American Multimodel Ensemble (NMME) (e.g., Kirtman et al. 2014), the World Meteorological Organization (WMO) Long-Range Forecast Multi-Model Ensemble (e.g., Graham et al. 2011), and the multimodel climate prediction of the Asian-Pacific Economic Cooperation Climate Center (APCC) (e.g., Min et al. 2014).

This paper introduces CanSIPsv2, which was implemented operationally in July 2019 to replace CanSIPsv1. Section 2 describes the two CanSIPsv2 models. Section 3 describes the forecast initialization. The reforecast is introduced in section 4. The overall performance of CanSIPsv2 is discussed in section 5, which includes systematic errors and skill evaluation. In section 6, we analyze atmospheric teleconnections related to ENSO and the Madden–Julian oscillation (MJO), which provide an important source of skill for subseasonal to seasonal predictions. A summary and discussion are given in section 7.

2. CanSIPv2 models

Like its predecessors (the HFP, HFP2, and CanSIPsv1), CanSIPsv2 is a multimodel system which takes advantage of the generally greater skill of multimodel ensembles compared to single-model systems for a given ensemble size (e.g., Krishnamurti et al. 1999; Kharin et al. 2009; Becker et al. 2014).

CanSIPsv2 consists of two global coupled models, CanCM4i and GEM-NEMO. CanCM4i applies an improved initialization to the same version of the CanCM4 model used previously in CanSIPsv1 (see section 3). This model is described in detail in Merryfield et al. (2013), and therefore only its most basic aspects are outlined here. On the other hand, GEM-NEMO is a new model introduced to CanSIPsv2 and will be described in detail.

a. CanCM4i

CanCM4i, like CanCM4, couples CCCma's fourth-generation atmospheric model CanAM4 to CCCma's

CanOM4 ocean component. The horizontal resolution of the CanAM4 atmospheric model is T63, corresponding to a 128×64 Gaussian grid (about 2.8°), with 35 vertical levels and a top at 1 hPa. The CanOM4 oceanic model uses a 256×192 grid which provides a horizontal resolution of approximately 1.4° in longitude and 0.94° in latitude, with 40 vertical levels. CanCM4 employs version 2.7 of the Canadian Land Surface Scheme (CLASS; Verseghy 2000), and a single-thickness category, cavitating-fluid sea ice model (Flato and Hibler 1992), both of which are formulated on the atmospheric model grid. CanAM4 and CanOM4 are coupled once per day, and as for CanCM4 no flux corrections are employed.

b. GEM-NEMO

GEM-NEMO, developed at RPN, is a fully coupled model with the atmospheric component GEM, the ocean component NEMO, and the CICE sea ice model.

The GEM version used is 4.8-LTS.16, with a horizontal resolution of 256×128 grid points (about 1.4°) evenly spaced on a latitude–longitude grid, and 79 vertical levels with the top at 0.075 hPa. The land scheme is ISBA (Noilhan and Planton 1989; Noilhan and Mahfouf 1996), where each grid point is assumed to be independent (no horizontal exchange). The atmospheric deep convection parameterization scheme of Kain–Fritsch (Kain and Fritsch 1990) is used. The Kuo-transient scheme (Bélair et al. 2005) is applied for shallow convection. The model represents climatological vegetation and ozone seasonal evolution during the integration. The evolution of anthropogenic radiative forcing is treated simply through specification of equivalent CO_2 concentrations with a linear trend. The GEM model is integrated with a time step of 1 h. Among the 10 ensemble members in both the forecast and the reforecast, an implicit surface flux numerical formulation is used in members 1–5 and an explicit surface flux formulation is applied in members 6–10.

The ocean model uses NEMOv3.1 on the ORCA 1 grid with a nominal horizontal resolution of $1^\circ \times 1^\circ$ ($1/3^\circ$ meridionally near the equator) and 50 vertical levels. The CICE 4.0 model is used for the sea ice component with five ice-thickness categories. The NEMO model is run with a 30-min time step.

GEM and NEMO are coupled once an hour through the Globally Organized System for Simulation Information Passing (GOSSIP) coupler developed at CCMEP. Further details on the GOSSIP coupler and coupling methodology can be found in Smith et al. (2018). No flux corrections are employed in the coupled model.

3. Forecast initialization

Similar to CanSIPsv1, each of the two models in CanSIPsv2, CanCM4i and GEM-NEMO, makes its

real-time forecast once a month, initialized at the beginning of each month. Each model produces forecasts of 10 ensemble members with all the integrations starting from the same date and lasting for 12 months. However, the two models take different approaches in the initialization of the forecasts.

a. CanCM4i

The only change in CanCM4i forecast initialization compared to CanCM4 initialization in CanSIPsv1 is that initial conditions for Northern Hemisphere (Arctic) sea ice thickness are obtained from the SMv3 statistical model of Dirkson et al. (2017). SMv3 extrapolates linear trends in ice thickness obtained from the Pan-Arctic Ice and Ocean Modeling and Assimilation System (PIOMAS; Zhang and Rothrock 2003), and includes interannual thickness variations that are proportional to linearly detrended ice concentration anomalies. Additional changes in reforecast initialization that are intended to improve consistency between reforecasts and real-time forecasts are described in section 4. Other aspects, common to CanCM4 in CanSIPsv1, are that each CanCM4i ensemble member is initialized from a separate assimilating model run in which atmospheric temperature, moisture, horizontal winds, SST and sea ice concentration are constrained by values from the Global Deterministic Prediction System (GDPS; Buehner et al. 2015) analysis. Additionally, land surface variables are initialized through the response of the land surface component to atmospheric conditions in the assimilating model runs, whereas subsurface ocean temperatures are constrained by values from the Global Ice Ocean Prediction System (GIOPS; Smith et al. 2016) analysis through an offline procedure. Details of the atmosphere and ocean initialization methods are provided in Merryfield et al. (2013).

b. GEM-NEMO

In the forecast, the atmospheric initial conditions of GEM-NEMO come from those of GEPS (Gagnon et al. 2015; Lin et al. 2016), which are generated from the ensemble Kalman filter (EnKF; Houtekamer et al. 2009, 2014) with data assimilation. Ten of the GEPS perturbed initial conditions are used for the GEM-NEMO seasonal forecast.

For the land surface initial conditions, in addition to taking advantage of the CCMEP analysis, one important consideration is to make it consistent with the reforecast. In the reforecast, which will be introduced in section 4b, we initialize the land surface by running the Surface Prediction System (SPS; Carrera et al. 2010) forced by the near-surface atmospheric and the precipitation fields of the ERA-Interim reanalysis

(Dee et al. 2011). In real-time forecasts, the land surface initial fields are generated by forcing the SPS with the CCMEP analysis.

The ocean and sea ice initial conditions in real-time forecasts come from the CCMEP GIOPS analysis. The near-surface air temperature from GEPS and the snow depth from SPS are used for the initialization of these variables over sea ice in CICE. Aside from the near-surface air temperature over ice, all ensemble members have the same ocean and sea ice initial conditions. For the reforecasts, the ocean and sea ice initial conditions are described in section 4b.

4. Reforecasts

For seasonal predictions, model drift and systematic model errors become a serious problem that contaminate the forecast quality. A common practice to mitigate such errors is to perform a historical reforecast (or hindcast) to estimate the model climatology and statistics, so that calibrated forecasts can be made. Seasonal predictions are usually presented as anomalies, which are departures from the model climatological seasonal averages estimated from the hindcast. Another purpose of the reforecast is to generate a long record of forecast data to quantify the performance of the forecast system. Therefore, the reforecast is an important component of the seasonal forecasting system. As in CanSIPsv1, the model climatology is estimated from hindcasts for each individual model, covering 30 years from 1981 to 2010. Like the real-time forecasts, 10 ensemble members are integrated from the beginning of each month for 12 months. This time period is consistent with the current WMO standards (WMO 2017). In the reforecasts, each individual model has the same configuration as in the real-time forecasts, as described in section 3.

a. CanCM4i

CanCM4i employs several changes to the CanCM4 reforecast initialization procedure described in Merryfield et al. (2013). First, as in the real time forecasts, modeled sea ice thickness is constrained to values derived from the SMv3 statistical model of Dirkson et al. (2017), which is much more realistic than the monthly varying (but otherwise stationary) model-based climatology employed for CanCM4, in part because it accounts for the long-term thinning of Arctic sea ice. Second, modeled sea ice concentration is constrained to values derived from Had2CIS, which is a product created by merging data from HadISST2.2, which employs an ice chart-based bias correction of the passive microwave record (Titchner and Rayner 2014), with data from digitized sea ice charts of the Canadian Ice Service

(CIS; Tivy et al. 2011). The monthly HadISST2.2 data, the digitized CIS weekly sea ice charts over the Arctic region, and the weekly CIS "Great Lakes ice charts" over Great Lakes are interpolated to daily data before being combined. Had2CIS provides improved temporal consistency across the reforecast period, as well as improved consistency with sea ice concentrations from the GDPS analysis used in real time, compared to the concentration product used to initialize CanCM4 reforecasts. Finally, initial subsurface ocean temperatures are constrained by the ORAP5 ocean reanalysis (Zuo et al. 2017), which is both closer in its formulation to the GIOPS analysis used in real time, and has more realistic trends than the ocean reanalysis used to initialize CanCM4 reforecasts.

b. GEM-NEMO

In the reforecasts, the atmospheric conditions for GEM-NEMO are initialized using the ERA-Interim reanalysis. Random isotropic perturbations are added to the reanalysis fields to create 10 different initial conditions. The atmospheric perturbations are homogeneous and isotropic as in Gauthier et al. (1999). Only the streamfunction and the unbalanced temperatures are perturbed as in EnKF (Houtekamer et al. 2009). These perturbed fields are then transformed to wind, temperature and surface pressure. The same approach of perturbation generation is used in the hindcast of the ECCC monthly prediction system (Lin et al. 2016). The ocean initial fields come from the ORAP5 ocean reanalysis (Zuo et al. 2017). The monthly values of temperature, salinity, zonal and meridional currents are interpolated to daily values. The sea ice concentration initial fields are from Had2CIS, the same as for CanCM4i. The sea ice thickness is interpolated from the monthly ORAP5 data. The near-surface air temperature from ERA-Interim and the snow depth from SPS are used for the initialization of these variables over sea ice in CICE.

For the land surface initialization, we make use of a historical dataset of surface fields produced by SPS (Carrera et al. 2010), which was previously known as GEM-surf and was used in several studies at high resolution (Separovic et al. 2014; Ioannidou et al. 2014; Bernier and Bélair 2012). SPS is simply the surface schemes (land, sea ice surface and glacier) of the GEM model used in offline mode to create surface fields compatible with the surface scheme of GEM. To initialize the land surface fields in the reforecasts, we have forced SPS for the period 1980–2010 by the near-surface atmospheric and the precipitation fields coming from the ERA-Interim reanalysis. The surface pressure, the 2-m temperature and dewpoint depression as well as the

solar and longwave downward radiative fluxes at the surface are provided to SPS at a 3-h interval. To limit the snow accumulation over glaciers, the maximum snow depth is set to 12 m in the initial condition.

It should be noted that the atmosphere, ocean and land are initialized separately which may cause some inconsistency and imbalance thus produce some initial shocks to the system. Coupled initialization is an important area of future development of the GEM-NEMO seasonal prediction system.

5. Overall performance of CanSIPSv2

Here we evaluate the general performance of CanSIPSv2. The verification data used are the ERA-Interim reanalysis for air temperatures and geopotential height, and the Global Precipitation Climatology Project (GPCP), version 2.3, dataset (Adler et al. 2018) for precipitation. The NOAA Optimum Interpolation Sea Surface Temperature, version 2 (OISST) (Reynolds et al. 2002), data are used for SST. For the sea ice concentration verification, the Had2CIS dataset is applied. In the following, systematic errors of the CanSIPSv2 models in SST, 2-m air temperature (T2m), and precipitation rate (PR) are first discussed. Then, forecast skill is assessed in comparison with CanSIPSv1.

a. Systematic errors

In this subsection, systematic errors consisting of the difference between the predicted and observed climatology over the reforecast period of 1981–2010 are analyzed for the CanCM4i and GEM-NEMO models. The objective is to demonstrate the general behavior of the two models related to model drift. When doing seasonal prediction, forecasts of anomalies are produced, where the model climate estimated from reforecasts is subtracted to remove the systematic error from the real-time forecast. Thus there is no direct impact of the systematic error on the forecast skill, although the systematic error can affect the forecast skill indirectly through the model dynamics. For example, a biased midlatitude westerly jet would lead to errors in Rossby wave propagation and teleconnection patterns forced by ENSO, resulting in an inaccurate anomaly forecast. Therefore, it is important to consider model bias in order to establish how such influences might be affected.

The systematic error presented here is for the 1-month lead seasonal mean (e.g., seasonal mean of DJF from 1 November initial conditions) averaged over the 12 initialization months. We first look at the model bias of SST, which is shown in Figs. 1a and 1b. CanCM4i has limited model drift, with warm biases only in localized areas near the west coast of South America, east coast of

East Asia and North America, and cold biases along the equatorial Pacific and particularly high latitude North Atlantic and Southern Oceans. On the other hand, GEM-NEMO has systematically larger cold biases in the tropical oceans, which could be an area of future improvement for GEM-NEMO, while biases in the high-latitude oceans are generally smaller than for CanCM4i.

Figures 1c and 1d show the geographical distributions of systematic error for T2m. The CanCM4i model is seen to have quite strong cold biases over the Northern Hemisphere Polar Region (Fig. 1c) that may be related to the treatment of sea ice physics or atmospheric physics over sea ice. Over the Antarctic, the two models have cold biases. It should be noted that the ERA-Interim reanalysis may not be best for validation over ice covered surfaces. Both models have warm T2m biases over the northern mid- to high-latitude continents, and cold biases over mountainous regions. CanCM4i has strong warm biases over the Amazon basin. Over the ocean, CanCM4i shows some warm biases near the west coast of South America, east coast of East Asia and North America, consistent with the SST bias. Cold T2m biases occur in GEM-NEMO in the tropical Pacific, which are related to the SST biases as discussed above.

Shown in Figs. 1e and 1f are the systematic errors in precipitation rate. The two models seem to have a similar precipitation bias distribution. Excessive precipitation is found over a large area of the tropics with insufficient precipitation over the equatorial Pacific, a feature of the double-intertropical convergence zone (ITCZ) which is common to many coupled general circulation models (e.g., Lin 2007). The magnitude of the wet bias in CanCM4i is stronger than GEM-NEMO. CanCM4i also has a strong wet bias in the western tropical Indian ocean, which is absent in GEM-NEMO. Over the tropical South American continent, CanCM4i has a stronger dry bias than GEM-NEMO.

Systematic errors of a climate model occur not only in seasonal means, but also in transient eddy activity. The synoptic-scale transient eddies, which have a time scale of about 2–8 days, arise primarily from baroclinic instability of the midlatitude westerly jet. They are often associated with cyclones, anticyclones and storm activity, which result in day-to-day changes in weather and thus in temperature and precipitation. The relevance of these high-frequency eddies to seasonal prediction is associated with their accumulated effect over time. On the other hand, many previous studies demonstrated that the midlatitude synoptic-scale transients, which are displaced or influenced by changes in seasonal mean atmospheric flow, feed back to reinforce the seasonal

FIG. 1. Systematic error at 1-month lead averaged over all 12 initialization months (a) CanCM4i SST, (b) GEM-NEMO SST, (c) CanCM4i T2m, (d) GEM-NEMO T2m, (e) CanCM4i PR, and (f) GEM-NEMO PR. The verification data are OISST V2 for SST, ERA-Interim for T2m, and GPCP V2.3 for PR.

mean flow anomaly through transient eddy vorticity flux convergence (e.g., Lau 1988; Klasa et al. 1992). Therefore, it is important for a seasonal forecast system to represent reasonably well the synoptic-scale transient eddies and storm track activities. Here we use the standard deviation of 2–8-day bandpass-filtered 500-hPa geopotential height (Z_{500}) during the DJF season to represent the synoptic-scale transient eddy activity. Shown in Fig. 2 are the transient eddy activities for CanCM4i and GEM-NEMO in DJF at 1-month lead and their biases with respect to the ERA-Interim reanalysis. As can be seen, both models capture the general features of the North Pacific and North Atlantic

storm tracks. CanCM4i tends to overestimate the storm track activity strength near the end of the two storm tracks, while the transient eddy activity in GEM-NEMO is weaker than in the analysis in the eastern Atlantic and the eastern Pacific.

b. Forecast skill

Here the forecast skill of CanSIPsv2 is evaluated in comparison with CanSIPsv1 for the hindcast period 1981–2010. The skill measures are first considered for deterministic forecasts, determined from the ensemble mean of the predicted anomalies. In addition to global maps for winter and summer forecasts, we consider area

FIG. 2. Storm track activity in DJF at 1-month lead for (a) CanCM4i and (b) GEM-NEMO. Contours represent the standard deviation of 2–8-day bandpass-filtered Z500. Shading represents the difference with ERA-Interim. Units are meters.

averages (over the globe, global land, North America and Canada for temperature and precipitation; over the globe, the Northern Hemisphere extratropics, tropics and Southern Hemisphere extratropics for 500-hPa geopotential height) of anomaly correlation (i.e., the correlation coefficient between predicted and observed anomalies). Ensemble forecast probabilistic skill over North America is also considered, using the continuous ranked probability skill score (CRPSS; e.g., [Bradley and Schwartz 2011](#)) which measures the fractional improvement in error between the forecast distribution and observed value, relative to a forecast based on the observed climatology. The CRPSS is presented for the calibrated probability forecasts following the methodology in [Kharin et al. \(2017\)](#). Variables considered include T2m, precipitation rate, Z500, and sea ice concentration. In addition, we discuss the forecast skill of SST and ENSO, as well as several climate indices that are important to seasonal predictions including those for the Pacific–North America (PNA) pattern (e.g., [Wallace and Gutzler 1981](#)), the North Atlantic Oscillation (NAO) (e.g., [Hurrell et al. 2003](#)), and the MJO (e.g., [Madden and Julian 1971](#)).

1) SURFACE AIR TEMPERATURE (T2M)

[Figure 3](#) shows the global geographical distribution of anomaly correlation skill for seasonal mean surface air temperature from the CanSIPsv1 and the CanSIPsv2 hindcasts for the DJF and JJA seasons at 1-month lead.

For the DJF forecast ([Figs. 3a,b](#)), the highest skill for both systems is in the tropics and especially the tropical Pacific, where SST anomalies are most persistent and ENSO imparts relatively high predictability. Skill is also appreciable in many locations over land, including parts of North America where much of the seasonal predictability particularly in winter and early spring is attributable to the teleconnected influence of ENSO, as will be discussed in more detail in [section 6](#). In general, the skill distributions of the two systems are comparable although the global average skill of CanSIPsv2 is higher than CanSIPsv1. Similar conclusions apply for the JJA season ([Figs. 3c,d](#)), as well as other seasons (not shown). The new CanSIPsv2 system is also showing evidently higher skill in the NAO sector in both DJF and JJA seasons.

Shown in [Fig. 4a](#) is the T2m anomaly correlation skill of seasonal hindcasts at 1-month lead for different regions averaged over all 12 initialization months. It is evident that GEM-NEMO alone has a performance comparable to or better than that of CanSIPsv1. By combining GEM-NEMO with CanCM4i, CanSIPsv2 skill is improved when compared to CanSIPsv1 over all the regions.

To measure the quality of probabilistic forecasts of the ensemble system, we present in [Fig. 5](#) the CRPSS over North America for DJF and JJA seasonal mean T2m forecasts at 1-month lead. In the DJF forecast, appreciable skill scores are observed in a large part of

FIG. 3. Correlation skill of seasonal mean T2m at 1-month lead: (a) CanSIPsv1 DJF, (b) CanSIPsv2 DJF, (c) CanSIPsv1 JJA, and (d) CanSIPsv2 JJA. The number on the top of each panel is the global average value.

northeast Canada, as well as west coastal regions and much of Mexico in both CanSIPsv1 (Fig. 5a) and CanSIPsv2 (Fig. 5b). Relatively low skill is seen in the central and eastern United States. The CRPSS skill distribution is consistent with the anomaly correlation of the ensemble mean forecast as shown in Fig. 3. As discussed in Kumar (2009), there is a strong relationship between the CRPSS and correlation skill. Comparing the two forecast systems, it is seen that CanSIPsv2 performs better overall than CanSIPsv1, especially in eastern Canada and Mexico. In JJA, the CRPSS skill scores seem comparable between CanSIPsv1 and CanSIPsv2 (Figs. 5c,d). Relatively high forecast skill is observed in the southwestern United States, especially in CanSIPsv2.

To establish a statistical significance of skill difference between two seasonal forecast systems, we follow the random walks method described in DelSole and Tippett (2016). To measure the skill of a given forecast for T2m, the percent correct score in terms of three categories (i.e., above normal, near normal, and below normal) is calculated as the percentage of area that is correctly predicted. The criterion used to define the categories is 0.43 times the standard deviation, so that the three categories have approximately an equal probability to occur. The ensemble mean forecast is used and the verification is against the ERA-Interim reanalysis.

Shown in Fig. 6a are the results for seasonal mean T2m forecast at 1-month lead. For the global and global land regions, CanSIPsv2 shows significantly better skill than CanSIPsv1 even with a small number of forecast cases. Over Canada, the significance can be established with about 15 years of hindcast data.

2) PRECIPITATION

The global distributions of anomaly correlation skill for seasonal mean precipitation forecast in DJF at 1-month lead are shown in Figs. 7a and 7b for CanSIPsv1 and CanSIPsv2 hindcasts, respectively. Again, high skill is mainly observed in the tropical regions, reflecting the contribution of ENSO. Some skill of precipitation is seen in western and eastern coastal regions of North America. The global average skill of CanSIPsv2 is higher than CanSIPsv1, although the distributions of the two systems are very similar. For the JJA forecast (Figs. 7c,d), the precipitation skill has a similar distribution as DJF but with a weaker magnitude. In this season, both systems produce skillful seasonal precipitation forecasts over parts of western North America. Again, the global average skill of precipitation in CanSIPsv2 is higher than that of CanSIPsv1.

Probabilistic skill scores for DJF and JJA seasonal mean precipitation forecasts at 1-month lead are shown in Fig. 8. Although the precipitation forecast skill is

FIG. 4. Anomaly correlation skill of seasonal means of (a) T2m and (b) Z500 at 1-month lead. The skills shown are averages over all 12 initialization months. The CanSIPsv2 hindcast skill is shown separately for the individual component models CanCM4i and GEM-NEMO as well as for the two-model CanSIPsv2 hindcasts. The scores are compared with CanSIPsv1. (c) Globally averaged anomaly correlation skill for seasonal mean Z500 as a function of lead time.

mostly low in the mid- and high-latitude North America, appreciable improvement in CanSIPsv2 can be identified for DJF particularly in the southwestern United States and adjoining parts of Mexico (Figs. 8a,b). For JJA (Figs. 8c,d), the two systems have relatively high skill in the northwestern United States, with peak CanSIPsv2 skills that are higher.

3) 500-hPa GEOPOTENTIAL HEIGHT

The anomaly correlation skills for DJF seasonal mean Z500 at 1-month lead are compared between CanSIPsv1 and CanSIPsv2 in Figs. 9a and 9b. Similar to T2m and precipitation, significant skill of Z500 is found in the tropics. In the Northern Hemisphere extratropics, the

correlation skill is high over the North Pacific, most parts of Canada and the region of southern United States and Mexico, resulting from variability of the PNA pattern associated with ENSO (e.g., Wallace and Gutzler 1981). Relatively high correlation skills are also observed over Greenland and the midlatitude North Atlantic, which appears to be related to the NAO (e.g., Hurrell et al. 2003). In general, the two forecast systems have a similar skill distribution.

For the season of JJA, as can be seen in Figs. 9c and 9d, the forecast skill of seasonal mean Z500 is also high in the tropics. In the Northern Hemisphere extratropics, besides over Greenland and the polar regions, four centers of high skill are found in a wave pattern along the midlatitudes, that are eastern Europe, China, the North Pacific, and western North America. The last is likely responsible for the relatively high temperature and precipitation skills in western North America as seen in Figs. 5c and 5d and 8c and 8d. It is possible that the skill in the subtropical Northern Hemisphere is associated with the circumglobal teleconnection (CGT; e.g., Branstator 2002), which is characterized by a wavenumber-5 structure in the upper troposphere along the subtropical westerly jet. The CGT variability was found to be linked to tropical convections (e.g., Ding et al. 2011; Lin et al. 2015). The Northern Hemisphere jet is located around a higher latitude in JJA than in DJF, which may explain why JJA skill (Figs. 9c,d) has a broader meridional extent than in DJF (Figs. 9a,b). Again, CanSIPsv2 has a higher global average of skill value than CanSIPsv1, although the two systems have a very similar skill distribution.

The anomaly correlation skill for Z500 averaged over all 12 initialization months for different regions [i.e., the globe, the Northern Hemisphere extratropics (30°–90°N), the tropics (30°S–30°N), and the Southern Hemisphere extratropics (30°–90°S)] are presented in Fig. 4b for the 1-month lead time. Again, the CanSIPsv2 system outperforms CanSIPsv1 in all the regions. To see the skill evolution with the increase of lead time, shown in Fig. 4c is the Z500 anomaly correlation skill averaged over the globe as a function of lead time. The skills shown are averages for all 12 initialization months. As can be seen, the improvement of forecast skill of CanSIPsv2 over that of CanSIPsv1 occurs for all lead times. The skill improvement appears larger at longer leads than at shorter leads.

To estimate the statistical significance of skill improvement for Z500, pattern correlation between the ensemble mean forecast at 1-month lead and the ERA-Interim anomaly is calculated for each forecast. Shown in Fig. 6b is the result obtained following the random

FIG. 5. CRPSS of seasonal mean T2m at 1-month lead over North America: (a) CanSIPsv1 DJF, (b) CanSIPsv2 DJF, (c) CanSIPsv1 JJA, and (d) CanSIPsv2 JJA.

walks approach (DelSole and Tippett 2016). As can be seen, for the globe and Southern Hemisphere (30° – 90° S), CanSIPsv2 shows significantly better skill than CanSIPsv1 with about six hindcast years, while over the Northern Hemisphere (30° – 90° N), about 10 years of data are needed.

4) SST AND ENSO

We now show the anomaly correlation skill of DJF seasonal mean SST at 1-month lead in Figs. 10a and 10b. The two systems have a very similar skill distribution. High skill of SST forecast can be found

over the global ocean, with the maximum skill in the ENSO region of the equatorial eastern Pacific. Over the North Atlantic, high forecast skill of SST appears in the tropical and high latitude regions with relatively low skill in the midlatitudes, especially the Gulf Stream extension. Similar features occur in the NCEP CFSv2 as reported in Hu et al. (2013). The global average skill of CanSIPsv2 (0.61) is higher than that of CanSIPsv1 (0.55). The skill for the 1-month lead SST forecast of JJA has a similar distribution as in DJF, but is weaker and less concentrated in the ENSO region (Figs. 10c,d).

(a) T2m Percent Correct

(b) Z500 Pattern Correlation

FIG. 6. Comparison of seasonal mean forecast skill at 1-month lead between CanSIPsv2 and CanSIPsv1 over different regions for (a) percent correct of T2m and (b) pattern correlation of Z500. The count increases by 1 when the CanSIPsv2 is less skillful than CanSIPsv1, and decreases by 1 otherwise. The count is accumulated forward in time over all initial months and years of the hindcast. The shaded area indicates the range of 5% significance level. A random walk extending above the shaded area indicates that CanSIPsv1 is more skillful than CanSIPsv2, whereas that extending below the shaded area indicates that CanSIPsv2 is more skillful than CanSIPsv1.

To demonstrate the performance of CanSIPsv2 in predicting ENSO, the correlation skill of the Niño-3.4 index, which is defined as the SST anomaly averaged in the area of 5°N – 5°S , 170° – 120°W , is calculated. Figure 11 provides a summary of the correlation skill of the seasonal mean Niño-3.4 index as a function of target season

(horizontal axis) and lead time in months (vertical axis). The skill scores averaged over all target seasons as a function of lead time are shown along the right-hand side, while those averaged over all lead times as a function of the target season are shown along the top of the figure. The grand average for all the target seasons

(a) CanSIPsv1 DJF 0.22

(b) CanSIPsv2 DJF 0.27

(c) CanSIPsv1 JJA 0.18

(d) CanSIPsv2 JJA 0.21

FIG. 7. Correlation skill of seasonal mean PR at 1-month lead: (a) CanSIPsv1 DJF, (b) CanSIPsv2 DJF, (c) CanSIPsv1 JJA, and (d) CanSIPsv2 JJA.

FIG. 8. CRPSS of seasonal mean PR at 1-month lead over North America: (a) CanSIPsv1 DJF, (b) CanSIPsv2 DJF, (c) CanSIPsv1 JJA, and (d) CanSIPsv2 JJA.

and lead times is indicated at the top-right corner of each panel. In general, the two systems have comparable ENSO forecast skill, although CanSIPsv2 skill is slightly better overall particularly for fall and winter target seasons. For example, for the DJF and JFM target seasons, the skill of 0.90 for CanSIPsv2 reaches 7-month lead, which is about 2 months longer than CanSIPsv1.

CanSIPsv2 performance for predicting ENSO across its 12-month range is illustrated in Fig. 12a, which shows reforecasts of seasonal-mean Niño-3.4 over 1981–2018 at lead times of 0, 3, 6, and 9 months. Strong ENSO

events (seasonal $|\text{Niño-3.4}| > 1.5^{\circ}\text{C}$) are predicted virtually without exception by CanSIPsv2 even at 9-month lead time (forecast months 10–12), although their amplitude tends to be underestimated at longer lead times. CanSIPsv2 is, however, prone to occasional “false alarms” (i.e., predicted strong ENSO events that are not realized in observations as in 1984–85, 1990–91, and 2017–18).

It is well known that individual seasonal forecast models tend to be underdispersive, that is, characterized by spread–error ratios (ensemble standard deviation

FIG. 9. Correlation skill of seasonal mean Z500 at 1-month lead: (a) CanSIPsv1 DJF, (b) CanSIPsv2 DJF, (c) CanSIPsv1 JJA, and (d) CanSIPsv2 JJA.

divided by root-mean square error or RMSE) that are <1 (e.g., [Ho et al. 2013](#)). This results in probabilistic forecasts that are overconfident, whereas for MMEs the spread–error ratio tends to be closer to 1, indicating more appropriate confidence and greater reliability (e.g., [Tompkins et al. 2017](#)). That this applies for CanSIPsv2 is illustrated in [Fig. 12b](#), which shows Niño-3.4 ensemble standard deviations and RMSE for CanCM4i, GEM-NEMO and the CanSIPsv2 MME. For the individual models, RMSE exceeds ensemble spread by up to a factor of 2 or more especially for shorter lead times, whereas the CanSIPsv2 MME has generally lower RMSE and larger ensemble spread than either model alone. Taking $\exp[|\log(\text{spread–error ratio})|]$ as a measure of the absolute distance of the spread–error ratio from 1, where $\langle \cdot \rangle$ denotes averaging over lead times and 1 is the minimum “ideal” value, this metric is 1.71, 1.56, and 1.36 for CanCM4i, GEM-NEMO, and CanSIPsv2, respectively, as compared to 1.71, 1.95, and 1.36 for CanCM4, CanCM3 and CanSIPsv1, respectively.

5) THE PNA AND NAO

The PNA and NAO are two leading modes of interannual variability of seasonal mean circulation in the Northern Hemisphere extratropics (e.g., [Wallace and Gutzler 1981](#)) that greatly influence the weather and

climate. Here we analyze the forecast skill of these two patterns.

The PNA and NAO loading patterns are obtained here as the first and second rotated empirical orthogonal function (REOF) modes of the monthly mean 500-hPa geopotential height over the Northern Hemisphere using the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis ([Kistler et al. 2001](#)), following [Barnston and Livezey \(1987\)](#). Monthly mean anomalies for all 12 calendar months are used. Since the PNA and NAO variability has the largest variance in the cold season, the patterns are thus dominated by characteristics in winter. The forecast PNA and NAO indices are calculated as the projections of the seasonal mean 500-hPa geopotential height anomalies onto the observed REOF patterns. These loading patterns of the PNA and NAO are similar to those in [Fig. 1](#) of [Johansson \(2007\)](#), and the NAO pattern was also shown in [Lin et al. \(2009, their Fig. 1\)](#). REOF calculation of the monthly mean 500-hPa geopotential height at 1-month lead from the hindcast indicates that both GEM-NEMO and CanCM4i have a reasonably good representation of the PNA and NAO spatial distribution (not shown). In the following discussion, PNA and NAO skills are compared between CanSIPsv1 and CanSIPsv2. The verification indices are obtained by projecting the corresponding ERA-Interim

FIG. 10. Correlation skill of seasonal mean SST at 1-month lead: (a) CanSIPsv1 DJF, (b) CanSIPsv2 DJF, (c) CanSIPsv1 JJA, and (d) CanSIPsv2 JJA.

seasonal mean Z500 anomaly onto the same PNA and NAO loading patterns.

The correlation skills of the PNA index as a function of target season are presented in Fig. 13a for the 1-month lead time. The PNA skill changes with target season, with high skill in winter, spring and early summer. In late summer and autumn when the PNA is not well defined, the skill is low. In general, the PNA skill of CanSIPsv2 is similar to that of CanSIPsv1. There is

some degradation of wintertime PNA skill, especially for DJF and JFM. The PNA skill in these two seasons in CanCM4i is lower than CanCM4 of CanSIPsv1 (Table 1). It is possible that the change of sea ice initial condition from CanCM4 to CanCM4i results in slightly different spatial distribution of the internal variability modes, which affects the projection to the observed PNA. This, however, does not affect the general model performance, as CanCM4i has a very similar

FIG. 11. Anomaly correlation reforecast skill of seasonal mean Niño-3.4 index as a function of target season (horizontal axis) and lead time (vertical axis). The verification dataset is OISST V2.

FIG. 12. (a) CanSIPsv2 reforecast values of seasonal-mean Niño-3.4 at 0-, 3-, 6-, and 9-month lead time (colors), compared to OISST observed values (black). Anomaly correlation values at each lead time are indicated in the legend. Dotted lines at $\pm 1.5^\circ\text{C}$ indicate thresholds for strong ENSO events as discussed in the main text. (b) Ensemble standard deviations for CanCM4i, GEM-NEMO, and CanSIPsv2 reforecasts of 1981–2010 Niño-3.4 vs lead time (solid), compared to corresponding values of RMSE (long-dashed curves). The black short-dashed curve indicates RMSE of a simple forecast based on persisting anomalies from the month before the initial forecast month. (c) September Arctic sea ice extent (SIE) from 1981 to 2018 CanSIPsv2 reforecasts at indicated lead times (solid colored) vs Had2CIS observations (black). Values from CanSIPsv1 at 3-month lead are shown for comparison (dotted blue). Anomaly correlation values with long-term trends included are indicated in the legend; corresponding values based on linearly detrended predictions and observations appear in parentheses. Reforecast SIE values have additive bias corrections applied so they have the same 1981–2010 means as the observations.

skill as CanCM4 for most variables including T2m, precipitation, and Z500.

The forecast skill of the NAO index is shown in Fig. 13b. In general, the NAO skill is lower than the PNA. The performance of CanSIPsv2 does not seem to be too different from CanSIPsv1. The NAO skill is relatively high in the winter and spring target seasons, and the skill for JFM is slightly improved in CanSIPsv2 comparing to CanSIPsv1 (Table 1). Skill for winter NAO is comparable to or higher than that of many other seasonal prediction systems (Butler et al. 2016; Kim et al. 2012), although is not as skillful as was

reported in Scaife et al. (2014) for GloSea5 of the Met Office.

6) THE MJO

The MJO is the dominant mode of tropical variability on the subseasonal time scale, which is characterized by a large-scale disturbance with zonal wavenumber 1–3 coupled with convection propagating eastward along the equator with a period of 30–60 days (e.g., Zhang 2005). Although with a shorter time scale than ENSO, the tropical convection anomaly of the MJO influences extratropical atmospheric circulation through teleconnections,

FIG. 13. Correlation skill of seasonal mean (a) PNA index and (b) NAO index as a function of target season at 1-month lead time. The vertical bars represent the range of 95% from a bootstrap resampling.

representing an important source of skill for subseasonal to seasonal predictions (e.g., Waliser et al. 2003).

The MJO index used here is the real-time multivariate MJO (RMM) index as defined in Wheeler and Hendon (2004). It is calculated as the projection of the combined fields of 15°S–15°N meridionally averaged outgoing longwave radiation (OLR) and zonal winds at 850 and 200 hPa onto the two leading empirical orthogonal function (EOF) structures as derived using the observed meridionally averaged values of the same variables. For verification, the observed MJO index is derived from the ERA-Interim zonal winds at 850 and 200 hPa and the daily averaged satellite-observed OLR from the National Oceanic and Atmospheric Administration (NOAA) polar-orbiting series of satellites (Liebmann and Smith 1996). The same approach as described in

Vitart (2017) is applied to calculate the forecast MJO index. To verify the skill of MJO, the bivariate correlation (COR) and root-mean-square error (RMSE) of the MJO index are calculated following Lin et al. (2008). COR measures the model's ability to predict the MJO propagating phase, whereas RMSE takes into account errors in both phase and amplitude.

The COR skill of the MJO index is shown in Figs. 14a and 14b for the ensemble forecast of CanSIPsv1 and CanSIPsv2 as a function of lead time in days for the initialization months of extended boreal winter (December–March) and all 12 initialization months, respectively. As can be seen, in the first 20 days of the forecast, CanSIPsv2 has a better MJO skill than CanSIPsv1. For both forecast systems, the MJO skill for the winter forecast is higher than that for the whole year. COR exceeds 0.6 for about 20 forecast days in winter and 12–14 days in all seasons, and exceeds 0.5 up to about 23 days for both winter and all seasons in both CanSIPsv1 and CanSIPsv2. This level of skill is similar to that in most of the subseasonal to seasonal (S2S) models (Vitart 2017). The RMSE skill (Figs. 14c,d) confirms that CanSIPsv2 has a better MJO forecast in the first 20 days. After that, the two systems have a similar MJO skill.

7) SEA ICE

CanSIPsv1 was one of the first seasonal prediction systems to have an interactive sea ice model component that in principle enables seasonal predictions of sea ice. In practice CanSIPsv1 Arctic sea ice prediction skill, though appreciable, is attributable mainly to the long-term trend that, however, tends to be underrepresented in the reforecasts (Sigmond et al. 2013). This suggests considerable room for improvement, and in fact deficiencies in CanSIPsv1 initialization of sea ice were identified that likely degrade skill. These include unrealistic Arctic sea ice extent trends in the data product used to initialize sea ice concentration in the reforecasts, as well as the use of a seasonally varying model-based thickness climatology having no long-term thinning trend.

A further deficiency of sea ice initialization in the CanSIPsv1 reforecasts is that the sea ice concentration (SIC) product that was used is biased low compared to the GDPS analysis, especially during the summer melt season. This contributed to a notable high bias in real

TABLE 1. Seasonal mean PNA and NAO skill at 1-month lead. Numbers in bold are statistically significant at the 0.05 level.

	CanSIPsv1			CanSIPsv2		
	CanCM3	CanCM4	CanSIPsv1	GEM-NEMO	CanCM4i	CanSIPsv2
PNA (DJF)	0.44	0.63	0.57	0.49	0.42	0.48
PNA (JFM)	0.60	0.66	0.69	0.49	0.54	0.54
NAO (DJF)	0.38	0.41	0.46	0.42	0.33	0.45
NAO (JFM)	0.32	0.35	0.35	0.44	0.38	0.45

FIG. 14. (a) MJO COR skill for the forecast initialized in DJFM, (b) MJO COR skill for all initialization months, (c) MJO RMSE for the forecast initialized in DJFM, and (d) MJO RMSE for all initialization months, as a function of lead time in days. The black and green dashed lines in (a) and (b) indicate the skill values of 0.5 and 0.6, respectively.

time CanSIPsv1 forecasts of Arctic sea ice extent, which were not useful largely for that reason. The modifications to the initialization of sea ice in the CanCM4i forecasts and reforecasts described in sections 3 and 4 are intended to address the above issues. As an example of how these modifications impact the predictive skill of sea ice, Figs. 15a and 15b compares anomaly correlation skills for reforecasts of September SIC at 4-month lead (initialized at the beginning of May) in CanCM4 and CanCM4i.

Further improvement in sea ice skill in CanSIPsv2 is provided by GEM-NEMO, which is even more skillful than CanCM4i (e.g., global mean anomaly correlation over ice-covered regions for September SIC at 4-month lead near 0.30). This overall improvement is illustrated in Figs. 15c and 15b, which compares anomaly correlation reforecast skills for prediction of September SIC at 4-month lead in CanSIPsv1 and CanSIPsv2. As can be seen, the correlation skill averaged over sea ice covered regions is far higher in CanSIPsv2.

Figure 12c shows Had2CIS observations and CanSIPsv2 reforecasts of Arctic sea ice extent (SIE, defined as the summed area of Northern Hemisphere grid cells for which $SIC \geq 15\%$) for September, normally the month of minimum SIE, for various lead times. Large anomaly correlation at all lead times is substantially attributable to the strong decreasing trend in SIE, although it remains >0.5 at a 3-month lead when linear trends are removed (values in parentheses). The decreasing trend is captured fairly well in the CanSIPsv2 reforecasts although longer-lead predictions for the most recent years tend to be lower than observed. By contrast, CanSIPsv1 reforecasts at lead 3 show little decreasing trend due to the sea ice initialization deficiencies discussed above.

6. Teleconnections

In this section, we analyze two processes that are important for the forecast skill on the subseasonal to

FIG. 15. Correlation skill for September monthly mean SIC at 4-month lead (initialized at 1 May). Verified against Had2CIS.

seasonal time scales, which are ENSO and the MJO. Both processes influence extratropical weather and climate through atmospheric teleconnections.

a. ENSO

A major source of skill for seasonal prediction is the ENSO variability (e.g., Kumar and Hoerling 1995; Shukla et al. 2000; Derome et al. 2001). In order for a seasonal prediction system to be skillful, it needs to capture and predict the ENSO variability and the remote influence of ENSO (i.e., its related teleconnections). Figure 16 compares the forecast SST variability of the CanSIPsv2 models with that of the observation in the tropical Pacific. Here we focus on the boreal winter season, considering that the ENSO-related teleconnection is the strongest in the Northern Hemisphere during this season, when a strong extratropical upper-tropospheric westerly jet stream favors the two-dimensional Rossby wave propagation (e.g., Hoskins and Karoly 1981). Standard deviation of the DJF mean SST anomalies over the 30 years of 1981–2010 is calculated. For the forecasting models, seasonal means at 1-month lead (i.e., initialized on 1 November) are used. Both CanCM4i and GEM-NEMO capture the main features of the observations with maximum SST variability in the central to eastern equatorial Pacific. CanCM4i has maximum SST variability related to ENSO in the equatorial Pacific that

is too strong and extended too far to the west, consistent with that in CanCM4 as discussed in Merryfield et al. (2013). GEM-NEMO appears to simulate well the SST variability pattern, but it also seems to extend a little too far west and its strength is slightly overestimated.

To analyze the ENSO-related teleconnection, linear regressions are calculated for DJF mean Z500 and T2m with respect to the Niño-3.4 index. Here we focus on the linear association between the atmospheric circulation and ENSO, although there exists a nonlinear component of the atmospheric response to tropical forcing which is normally secondary (e.g., Hoerling et al. 1997; Lin et al. 2007). Shown in Fig. 17 is the regression of DJF mean Z500 to the Niño-3.4 index. Both CanCM4i and GEM-NEMO are able to simulate the main features of the atmospheric circulation anomaly pattern associated with ENSO, which is similar to the PNA (e.g., Wallace and Gutzler 1981). CanCM4i is seen to overestimate the amplitude of the observed regression. The North Pacific negative Z500 anomaly center of CanCM4i tends to be too strong and shifted about 15° westward compared to the observations. Both CanCM4i and GEM-NEMO simulate the positive anomaly center over Canada too far to the west of that in the observations. Figure 18 shows the regression of DJF mean T2m to the Niño-3.4 index. The general patterns of the surface air temperature anomaly associated with ENSO in the two forecasting

FIG. 16. Standard deviation of DJF mean SST anomalies for (a) OISST V2, (b) CanCM4i at 1-month lead, and (c) GEM-NEMO at 1-month lead.

models are similar to the observations. Warm anomalies are seen over Canada during the positive phase of ENSO, as observed in previous studies (e.g., Ropelewski and Halpert 1986; Shabbar and Barnston 1996). However, the two forecasting models simulate the warm anomaly center in western Canada, instead of central Canada as in the observations. CanCM4i predicts significant cold anomalies in the northern Eurasian continent during El Niño winter, but in the observations the signal is weak. It should be noted that the observational estimate of regression pattern (Fig. 18a) has more uncertainty than the model counterparts that are based on ensembles of 10 realizations.

To assess the impact of ENSO on the forecast skill, we examine the forecast skill of CanSIPsv2 in years with strong and weak ENSO signals, following an approach similar to Derome et al. (2001). Among the 30 years of hindcast, 15 years are selected as the extreme phase of ENSO (EPSO) years, which correspond to those years with absolute values of DJF mean Niño-3.4 index greater

than 0.7 standard deviation. The remaining 15 years with small absolute values of Niño-3.4 are grouped as the non-EPSO (or NEPSO) years. Figure 19 shows the correlation skill of 1-month lead DJF mean Z500 for the EPSO and NEPSO years in the CanSIPsv1 and CanSIPsv2 systems. It is clear that in the extratropics especially in the PNA region, the forecast skill in the EPSO years is superior to that in the NEPSO years. The influence of ENSO on the forecast skill is evident in both CanSIPsv1 and CanSIPsv2. Table 2 summarizes the forecast skill of 1-month lead DJF mean Z500 averaged over the globe, Northern Hemisphere (30°–90°N) (NH), tropics (30°S–30°N), and Southern Hemisphere (90°–30°S) (SH) during the EPSO and NEPSO years for CanSIPsv1 and CanSIPsv2. Under each correlation skill in the table is the percentage of the area with significant skill at the 0.05 level for that area. As can be seen, DJF mean Z500 forecasts in EPSO years are more skillful than NEPSO years in all regions. The difference is especially large for the extratropical Northern and Southern Hemispheres. For CanSIPsv2, the percentages of area with significant skill are 30.9% and 42.1% in NH and SH, respectively, in EPSO years comparing to 4.8% and 7.1% in NEPSO years. In general, CanSIPsv2 outperforms CanSIPsv1 in both EPSO and NEPSO years.

The ENSO impact also extends to the forecast skill of the teleconnection patterns. As shown in Figs. 19c and 19d, the DJF mean Z500 skill in EPSO years has centers in the Pacific and North American region, contributing to the PNA skill. In CanSIPsv2, the correlation skill of the DJF (JFM) mean PNA index at 1-month lead is 0.67 (0.71) in EPSO years, which is statistically significant at the 0.01 level, comparing to 0.11 (0.25) in NEPSO years, which is not statistically significant. Although there are multiple possible sources for the NAO prediction skill, including anomalies in SST, snow cover, sea ice and stratospheric processes, it is possible that the ENSO variability is partly contributing to the NAO skill as well. In the ENSO teleconnection maps shown in Fig. 17, a band of negative Z500 anomaly is seen along the mid-latitude North Atlantic associated with El Niño, which may be related to the southern branch of the NAO dipole structure. From Figs. 19c and 19d, we see that there is also a high skill area in the midlatitude North Atlantic during EPSO years, indicating that ENSO is possibly contributing to the forecast skill of the NAO. The correlation skill of the DJF mean NAO index at 1-month lead in CanSIPsv2 is 0.55 in EPSO years, which is statistically significant at the 0.05 level, comparing to 0.08 in NEPSO years, which is not statistically significant.

b. MJO

Besides ENSO, the MJO also influences the global circulation and provides a source of skill on the subseasonal

FIG. 17. Regression of DJF mean Z500 with Niño-3.4 for (a) ERA-Interim, (b) CanCM4i at 1-month lead, and (c) GEM-NEMO at 1-month lead. The value corresponds to one standard deviation of the Niño-3.4 index. The shaded areas are statistically significant at the 0.05 level according to a Student's *t* test.

to seasonal time scale (e.g., [NAS 2016](#)). For example, [Lin and Brunet \(2009\)](#) found that surface air temperature over Canada and eastern United States in winter tends to be anomalously warm 10–20 days following the MJO phase 3, which corresponds to enhanced convection in the equatorial Indian Ocean and suppressed convection in the western tropical Pacific. Using statistical models with the MJO as a predictor it is possible to obtain useful skill in predicting North American temperature anomalies beyond 20 days, especially for strong MJO cases (e.g., [Yao et al. 2011](#); [Rodney et al. 2013](#); [Johnson et al. 2014](#)). The MJO influences the variability of the NAO through its teleconnection, which in turn

modulates European weather. [Lin et al. \(2009\)](#) observed that 10–15 days following MJO phase 3 (7), there is a significant increase of probability of occurrence of positive (negative) NAO. Similar results were reported in [Cassou \(2008\)](#).

Here we evaluate how the two CanSIPsv2 models simulate the MJO related teleconnections. Forecast daily Z500 anomalies for each ensemble member during the 30 years of hindcast from 1981 to 2010 are first calculated by removing the ensemble mean model climatology which is a function of lead time. Calculation is made separately for CanCM4i and GEM-NEMO. The forecast MJO index and Z500 anomalies are then

FIG. 18. As in Fig. 17, but for DJF mean T2m.

averaged over consecutive five day periods (pentads). We calculate the lagged composites of Z500 anomaly two pentads following the pentad-averaged MJO phase 3 and phase 7 that have amplitude greater than 1. This means that the composite Z500 anomaly corresponds to the period 11–15 days after the occurrence of MJO phase 3 or 7. The MJO cases are selected from individual ensemble members of the forecast during the first 20 days of forecast when there exists some useful skill of MJO in the two forecasting models as is seen in Fig. 14. To focus on the extended boreal winter season, the forecasts with initialization months from November to March are used. For CanCM4i, 447 cases of MJO phase 3 and 328 cases of MJO phase 7 are selected for the composites, whereas for GEM-NEMO the numbers are

467 and 389, respectively. For the ERA-Interim reanalysis, the composites are made with 60 MJO phase 3 cases and 52 phase 7 cases. The results are compared in Fig. 20. Shown at the upper-right corner of each panel is the composite NAO index. As can be seen, both CanCM4i and GEM-NEMO are able to capture the general features of the observed MJO-related teleconnection pattern (i.e., positive NAO 11–15 days after MJO phase 3 and negative NAO 11–15 days after MJO phase 7). However, the forecast NAO strength is weaker than the observations. Vitart (2017) reported that most S2S models tend to strongly overestimate the intensity of the MJO teleconnections in the North Pacific and underestimate its projections onto the positive or negative phase of the NAO over the North Atlantic basin.

FIG. 19. Correlation skill for 1-month lead DJF Z500 forecasts (initialized on 1 Nov). (a) CanSIPsv1 NEPSO years ($|\text{Niño-3.4}| < 0.7$, 15 years), (b) CanSIPsv2 NEPSO, (c) CanSIPsv1 EPSO ($|\text{Niño-3.4}| > 0.7$, 15 years), and (d) CanSIPsv2 EPSO. The black contour line is for the 0.05 significance level.

This seems to be true for the teleconnections of MJO phase 7 for both CanSIPsv2 models, where the strength of the negative Z500 anomaly in the North Pacific is overestimated and the intensity of the positive Z500 anomaly near Greenland is underestimated. For the teleconnection after MJO phase 3, GEM-NEMO is able to simulate the North Pacific positive Z500 anomaly with a similar amplitude as the observations, while that in CanCM4i is still too strong. GEM-NEMO appears to capture the observed MJO teleconnection pattern and the intensity of the projected NAO better than CanCM4i. By comparing the MJO teleconnections in 10 S2S models, Vitart (2017) suggested that the strength of MJO teleconnections might be dependent on the horizontal resolution of the atmospheric component of the models, and that models with a higher horizontal resolution tend to produce stronger MJO teleconnections. It is possible that the stronger MJO teleconnections in GEM-NEMO than CanCM4i is related to the fact that the

former has a higher horizontal resolution in the atmospheric component.

7. Summary and discussion

In this paper, the second version of the Canadian Seasonal to Interannual Prediction System (CanSIPsv2) is introduced. CanSIPsv2 became operational to produce seasonal predictions at the Canadian Meteorological Centre in July 2019, replacing CanSIPsv1. Like CanSIPsv1, the new system is a multimodel ensemble system, consisting of two global atmosphere–ocean coupled models. One of them has an NWP atmospheric component (i.e., the GEM model) coupled with the NEMO ocean model, to become GEM-NEMO. The other is previously existing CanCM4, the better performing coupled climate model in CanSIPsv1, with improved sea ice initialization, to become CanCM4i.

The ECCO official seasonal prediction is made by running GEM-NEMO and CanCM4i for 12 months near the beginning of each month. A total of 20 ensemble

TABLE 2. Area-averaged correlation skill of 1-month lead DJF mean Z500 and percentage of area with significant correlation skill at the 0.05 level according to Student's t test for EPSO and NEPSO years.

	Global		NH		Tropics		SH	
	NEPSO	EPSO	NEPSO	EPSO	NEPSO	EPSO	NEPSO	EPSO
CanSIPsv1	0.41, 41.9%	0.57, 59.4%	0.13, 5.8%	0.33, 26.1%	0.68, 76.5%	0.79, 88.2%	0.12, 4.5%	0.36, 35.3%
CanSIPsv2	0.45, 44.4%	0.59, 62.2%	0.15, 4.8%	0.36, 30.9%	0.72, 81.8%	0.80, 87.9%	0.20, 7.1%	0.40, 42.1%

FIG. 20. Two-pentad lagged composite of Z500 anomaly following MJO (left) phase 3 and (right) phase 7. The number on the top-right corner is the NAO index.

members, 10 from each model, are produced. The two models take different approaches in the initialization for the forecasts. To calculate the forecast anomaly and to calibrate the seasonal forecast, the model climatology is estimated from the hindcast from the beginning of each month of the 30 years from 1981 to 2010 for each model.

The performance of CanSIPsv2 is evaluated and compared with CanSIPsv1, based on the hindcast of 30 years from 1981 to 2010. Ensemble forecasts of seasonal averages of 2-m air temperature, 500-hPa geopotential height, precipitation rate, sea surface temperature, and sea ice concentration are compared with that of the observations. Verification is also performed on the forecast skill of several climate indices, including the Niño-3.4, PNA, NAO, and MJO indices. It is found that CanSIPsv2 generally outperforms CanSIPsv1. The difference in the initialization procedure between the hindcast and forecast may affect the forecast quality.

The ability of the CanSIPsv2 models in capturing the ENSO and MJO teleconnections is assessed using the data of hindcasts. Both GEM-NEMO and CanCM4i are able to simulate the general features of Z500 and T2m anomalies, especially in the Pacific and North American region, associated with ENSO. Like in other seasonal forecasting systems (e.g., Derome et al. 2001), ENSO is the dominant contributor for seasonal forecast skill in CanSIPsv2, at least in boreal winter. The forecast skill of PNA is influenced by the strength of ENSO, as is the NAO to a lesser extent. On the subseasonal time scale, both GEM-NEMO and CanCM4i do a reasonably good job capturing the Northern Hemisphere extratropical Z500 anomalies following MJO phases that correspond to a dipole structure of convection anomaly in the equatorial Indian Ocean and western Pacific area (i.e., MJO phase 3 or 7).

The improved forecast skill of CanSIPsv2 over CanSIPsv1 results mainly from the replacement of CanCM3 by a better performing model, GEM-NEMO. The new sea ice initialization of CanCM4i clearly leads to better sea ice forecast skill than CanCM4, as demonstrated in section 5b(7) (Fig. 15). However, outside the polar regions, the performance of CanCM4i is in general similar to CanCM4 of CanSIPsv1. Although there is indication that changes in the Arctic sea ice can potentially influence the midlatitude atmospheric circulation and weather (e.g., Cohen et al. 2014; Coumou et al. 2018), it does not appear that the midlatitude forecast skill in CanCM4i benefits from the improved Arctic sea ice initial condition. It would be interesting for future studies to identify and understand factors that might influence the interaction between the Arctic sea ice and midlatitude weather.

The performance of CanSIPsv2 is in general comparable to the seasonal forecast systems of other operational

centers. There are some aspects where CanSIPsv2 is outperforming. For example, Kim et al. (2012) assessed the DJF forecast skill of T2m at 1-month lead time for the ECMWF System 4 and NCEP CFSv2 and found that there is almost no skill near the east coast of North America (their Figs. 2a,b). Although there is some improvement in System 5 of ECMWF (Fig. 19a of Johnson et al. 2019), the correlation skill in eastern Canada is still less than 0.4. From our Fig. 3b, it is seen that CanSIPsv2 does produce correlation skill above 0.6 in eastern Canada and the adjacent North Atlantic area. Further studies would be needed to explore the source of skill over that region.

The hindcast output of CanSIPsv2 provides a large set of data for research on climate variability, seasonal predictability and prediction, as well as on model intercomparisons. The hindcast data for all initialization months from 1981 to 2010 are available at <http://dd.weather.gc.ca/ensemble/cansips/grib2/hindcast/raw/>. These data are monthly averages for each member of CanCM4i and GEM-NEMO, with variables including Z500, precipitation rate, mean sea level pressure, 850-hPa temperature, T2m, zonal and meridional winds at 850 and 200 hPa, and SST. They are provided at $1^\circ \times 1^\circ$ resolution in grib2 format. In addition to the 30 years of 1981–2010 that is used to estimate the model climatology in the operation, we have extended the hindcast for all the initialization months from 2011 to 2018 for both CanCM4i and GEM-NEMO. Comparable hindcast and forecast data produced for the NMME and archived by International Research Institute for Climate and Society (IRI) is available at <https://iridl.ldeo.columbia.edu/SOURCES/.Models/.NMME/>.

Acknowledgments. Members of the Seasonal Forum and many colleagues at RPN, CCCma, and CCMEP contributed to the development and implementation of CanSIPsv2. We thank the following colleagues for their various contributions to this project: Benoit Archambault, Jean-Marc Bélanger, Normand Gagnon, Peter Houtekamer, Ron McTaggart-Cowan, Ayrton Zadra, Paul Vaillancourt, Michel Roch, Stéphane Chamberland, Vivian Lee, Michel Desgagné, Stéphane Bélair, Maria Abrahamowicz, Marco Carrera, Nicola Gasset, and Katja Winger.

REFERENCES

- Adler, R., and Coauthors, 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>.
- Barnston, A. G., and R. E. Livezey, 1987: Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Mon. Wea. Rev.*, **115**, 1083–1126, [https://doi.org/10.1175/1520-0493\(1987\)115<1083:CSAPOL>2.0.CO;2](https://doi.org/10.1175/1520-0493(1987)115<1083:CSAPOL>2.0.CO;2).

- Becker, E., H. Van den Dool, and Q. Zhang, 2014: Predictability and forecast skill in NMME. *J. Climate*, **27**, 5891–5906, <https://doi.org/10.1175/JCLI-D-13-00597.1>.
- Bélair, S., J. Mailhot, C. Girard, and P. Vaillancourt, 2005: Boundary layer and shallow cumulus clouds in a medium-range forecast of a large-scale winter system. *Mon. Wea. Rev.*, **133**, 1938–1960, <https://doi.org/10.1175/MWR2958.1>.
- Bernier, N. B., and S. Bélair, 2012: High horizontal and vertical resolution limited-area model: Near surface and wind energy forecast applications. *J. Appl. Meteor. Climatol.*, **51**, 1061–1078, <https://doi.org/10.1175/JAMC-D-11-0197.1>.
- Bradley, A. A., and S. S. Schwartz, 2011: Summary verification measures and their interpretation for ensemble forecasts. *Mon. Wea. Rev.*, **139**, 3075–3089, <https://doi.org/10.1175/2010MWR3305.1>.
- Branstator, G. W., 2002: Circumglobal teleconnections, the jet stream waveguide, and the North Atlantic Oscillation. *J. Climate*, **15**, 1893–1910, [https://doi.org/10.1175/1520-0442\(2002\)015<1893:CTTJSW>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1893:CTTJSW>2.0.CO;2).
- Buehner, M., and Coauthors, 2015: Implementation of deterministic weather forecasting systems based on ensemble-variational data assimilation at Environment Canada. Part I: The global system. *Mon. Wea. Rev.*, **143**, 2532–2559, <https://doi.org/10.1175/MWR-D-14-00354.1>.
- Butler, A. H., and Coauthors, 2016: The climate-system historical forecast project: Do stratosphere-resolving models make better seasonal climate predictions in boreal winter? *Quart. J. Roy. Meteor. Soc.*, **142**, 1413–1427, <https://doi.org/10.1002/qj.2743>.
- Carrera, M. L., S. Bélair, V. Fortin, B. Bilodeau, D. Charpentier, and I. Doré, 2010: Evaluation of snowpack simulations over the Canadian Rockies with an experimental hydrometeorological modeling system. *J. Hydrometeorol.*, **11**, 1123–1140, <https://doi.org/10.1175/2010JHM1274.1>.
- Cassou, C., 2008: Intraseasonal interaction between the Madden-Julian Oscillation and the North Atlantic oscillation. *Nature*, **455**, 523–527, <https://doi.org/10.1038/nature07286>.
- Cohen, J., and Coauthors, 2014: Recent Arctic amplification and extreme mid-latitude weather. *Nat. Geosci.*, **7**, 627–637, <https://doi.org/10.1038/ngeo2234>.
- Côté, J., S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998: The operational CMC-MRB Global Environmental Multiscale (GEM) model: Part I—Design considerations and formulation. *Mon. Wea. Rev.*, **126**, 1373–1395, [https://doi.org/10.1175/1520-0493\(1998\)126<1373:TOCMGE>2.0.CO;2](https://doi.org/10.1175/1520-0493(1998)126<1373:TOCMGE>2.0.CO;2).
- Coumou, D., G. Di Capua, S. Vavrus, I. Wang, and S. Wang, 2018: The influence of Arctic amplification on mid-latitude summer circulation. *Nat. Commun.*, **9**, 2959, <https://doi.org/10.1038/s41467-018-05256-8>.
- Dee, D. P., and Coauthors, 2011: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quart. J. Roy. Meteor. Soc.*, **137**, 553–597, <https://doi.org/10.1002/qj.828>.
- DelSole, T., and M. K. Tippett, 2016: Forecast comparison based on random walks. *Mon. Wea. Rev.*, **144**, 615–626, <https://doi.org/10.1175/MWR-D-15-0218.1>.
- Derome, J., G. Brunet, A. Plante, N. Gagnon, G. J. Boer, F. W. Zwiers, S. Lambert, and H. Ritchie, 2001: Seasonal predictions based on two dynamical models. *Atmos.–Ocean*, **39**, 485–501, <https://doi.org/10.1080/07055900.2001.9649690>.
- Ding, Q., B. Wang, J. M. Wallace, and G. Branstator, 2011: Tropical–extratropical teleconnections in boreal summer: Observed interannual variability. *J. Climate*, **24**, 1878–1896, <https://doi.org/10.1175/2011JCLI3621.1>.
- Dirkson, A., W. J. Merryfield, and A. Monahan, 2017: Impacts of sea ice thickness initialization on seasonal Arctic sea ice predictions. *J. Climate*, **30**, 1001–1017, <https://doi.org/10.1175/JCLI-D-16-0437.1>.
- Flato, G. M., and W. D. Hibler, 1992: Modelling pack ice as a cavitating fluid. *J. Phys. Oceanogr.*, **22**, 626–651, [https://doi.org/10.1175/1520-0485\(1992\)022<0626:MPIAAC>2.0.CO;2](https://doi.org/10.1175/1520-0485(1992)022<0626:MPIAAC>2.0.CO;2).
- Gagnon, N., and Coauthors, 2015: Global Ensemble Prediction System (GEPS): Update from version 4.0.1 to version 4.1.1. Canadian Meteorological Centre Tech. Note, Environment Canada, 36 pp., http://collaboration.cmc.ec.gc.ca/cmc/cmoin/product_guide/docs/lib/technote_geps-411_20151215_e.pdf.
- Gauthier, P., M. Buehner, and L. Fillion, 1999: Background-error statistics modelling in a 3D variational data assimilation scheme: Estimation and impact on the analyses. *Proc. ECMWF Workshop on Diagnosis of Data Assimilation Systems*, Reading, United Kingdom, ECMWF, 131–145.
- Girard, C., and Coauthors, 2014: Staggered vertical discretization of the Canadian Environmental Multiscale (GEM) model using a coordinate of the log-hydrostatic-pressure type. *Mon. Wea. Rev.*, **142**, 1183–1196, <https://doi.org/10.1175/MWR-D-13-00255.1>.
- Graham, R., and Coauthors, 2011: Long-range forecasting and global framework for climate services. *Climate Res.*, **47**, 47–55, <https://doi.org/10.3354/cr00963>.
- Ho, C. K., E. Hawkins, L. Shaffrey, J. Broecker, L. Hermanson, J. M. Murphy, and D. M. Smith, 2013: Examining reliability of seasonal to decadal sea surface temperature forecasts: The role of ensemble dispersion. *Geophys. Res. Lett.*, **40**, 5770–5775, <https://doi.org/10.1002/2013GL057630>.
- Hoerling, M. P., A. Kumar, and M. Zhong, 1997: El Niño, La Niña, and the nonlinearity of their teleconnections. *J. Climate*, **10**, 1769–1786, [https://doi.org/10.1175/1520-0442\(1997\)010<1769:ENOLNA>2.0.CO;2](https://doi.org/10.1175/1520-0442(1997)010<1769:ENOLNA>2.0.CO;2).
- Hoskins, B. J., and D. J. Karoly, 1981: The steady linear response of a spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci.*, **38**, 1179–1196, [https://doi.org/10.1175/1520-0469\(1981\)038<1179:TSLROA>2.0.CO;2](https://doi.org/10.1175/1520-0469(1981)038<1179:TSLROA>2.0.CO;2).
- Houtekamer, P. L., H. L. Mitchell, and X. Deng, 2009: Model error representation in an operational ensemble Kalman filter. *Mon. Wea. Rev.*, **137**, 2126–2143, <https://doi.org/10.1175/2008MWR2737.1>.
- , X. Deng, H. L. Mitchell, S.-J. Baek, and N. Gagnon, 2014: Higher resolution in an operational ensemble Kalman filter. *Mon. Wea. Rev.*, **142**, 1143–1162, <https://doi.org/10.1175/MWR-D-13-00138.1>.
- Hu, Z.-Z., A. Kumar, B. Huang, W. Wang, J. Zhu, and C. Wen, 2013: Prediction skill of monthly SST in the North Atlantic Ocean in NCEP climate forecast system version 2. *Climate Dyn.*, **40**, 2745–2759, <https://doi.org/10.1007/s00382-012-1431-z>.
- Hunke, E. C., and W. H. Lipscomb, 2010: CICE: The Los Alamos sea ice model, documentation and software user's manual, version 4.1. Doc. LA-CC-06-012, 76 pp., http://csdms.colorado.edu/w/images/CICE_documentation_and_software_user's_manual.pdf.
- Hurrell, J. W., Y. Kushnir, M. Visbeck, and G. Ottersen, 2003: An overview of the North Atlantic oscillation. *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*, *Geophys. Monogr.*, Vol. 134, Amer. Geophys. Union, 1–35.
- Ioannidou, L., W. Yu, and S. Bélair, 2014: Forecasting of surface winds over eastern Canada using the Canadian offline land surface modeling system. *J. Appl. Meteor. Climatol.*, **53**, 1760–1774, <https://doi.org/10.1175/JAMC-D-12-0284.1>.
- Johansson, A., 2007: Prediction skill of the NAO and PNA from daily to seasonal time scales. *J. Climate*, **20**, 1957–1975, <https://doi.org/10.1175/JCLI4072.1>.

- Johnson, N. C., D. C. Collins, S. B. Feldstein, M. L. L'Heureux, and E. E. Riddle, 2014: Skillful wintertime North American temperature forecasts out to four weeks based on the state of ENSO and the MJO. *Wear. Forecasting*, **29**, 23–38, <https://doi.org/10.1175/WAF-D-13-00102.1>.
- Johnson, S. J., and Coauthors, 2019: SEAS5: The new ECMWF seasonal forecast system. *Geosci. Model Dev.*, **12**, 1087–1117, <https://doi.org/10.5194/gmd-12-1087-2019>.
- Kain, J. S., and J. M. Fritsch, 1990: A one-dimensional entraining detraining plume model and its application in convective parameterization. *J. Atmos. Sci.*, **47**, 2784–2802, [https://doi.org/10.1175/1520-0469\(1990\)047<2784:AODEPM>2.0.CO;2](https://doi.org/10.1175/1520-0469(1990)047<2784:AODEPM>2.0.CO;2).
- Kharin, V. V., Q. Teng, F. W. Zwiers, G. J. Boer, J. Derome, and J. S. Fontecilla, 2009: Skill assessment of seasonal hindcasts from the Canadian historical forecast project. *Atmos.–Ocean*, **47**, 204–223, <https://doi.org/10.3137/AO1101.2009>.
- , W. J. Merryfield, G. J. Boer, and W.-S. Lee, 2017: A post-processing method for seasonal forecasts using temporally and spatially smoothed statistics. *Mon. Wea. Rev.*, **145**, 3545–3561, <https://doi.org/10.1175/MWR-D-16-0337.1>.
- Kim, H.-M., P. J. Webster, and J. A. Curry, 2012: Seasonal prediction skill of ECMWF system 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere winter. *Climate Dyn.*, **39**, 2957–2973, <https://doi.org/10.1007/s00382-012-1364-6>.
- Kirtman, B. P., and Coauthors, 2014: The North American multimodel ensemble: Phase-1: Seasonal to interannual prediction; phase-2 toward developing intraseasonal prediction. *Bull. Amer. Meteor. Soc.*, **95**, 585–601, <https://doi.org/10.1175/BAMS-D-12-00050.1>.
- Kistler, R., and Coauthors, 2001: The NCEP–NCAR 50-Year Reanalysis: Monthly means CD-ROM and documentation. *Bull. Amer. Meteor. Soc.*, **82**, 247–268, [https://doi.org/10.1175/1520-0477\(2001\)082<0247:TNNYRM>2.3.CO;2](https://doi.org/10.1175/1520-0477(2001)082<0247:TNNYRM>2.3.CO;2).
- Klasa, M., J. Derome, and J. Sheng, 1992: On the interaction between the synoptic-scale eddies and the PNA teleconnection pattern. *Beitr. Phys. Atmos.*, **65**, 211–222.
- Krishnamurti, T. N., C. M. Kishtawal, T. E. LaRow, D. R. Bachiocchi, Z. Zhang, C. E. Williford, S. Gadgil, and S. Surendran, 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., 2009: Finite samples and uncertainty estimates for skill measures for seasonal predictions. *Mon. Wea. Rev.*, **137**, 2622–2631, <https://doi.org/10.1175/2009MWR2814.1>.
- , and M. P. Hoerling, 1995: Prospects and limitations of atmospheric GCM climate predictions. *Bull. Amer. Meteor. Soc.*, **76**, 335–345, [https://doi.org/10.1175/1520-0477\(1995\)076<0335:PALOSA>2.0.CO;2](https://doi.org/10.1175/1520-0477(1995)076<0335:PALOSA>2.0.CO;2).
- Lau, N.-C., 1988: Variability of the observed midlatitude storm tracks in relation to low-frequency changes in the circulation pattern. *J. Atmos. Sci.*, **45**, 2718–2743, [https://doi.org/10.1175/1520-0469\(1988\)045<2718:VOTOMS>2.0.CO;2](https://doi.org/10.1175/1520-0469(1988)045<2718:VOTOMS>2.0.CO;2).
- Liebmann, B., and C. A. Smith, 1996: Description of a complete (interpolated) outgoing longwave radiation dataset. *Bull. Amer. Meteor. Soc.*, **77**, 1275–1277.
- Lin, H., and G. Brunet, 2009: The influence of the Madden–Julian oscillation on Canadian wintertime surface air temperature. *Mon. Wea. Rev.*, **137**, 2250–2262, <https://doi.org/10.1175/2009MWR2831.1>.
- , J. Derome, and G. Brunet, 2007: The nonlinear transient atmospheric response to tropical forcing. *J. Climate*, **20**, 5642–5665, <https://doi.org/10.1175/2007JCLI1383.1>.
- , G. Brunet, and J. Derome, 2008: Forecast skill of the Madden–Julian oscillation in two Canadian atmospheric models. *Mon. Wea. Rev.*, **136**, 4130–4149, <https://doi.org/10.1175/2008MWR2459.1>.
- , —, and —, 2009: An observed connection between the North Atlantic Oscillation and the Madden–Julian oscillation. *J. Climate*, **22**, 364–380, <https://doi.org/10.1175/2008JCLI2515.1>.
- , —, and B. Yu, 2015: Interannual variability of the Madden–Julian Oscillation and its impact on the North Atlantic oscillation in the boreal winter. *Geophys. Res. Lett.*, **42**, 5571–5576, <https://doi.org/10.1002/2015GL064547>.
- , N. Gagnon, S. Beauregard, R. Muncaster, M. Markovic, B. Denis, and M. Charron, 2016: GEPS based monthly prediction at the Canadian Meteorological Centre. *Mon. Wea. Rev.*, **144**, 4867–4883, <https://doi.org/10.1175/MWR-D-16-0138.1>.
- Lin, J.-L., 2007: The double-ITCZ problem in IPCCAR4 coupled GCMs: Ocean-atmosphere feedback analysis. *J. Climate*, **20**, 4497–4525, <https://doi.org/10.1175/JCLI4272.1>.
- MacLachlan, C., and Coauthors, 2015: Global Seasonal forecast system version 5 (GloSea5): A high-resolution seasonal forecast system. *Quart. J. Roy. Meteor. Soc.*, **141**, 1072–1084, <https://doi.org/10.1002/qj.2396>.
- Madden, R. A., and P. R. Julian, 1971: Description of a 40–50 day oscillation in the zonal wind in the tropical Pacific. *J. Atmos. Sci.*, **28**, 702–708, [https://doi.org/10.1175/1520-0469\(1971\)028<0702:DOADOI>2.0.CO;2](https://doi.org/10.1175/1520-0469(1971)028<0702:DOADOI>2.0.CO;2).
- Merryfield, W. J., and Coauthors, 2013: The Canadian Seasonal to Interannual Prediction System. Part I: Models and initialization. *Mon. Wea. Rev.*, **141**, 2910–2945, <https://doi.org/10.1175/MWR-D-12-00216.1>.
- Min, Y.-M., V. N. Kryjov, and S. M. Oh, 2014: Assessment of APCC multimodel ensemble prediction in seasonal climate forecasting: Retrospective (1983–2003) and real-time forecasts (2008–2013). *J. Geophys. Res. Atmos.*, **119**, 12 132–12 150, <https://doi.org/10.1002/2014JD022230>.
- NAS, 2016: *Next Generation Earth System Prediction: Strategies for Subseasonal to Seasonal Forecasts*. National Academies Press, 350 pp., <https://doi.org/10.17226/21873>.
- Noilhan, J., and S. Planton, 1989: A simple parameterization of land surface processes for meteorological models. *Mon. Wea. Rev.*, **117**, 536–549, [https://doi.org/10.1175/1520-0493\(1989\)117<0536:ASPOLS>2.0.CO;2](https://doi.org/10.1175/1520-0493(1989)117<0536:ASPOLS>2.0.CO;2).
- , and J. F. Mahfouf, 1996: The ISBA land surface parameterisation scheme. *Global Planet. Change*, **13**, 145–159, [https://doi.org/10.1016/0921-8181\(95\)00043-7](https://doi.org/10.1016/0921-8181(95)00043-7).
- Reynolds, R. W., N. A. Rayner, T. M. Smith, D. C. Stokes, and W. Wang, 2002: An improved in situ and satellite SST analysis for climate. *J. Climate*, **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).
- Ritchie, H., 1991: Application of the semi-Lagrangian method to a multilevel spectral primitive-equation model. *Quart. J. Roy. Meteor. Soc.*, **117**, 91–106, <https://doi.org/10.1002/qj.49711749705>.
- Rodney, M., H. Lin, and J. Derome, 2013: Subseasonal prediction of wintertime North American surface air temperature during strong MJO events. *Mon. Wea. Rev.*, **141**, 2897–2909, <https://doi.org/10.1175/MWR-D-12-00221.1>.
- Ropelewski, C. F., and M. S. Halpert, 1986: North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO). *Mon. Wea. Rev.*, **114**, 2352–2362, [https://doi.org/10.1175/1520-0493\(1986\)114<2352:NAPATP>2.0.CO;2](https://doi.org/10.1175/1520-0493(1986)114<2352:NAPATP>2.0.CO;2).

- Saha, S., and Coauthors, 2014: The NCEP Climate Forecast System version 2. *J. Climate*, **27**, 2185–2208, <https://doi.org/10.1175/JCLI-D-12-00823.1>.
- Scaife, A. A., and Coauthors, 2014: Skillful long-range prediction of European and North American winters. *Geophys. Res. Lett.*, **41**, 2514–2519, <https://doi.org/10.1002/2014GL059637>.
- Scinocca, J. F., N. A. McFarlane, M. Lazare, and J. Li, 2008: The CCCma third generation AGCM and its extension into the middle atmosphere. *Atmos. Chem. Phys.*, **8**, 7055–7074, <https://doi.org/10.5194/acp-8-7055-2008>.
- Separovic, L., S. Z. Husain, W. Yu, and D. Fernig, 2014: High-resolution surface analysis for extended-range downscaling with limited-area atmospheric models. *J. Geophys. Res. Atmos.*, **119**, 13 651–13 682, <https://doi.org/10.1002/2014JD022387>.
- Shabbar, A., and A. G. Barnston, 1996: Skill of seasonal climate forecasts in Canada using canonical correlation analysis. *Mon. Wea. Rev.*, **124**, 2370–2385, [https://doi.org/10.1175/1520-0493\(1996\)124<2370:SOSCFI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1996)124<2370:SOSCFI>2.0.CO;2).
- Shukla, J. S., and Coauthors, 2000: Dynamical seasonal prediction. *Bull. Amer. Meteor. Soc.*, **81**, 2593–2606, [https://doi.org/10.1175/1520-0477\(2000\)081<2593:DSP>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<2593:DSP>2.3.CO;2).
- Sigmond, M., J. C. Fyfe, G. M. Flato, V. V. Kharin, and W. J. Merryfield, 2013: Seasonal forecast skill of Arctic sea ice area in a dynamical forecast system. *Geophys. Res. Lett.*, **40**, 529–534, <https://doi.org/10.1002/grl.50129>.
- Smith, G. C., F. Davidson, and Y. Lu, 2013: The CONCEPTS initiative: Canadian operational network of coupled environmental prediction systems. *J. Ocean Technol.*, **8**, 80–81.
- , and Coauthors, 2016: Sea ice forecast verification in the Canadian global ice ocean prediction system. *Quart. J. Roy. Meteor. Soc.*, **142**, 659–671, <https://doi.org/10.1002/qj.2555>.
- , and Coauthors, 2018: Impact of coupling with an ice-ocean model on global medium-range NWP forecast skill. *Mon. Wea. Rev.*, **146**, 1157–1180, <https://doi.org/10.1175/MWR-D-17-0157.1>.
- Titchner, H. A., and N. A. Rayner, 2014: The Met Office Hadley Centre sea ice and sea surface temperature data set, version 2: 1. Sea ice concentrations. *J. Geophys. Res. Atmos.*, **119**, 2864–2889, <https://doi.org/10.1002/2013JD020316>.
- Tivy, A., S. E. L. Howell, B. Alt, S. McCourt, R. Chagnon, G. Crocker, T. Carrieres, and J. J. Yackel, 2011: Trends and variability in summer sea ice cover in the Canadian Arctic based on the Canadian Ice Service Digital Archive, 1960–2008 and 1968–2008. *J. Geophys. Res.*, **116**, C03007, <https://doi.org/10.1029/2009JC005855>.
- Tompkins, A. M., and Coauthors, 2017: The Climate-System Historical Forecast Project: Providing open access to seasonal forecast ensembles from centers around the globe. *Bull. Amer. Meteor. Soc.*, **98**, 2293–2301, <https://doi.org/10.1175/BAMS-D-16-0209.1>.
- Versegny, D. L., 2000: The Canadian Land Surface Scheme (CLASS): Its history and future. *Atmos.–Ocean*, **38**, 1–13, <https://doi.org/10.1080/07055900.2000.9649637>.
- Vitart, F., 2017: Madden-Julian Oscillation prediction and teleconnections in the S2S database. *Quart. J. Roy. Meteor. Soc.*, **143**, 2210–2220, <https://doi.org/10.1002/qj.3079>.
- Waliser, D. E., K.-M. Lau, W. Stern, and C. Jones, 2003: Potential predictability of the Madden-Julian oscillation. *Bull. Amer. Meteor. Soc.*, **84**, 33–50, <https://doi.org/10.1175/BAMS-84-1-33>.
- Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. 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).
- Wheeler, M., and H. H. Hendon, 2004: An all-season real-time multivariate MJO index: Development of an index for monitoring and prediction. *Mon. Wea. Rev.*, **132**, 1917–1932, [https://doi.org/10.1175/1520-0493\(2004\)132<1917:AARMMI>2.0.CO;2](https://doi.org/10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2).
- World Meteorological Organization, 2017: WMO guidelines on the calculation of climate normals. WMO-1203, 29 pp., https://library.wmo.int/doc_num.php?explnum_id=4166.
- Yao, W., H. Lin, and J. Derome, 2011: Submonthly forecasting of winter surface air temperature in North America based on organized tropical convection. *Atmos.–Ocean*, **49**, 51–60, <https://doi.org/10.1080/07055900.2011.556882>.
- Zhang, C., 2005: Madden-Julian Oscillation. *Rev. Geophys.*, **43**, RG2003, <https://doi.org/10.1029/2004RG000158>.
- Zhang, J., and D. A. Rothrock, 2003: Modeling global sea ice with a thickness and enthalpy distribution model in generalized curvilinear coordinates. *Mon. Wea. Rev.*, **131**, 845–861, [https://doi.org/10.1175/1520-0493\(2003\)131<0845:MGSIWA>2.0.CO;2](https://doi.org/10.1175/1520-0493(2003)131<0845:MGSIWA>2.0.CO;2).
- Zuo, H., M. A. Balmaseda, and K. Mogensen, 2017: The new eddy-permitting ORAP5 ocean reanalysis: Description, evaluation and uncertainties in climate signals. *Climate Dyn.*, **49**, 791–811, <https://doi.org/10.1007/s00382-015-2675-1>.