Evaluation of Soil Moisture in the Canadian Seasonal to Interannual Prediction System, Version 2.1 (CanSIPSv2.1)

REINEL SOSPEDRA-ALFONSO,* WILLIAM J. MERRYFIELD,* VIATSHESLAV V. KHARIN,* WOO-SUNG LEE,* HAI LIN,* GULILAT T. DIRO,* AND RYAN MUNCASTER*

*a Canadian Centre for Climate Modelling and Analysis, Environment and Climate Change Canada, Victoria, British Columbia, Canada
*b Recherche en Prévision Numérique Atmosphérique, Environment and Climate Change Canada, Dorval, Quebec, Canada
*c Canadian Meteorological and Environmental Prediction Center, Environment and Climate Change Canada, Dorval, Quebec, Canada

ABSTRACT: We evaluate the soil moisture hindcasts and the reconstruction runs giving the hindcasts initial conditions in version 2.1 of the Canadian Seasonal to Interannual Prediction System (CanSIPSv2.1). Different strategies are used to initialize the hindcasts for the two CanSIPSv2.1 models, CanCM4i and the coupled Global Environmental Multiscale, version 5.1, (GEM5)-NEMO model (GEM5-NEMO), with contrasting impacts on the soil moisture initial conditions and forecast performance. Forecast correlation skill is decomposed into contributions from persistence of the initial anomalies and contributions not linked to persistence, with performance largely driven by the accuracy of the initial conditions in regions of strong persistence. Seasonal soil moisture correlation skill is significant for several months into the hindcasts depending on initial and target months, with contributions not linked to persistence becoming more notable at longer lead times. For the first 2–4 months, the quality of CanSIPSv2.1 ensemble mean forecasts tends to be higher on average during summer and fall and is comparable to that of the best performing model, whereas CanSIPSv2.1 outperforms the single models during spring and winter. For longer lead times, remote climate influences from the Pacific Ocean are notable and contribute to predictable soil moisture variability in teleconnected regions.

KEYWORDS: Climate prediction; Ensembles; Forecast verification/skill; Hindcasts; Operational forecasting; Seasonal forecasting

1. Introduction

Soil moisture refers to the total amount of water, including ice and water vapor, in an unsaturated soil. The water stored in the soil controls many processes and feedback interactions within the climate system (Seneviratne et al. 2010), and its temporal and spatial variability is largely responsible for the complexities of these interactions. Soil moisture plays a key role in the land surface water and energy balance through evapotranspiration, the modulation of thermal properties at the soil–atmosphere interface, and the recycling of precipitation. By controlling the energy partition between latent and sensible heat fluxes at the soil–atmosphere interface, soil moisture can affect air temperature and the occurrence of convective precipitation (Crow et al. 2020; Eltahir 1998). Soil moisture can amplify climate extremes, including heat waves and droughts (Jaeger and Seneviratne 2011; Roundy and Santanello 2017; Zeri et al. 2022), and can contribute to surface and subsurface runoff with a direct impact to floods (Beljaars et al. 1996; Massari et al. 2014; Grillakis et al. 2016). The ability to predict soil moisture is therefore critical for drought and flood prediction (Esit et al. 2021; Wasko et al. 2020), streamflow prediction (Mahanama et al. 2008), forest fire prediction (O et al. 2020), water resource allocation (Pezij et al. 2019), agricultural monitoring (Bolten and Crow 2012), and irrigation management (Hanson et al. 2000).

Soil moisture has a distinctive persistence, with anomalous states that can last for several weeks, months, and years after the atmospheric conditions responsible for the wet or dry anomaly have dissipated (Koster and Suarez 2001; Seneviratne et al. 2006; Orth and Seneviratne 2012). This soil moisture “memory” is associated with the weak damping effect of soil hydrological processes, and it has major implications for subseasonal and seasonal forecasting (van den Hurk et al. 2012; Guo et al. 2012; Sospedra-Alfonso and Merryfield 2018). Because soils can integrate precipitation and evaporation noise to produce a predictable signal (Chikamoto et al. 2015), soil moisture can enhance predictability and boost forecast skill. A proper land initialization, and particularly soil moisture initialization, is therefore expected to have a significant impact on subseasonal, seasonal or even decadal prediction skill (Koster et al. 2010; Esit et al. 2021).

Drivers of soil moisture variability can be local or remote. Ocean surface conditions, for instance, are associated to temperature and precipitation variations that affect soil water content (Nicolai-Shaw et al. 2016; Sospedra-Alfonso and Merryfield 2018). El Niño–Southern Oscillation (ENSO), combined with the slower variations of the Pacific decadal oscillation (PDO), has been shown to play a significant role in summer moisture availability in North America (e.g., Shabbar and Skinner 2004; Rippey 2015). El Niño events, which are characterized by the

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

Supplemental information related to this paper is available at the Journals Online website: https://doi.org/10.1175/JAMC-D-23-0034.s1.

Corresponding author: Reinel Sospedra-Alfonso, reinel.sospedra-alfonso@ec.gc.ca

DOI: 10.1175/JAMC-D-23-0034.1

For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).
warming of the central and/or eastern equatorial Pacific Ocean, are linked to summer moisture deficit in western Canada (Shabbar and Skinner 2004), whereas La Niña events, characterized by the cooling of equatorial Pacific waters, are associated to an excess of summer moisture in western Canada and the Canadian prairies, and the occurrence of droughts across the southern United States (Rippey 2015). A proper representation of global climate variability and tropical–extratropical interactions can thus have a positive impact on predictions of surface land conditions, in particular, of soil moisture.

Prediction skill of soil moisture forecasts on seasonal time scales have been examined by Kanamitsu et al. (2003) using the NCEP forecasting system, and by Yao and Yuan (2018) using predictions from the North American Multi-Model Ensemble (NMME) over China. Sospedra-Alfonso and Merryfield (2018) assessed potential predictability of soil moisture in a previous version of the Canadian seasonal prediction system, although a comprehensive assessment of prediction skill has, until now, been lacking. On seasonal-to-decadal time scales, Esit et al. (2021) studied drought and soil moisture forecasting using predictions and historical simulations from the Community Earth System Model (CESM) large ensembles, whereas Chikamoto et al. (2015) examined soil moisture predictability with CESM “perfect model” initialized simulations. These studies address the effects of ocean and land initialization on soil moisture predictability using different methodologies, regions, evaluation periods, soil depth, and so forth.

In this work, we examine the representation of soil moisture in retrospective forecasts (hindcasts) produced with the Canadian Seasonal to Interannual Prediction System, version 2.1 (CanSIPSv2.1; Lin et al. 2020, 2021). In doing so, we assess the roles of anomaly persistence and ocean-driven variability in the performance of the forecasts. CanSIPSv2.1, implemented in December 2021, is the Environment and Climate Change Canada’s (ECCC) two-model seasonal forecasting system, which combines forecasts from coupled climate models CanCM4i, developed at the Canadian Centre for Climate Modeling and Analysis (CCCma), and coupled Global Environmental Multiscale, version 5.1—Nucleus for European Modeling of the Ocean (NEMO), version 3.1, model (GEMS-NEMO), developed at Recherche en Prévision Numérique Atmosphérique (RPN-A). Specifically, we assess the representation of volumetric fraction of soil moisture (VFSM), defined as the ratio of the volume of water to the total volume in the soil column. We evaluate VFSM initial conditions and examine the contribution of the persistence of initial anomalies to forecast correlation skill. We further quantify the effects of remote surface oceanic states on the VFSM forecast performance by computing the association between precedent oceanic climate indices and hindcast soil moisture while controlling for the persistence of the initial anomalies. This work provides a comprehensive evaluation of soil moisture forecast performance in CanSIPSv2.1 that complements assessments of potential predictability by Sospedra-Alfonso and Merryfield (2018), and evaluations of snow water equivalent (Sospedra-Alfonso et al. 2016a,b).

The remainder of this paper is organized as follows. Section 2 presents a brief description of CanSIPSv2.1 models and verification data. Evaluation methods are presented in section 3 and appendices A–C. Section 4 examines the soil moisture reconstructions that provide hindcasts initial conditions, and section 5 evaluates the quality of the hindcasts. Section 6 examines contributions to soil moisture forecast correlation skill from persistence of initial anomalies and its residual, including the impact of remote teleconnections on correlation skill. A summary and conclusions are provided in section 7.

2. Models and data

a. Model data and overview of CanSIPSv2.1

CanSIPSv2.1 is a version of ECCC’s seasonal forecasting system introduced in December 2021. It employs two models, CanCM4i and GEM5-NEMO, with 10 ensemble members each, to produce 12-month 20-member ensemble forecasts initialized monthly, in accordance with the view that multimodel skill is typically higher (but not always, as shown here) than that from single models (e.g., Palmer et al. 2005; Kharin et al. 2009; Kirtman and Min 2009; Becker et al. 2014). Because CanSIPSv2.1 and its model components are discussed at length by Lin et al. (2020), we provide a brief overview here emphasizing aspects of its soil moisture representation. Properties of the two models are summarized in Table 1 in the online supplemental material.

1) CANCM4I

CanCM4i employs the CanCM4 model (Merryfield et al. 2013) with improved initialization. Like CanCM4, CanCM4i couples the fourth-generation Canadian atmospheric (CanAM4) and ocean (CanOM4) models. CanAM4 (von Salzen et al. 2013) is a T63 spectral model with 35 vertical levels extending to 1 hPa and incorporates the Canadian Land Surface Scheme, version 2.7 (CLASS 2.7; Verseghy 2000). CanOM4 has a horizontal resolution of approximately 1.4° in longitude and 0.94° in latitude, and 40 vertical levels. CanAM4 and CanOM4 are coupled once per day with daily mean surface heat, freshwater and momentum fluxes transferred from the atmosphere to the ocean component, and the updated daily mean surface temperature values passed back to the atmosphere on the subsequent day. The Canadian sea ice (CanSI) model having one thickness category and cavitating fluid rheology is employed. CanCM4i is initialized from coupled assimilating runs, one for each ensemble member, which are constrained by observation-based values of atmospheric temperature, specific humidity, and horizontal winds, as well as sea surface temperature and sea ice concentration, similarly to CanCM4 (Merryfield et al. 2013). Improvements to the initialization of CanCM4i include a more realistic treatment of Arctic sea ice concentration and thickness as described by Lin et al. (2020); and use of the Ocean Reanalysis Pilot 5 (ORAP5) reanalysis (Zuo et al. 2017) to constrain subsurface ocean temperatures. Land initial conditions, including soil moisture, are obtained indirectly through the response of CLASS to the coupled-model data-constrained atmospheric fields. These data-constrained runs are referred to subsequently as the reconstruction runs for CanCM4i.

CLASS 2.7 in CanCM4i represents soil liquid and frozen water contents of three soil layers having maximum depths (from the top) of 10, 35, and 410 cm, with actual total soil depth depending on gridcell location. VFSM for the soil column is...
obtained from the total water content (liquid + frozen) in each layer by weighting the respective volumetric water fractions by the layer thicknesses.

2) GEM5-NEMO

GEM5-NEMO couples version 5.1 of the GEM model (Côté et al. 1998; Girard et al. 2014) and version 3.1 of the NEMO ocean model (http://www.nemo-ocean.eu). GEM5.1 employs a global yin–yang grid with an effective horizontal resolution of 1\degree, and 85 vertical levels extending from the surface to 0.1 hPa. GEM uses the ISBA land scheme (Noilhan and Planton 1989; Noilhan and Mahfouf 1996) and employs climatological vegetation during the model integration. NEMO3.6 is configured on the ORCA1 C grid with 50 vertical levels and a 1° horizontal resolution refined meridionally to 1/3° near the equator. GEM and NEMO are coupled once every half-an-hour without flux adjustment. Sea ice is simulated within the NEMO framework using version 4.0 of the CICE model with five ice-thickness categories.

Similar to the ECCC’s monthly prediction system (Lin et al. 2016), the atmospheric initial conditions of GEM5-NEMO hindcasts are obtained by applying random isotropic perturbations of reanalysis atmospheric fields. Only the streamfunction and the unbalanced temperatures are perturbed as in EnKF (Houtekamer et al. 2009). The perturbed fields are ultimately transformed to wind, temperature and surface pressure creating 10 different initial conditions. Sea ice is simulated within the NEMO framework using version 4.0 of the CICE model with five ice-thickness categories.

For verification, we employ ERA5 reanalysis (Hersbach et al. 2020). Li et al. (2020) report that among five reanalysis products compared with in situ measurements from 25 monitoring networks worldwide, ERA5 shows better representation of soil moisture and soil temperature variations for the 1979–2017 period. Overall, ERA5 is shown to reproduce well the soil moisture annual cycle and the timing of strong dry–wet transition, as well as the spatial structure of variability. Li et al. (2020) conclude that ERA5 is well constrained by water–energy interactions and is robust in regions of homogeneous topography, although it presents issues in regions of complex topography and landforms. In addition, evaluations against independent in situ observations by Muñoz-Sabater et al. (2021) show that the representation of soil moisture in ERA5 is not conspicuously worse than the higher-resolution specialized reanalysis ERA5-Land.

For verification, we employ ERA5 reanalysis (Hersbach et al. 2020). Li et al. (2020) report that among five reanalysis products compared with in situ measurements from 25 monitoring networks worldwide, ERA5 shows better representation of soil moisture and soil temperature variations for the 1979–2017 period. Overall, ERA5 is shown to reproduce well the soil moisture annual cycle and the timing of strong dry–wet transition, as well as the spatial structure of variability. Li et al. (2020) conclude that ERA5 is well constrained by water–energy interactions and is robust in regions of homogeneous topography, although it presents issues in regions of complex topography and landforms. In addition, evaluations against independent in situ observations by Muñoz-Sabater et al. (2021) show that the representation of soil moisture in ERA5 is not conspicuously worse than the higher-resolution specialized reanalysis ERA5-Land.

ERA5 uses the revised Tiled ECMWF Scheme for Surface Exchanges over Land (HITESSEL) model with the soil discretized in four layers of depths (from top to bottom) 7, 28, 100, and 289 cm (https://confluence.ecmwf.int/display/CKB/ERA5:+data+documentation, last accessed 20 January 2023). VFSM is provided for each layer and for each CanSIPSv2.1 model we combine these values weighted by the layer thicknesses to obtain VFSM for the whole soil column.
3. Evaluation methods

Hindcasts $Y$, reconstruction runs $Z$ giving the initial conditions, and verification data $X$ of monthly VFSM are first converted to anomalies relative to their 1980–2018 climatology, and regrided to a common 1° horizontal resolution using bilinear interpolation. For the hindcasts, climatology depends on lead time. Monthly anomaly hindcasts are subsequently denoted as $Y_{kjml}$ where $k = 1, \ldots, K$ indicates ensemble member; $j = 1, \ldots, J$ indicates initialization year; $m = 1, \ldots, M$ denotes initialization month; and $l = 1, \ldots, L$ specifies forecast or target month. For brevity, we omit the dependence on grid location throughout. Reconstruction run anomalies are denoted as $Z_{kjml}$ referring to the values for the month immediately preceding the first month $l = 1$ of the hindcast. For each CanSIPSv2.1 model, we employ $K = 10$ ensemble members, $J = 39$ initialization years over the period 1980–2018, $M = 12$ initialization months and $L = 12$ months encompassing the CanSIPSv2.1 forecast range. Similarly, we denote the verifying anomalies as $X_{jml}$ with the subindex $k$ absent to indicate a single realization and $j$ spanning the same number of initialization years as the hindcasts. Observations contemporary with $Z_{kjml}$ are denoted as $X_{jml}$. Seasonal values for hindcasts and observations are computed as rolling three-month averages, with forecast seasons $l = 1, L - 2$. For the reconstruction runs

Fig. 2. (a),(b) Climatological bias; (c),(d) correlation; and (e),(f) RMSE of JJA soil moisture relative to ERA5 for (left) CanCM4i and (right) GEM5-NEMO reconstruction runs during 1980–2018. Gray regions over land correspond to VFSM with near-zero interannual variability. Cross hatching indicates statistical significance at the 95% confidence level.
and their contemporary observations, the seasonal averages are taken with respect to $m$. We use the same notation for seasonal and monthly anomalies. Also, we refer to forecast months month 1, month 2, and so forth to indicate the first, second, and subsequent months in the forecasts, and similarly, months 1–3, months 2–4, and so forth for the first, second and subsequent seasons. We use month 0 to indicate the month preceding the first hindcast month. Seasons are referred to according to the standard boreal seasons, that is, December–February (DJF) for winter, March–May (MAM) for spring, June–August (JJA) for summer, and September–November (SON) for fall. Ensemble mean anomalies for hindcasts and reconstruction runs are denoted by dropping the subindex $k$, that is, $Y_{ml}$ and $Z_{ml}$, respectively, whereas climatological averages of the ensemble mean are denoted as $Y_{ml}$ and $Z_{ml}$ (and equal 0 for the anomalies). For simplicity, we subsequently omit the subindex $m$. Explicit formulas are given in appendix A.

a. Measures of performance

Several evaluation metrics are employed to assess soil moisture representation in CanSIPSv2.1, including model bias given by the difference of the full field ensemble mean and verifying climatologies,

$$\text{bias} = Y_I - X_I,$$  \hspace{1cm}  (1)

the anomaly correlation coefficient of the ensemble mean hindcast,

$$r_{Y_I Y_I} = \frac{\bar{X}Y_I}{\sigma_X \sigma_Y},$$  \hspace{1cm}  (2)

where the overbar denotes covariance and $\sigma$ denotes standard deviation, and the unbiased root-mean-square error,
\[
\text{RMSE} = \sqrt{\frac{J-1}{J} \left( \sigma_x^2 + \sigma_y^2 - 2X_l Y_l \right)}.
\]

In general, \(-\infty < \text{bias} < \infty, -1 \leq r_{X,Y} \leq 1\), and \(0 \leq \text{RMSE} < \infty\), with best performance achieved when bias = 0, \(r_{X,Y} = 1\), and \(\text{RMSE} = 0\). Analogous measures are used to assess the ensemble mean of the reconstruction runs \(Z_0\) (or the one member for GEM5-NEMO) relative to \(X_0\). We also assess the potentially predictable variance fraction (Boer et al. 2013, 2019a,b)

\[
\text{ppvf} = \frac{\sigma^2_{Y_0} + \sigma^2_{Y_0} - 2X_l Y_l}{\text{RMSE}^2},
\]

where \(\sigma^2_{Y_0} = \sigma^2_{Y_0} + \sigma^2_{Y_0}\) is the ensemble hindcast total variance, equal to the ensemble mean of the individual ensemble members variances. Here, the potentially predictable variance \(\sigma^2_{Y_0}\) measures the amplitude of variations that are common to the hindcast ensemble and the noise variance \(\sigma^2_{Y_0}\) measures the spread of the ensemble around the ensemble mean and represents the amplitude of unpredictable variations. It follows that \(0 \leq \text{ppvf} \leq 1\) with the two extreme cases indicating no predictability (i.e., \(\sigma^2_{Y_0} = 0\)) and maximum potential predictability (i.e., \(\sigma^2_{Y_0} = \sigma^2_{Y_0}\), or equivalently \(\sigma^2_{Y_0} = 0\)), respectively. For an ensemble of size \(K\), the noise variance and mean-square error (MSE) define the hindcast spread-over-error ratio (SOE) according to (Ho et al. 2013).

\[
\text{SOE} = \frac{\sqrt{K + 1 \sigma^2_{Y_0}}}{\frac{1}{K} \cdot \text{MSE}},
\]

which measures the reliability of the hindcast ensemble, where \(\text{SOE} = 1\) indicates that the ensemble members and observations are statistically indistinguishable, whereas \(\text{SOE} < 1\) or \(\text{SOE} > 1\) indicates that the hindcast is overconfident or underconfident, respectively. For simplicity, the forecast month or season dependence for \(\sigma^2_{Y_0}, \sigma^2_{Y_0}, \sigma^2_{Y_0}\), bias, RMSE, ppvf, and SOE is not shown explicitly. Explicit formulas for the computation of Eqs. (1)-(5) are provided in appendix A.

b. Contributions to correlation skill from persistence of initial anomalies and its residual

An important feature of soil moisture is its tendency to persist in time (e.g., Sospedra-Alfonso and Merryfield 2018). Hindcast persistence is measured here as the lag-\(l\) correlation relative to the reconstruction runs giving the initial conditions,

\[
r_{Y_0,Z_0} = \frac{Y_{0,Y_0}}{\sigma_{Y_0} \sigma_{Z_0}},
\]

and similarly for the observations, denoted as \(r_{X_0,X_0}\). As shown in appendix B, the variance \(\sigma^2_{Y_0}\) of the ensemble mean hindcast can be decomposed into contributions \(r_{X_0,Z_0} \sigma_{Y_0}^2\) from the persistence of the initial anomalies and a residual \(\sigma^2_{Y_0}\),

\[
\sigma^2_{Y_0} = r_{X_0,Z_0}^2 \sigma_{Y_0}^2 + \sigma^2_{Y_0},
\]

whereas the correlation skill [Eq. (2)] of the ensemble mean hindcast can be similarly decomposed into contributions from persistence \(P_P\) of the initial anomalies and a residual \(R_P\),

\[
r_{X_0,Y_0} = r_{X_0,Z_0} r_{Y_0,Z_0} + \frac{\sigma_{X_0} \sigma_{Y_0}}{\sigma_{X_0} \sigma_{Y_0}} r_{X_0,Y_0} = P_P + R_P.
\]
errors in the phase of the initial conditions (appendix C). To the extent that model initialization does not affect the forecast persistence, this indicates, and quantifies, the benefits of initialization for highly persistent variables.

On the other hand, \( R_{ll} \) quantifies the contributions to correlation skill not linked to the persistence of the initial anomalies, and is determined by the correlation of the residuals \( r_{X_l' Y_l'} \) (i.e., the partial correlation of \( Y_l \) and \( X_l \) after controlling for the variations of the initial conditions \( Z_0 \)), weighted by the contributions \( \sigma_{y_l'}/\sigma_y \) and \( \sigma_{x_l'}/\sigma_x \) of the residuals standard deviations. Realistic soil moisture residuals can result from the ability of hindcasts to predict future climate-driven variations, such as those forced by precipitation and/or evapotranspiration linked to large-scale atmospheric variability (Nicolai-Shaw et al. 2016) and sea surface temperature (Sospedra-Alfonso and Merryfield 2018).

c. Statistical significance

Statistical significance for correlations and partial correlations (i.e., correlation of residuals after controlling for a specified variable, e.g., \( r_{X_l' Y_l'} \)) is computed at the \( p = 0.05 \) significance level based on the two-tailed \( t \)-test statistics. Specifically, we reject the null hypothesis of zero (partial) correlation \( r \) if the absolute value of

\[
 t = r \sqrt{\left(\frac{n_{eff} - 2}{1 - r^2}\right)}
\]

is larger than an established critical value \( t_{n_{eff} - 2, p/2} \) that depends on the effective degrees of freedom \( n_{eff} \). For correlations, \( n_{eff} = J \) is the number of initialization years, whereas for partial correlations \( n_{eff} = J - 1 \) to account for the controlled variable.

For \( P_{ll} \) and \( R_{ll} \), statistical significance is evaluated using a nonparametric bootstrap approach (Goddard et al. 2013) to generate sampling distributions based on 10000 repetitions. For every grid cell, \( P_{ll} \) and \( R_{ll} \) are generated by resampling the data, with replacement, along the time and (when possible) ensemble dimensions of hindcasts, reconstruction runs and observations. The 2.5%–97.5% quantile estimates of the bootstrapping distributions determine \( P_{ll} \) and \( R_{ll} \) 95% confidence intervals. If the confidence interval does not include zero, they are deemed statistically significant with 95% confidence and the associated grid cell is cross hatched in the maps.

Fig. 5. Representation of seasonal soil moisture in CanSIPSv2.1 hindcasts during 1980–2018, showing globally averaged (left) correlation skill and (right) RMSE for (a),(b) multimodel CanSIPSv2.1; (c),(d) CanCM4i; and (e),(f) GEM5-NEMO ensemble mean hindcasts as a function of forecast month.
4. Evaluation of the reconstruction runs

Reconstruction runs that provide soil moisture initial conditions in CanSIPSv2.1 are obtained by the response of CLASS to the observational-constrained atmospheric model fields for CanCM4i and from the response of SPS to reanalysis fields for GEM5-NEMO (section 2). Figure 1 shows evaluations of the ensemble mean VFSM reconstructions runs for the full soil column against ERA5 reanalysis. Both CanCM4i and GEM5-NEMO exhibit a negative bias for all seasons indicating a deficit in fractional soil moisture in the hindcast initial conditions (Fig. 1a). These biases are seasonally dependent for GEM5-NEMO, with the strongest biases occurring in winter and spring. Biases in CanCM4i during summer and winter tend to be stronger in the northern latitudes, in higher elevations including the Himalayas and Karakoram Range, in sub-tropical regions and the Amazon basin, and in eastern North America extending into the Hudson Bay lowlands of Canada (Figs. 2a and 3a). Lack of bias seasonality for CanCM4i is evident. Strong biases in GEM5-NEMO are seen in most of these regions and extend broadly over most of the Northern Hemisphere during winter (Figs. 2b and 3b). For the top soil layer, both models present a reduced negative bias on average, with GEM5-NEMO exhibiting less seasonality and biases than CanCM4i (Fig. 1a in the online supplemental material).

Globally, GEM5-NEMO VFSM interannual variance is largely in agreement with ERA5 for most seasons, with slightly larger values during spring. CanCM4i variance exceeds that of ERA5, most notably during winter when it is almost twice as strong (Fig. 1b). Larger VFSM variance in CanCM4i occurs mainly over mountain ranges and in northern latitudes (not shown), and also in the Amazon basin, possibly linked to larger variance in potential evapotranspiration than observed (Fig. 3 in the online supplemental material; Sospedra-Alfonso and Merryfield 2018). For the top soil layer, modeled VFSM variance is larger than observed during winter, and about the same as observed during summer (Fig. 1b in the online supplemental material).

The representation of ERA5 interannual variability is generally better in GEM5-NEMO than CanCM4i (Fig. 1c), which shows global mean correlations above 0.5 for all seasons and up to 0.7 during summer. GEM5-NEMO correlations present a marked seasonality with highest values in summer and lowest in winter, which is somewhat opposite to the behavior of CanCM4i. The relatively high correlation for GEM5-NEMO during summer extends over most of the globe, excluding the main deserts and Arctic regions, with larger values in the tropics (Fig. 2d). By contrast, correlation skill in the mid- to high latitudes is much reduced during winter, when it can be negative in scattered regions, most notably in eastern Europe (Fig. 3d).

CanCM4i, on the other hand, has global mean correlation skill between 0.3 and 0.4 for all seasons (Fig. 1c), with highest values during winter. Poor performance is seen in western South America during summer and winter (Figs. 2c and 3c), particularly in the Amazon basin where reconstruction runs are unrealistically dry and interannual variations of precipitation are not well represented (Fig. 3, along with Fig. 4 in the online supplemental material).
supplemental material; Sospedra-Alfonso and Merryfield 2018). Both models present higher correlation skill for the top soil layer, with globally averaged values between 0.6 and just over 0.8 for GEM5-NEMO, and between 0.4 and 0.6 for CanCM4i (Fig. 1c in the online supplemental material).

Globally averaged RMSE has similar seasonality in CanCM4i and GEM-NEMO, with the lowest values occurring in summer (Fig. 1d). RMSE is generally lower for GEM5-NEMO, with highest values occurring in the midlatitudes during winter, most notably across the eastern United States and eastern Europe (Fig. 3f). RMSE geographic patterns for CanCM4i are largely similar in summer and winter (Figs. 2e and 3e), with the largest values occurring in Asia, particularly over the Himalayas and Siberia. For the top soil layer, GEM5-NEMO has smaller globally averaged RMSE and similar seasonality than CanCM4i, with lowest values in summer (Fig. 1d in the online supplemental material).

The persistence of ERA5 monthly soil moisture anomalies is shown in Fig. 4a for lags of 1 to 11 months. At 1-month lag, globally averaged autocorrelation ranges from above 0.8 to 0.9 with the larger values occurring for initial months in fall and winter decaying to values between 0.3 and 0.4 at 11-month lag. A relatively faster decay occurs for initial months in winter compared to those in summer, with a very little change or even a slight increase in autocorrelation for the initial summer months at lags longer than about 9 months. This high autocorrelation indicates strong anomaly persistence in ERA5 and implies comparable correlation skill for simple forecasts based on persisting the initial anomaly. For CanCM4i (Fig. 4b), globally averaged VFSM autocorrelation at 1-month lag is relatively high (between 0.6 and just above 0.8), although lower than for ERA5 and decaying faster for up to about 6-month lag, particularly for initial anomalies during spring and early summer. For GEM5-NEMO (Fig. 4c), the values are also lower than for ERA5 and slightly reduced relative to CanCM4i except for initial anomalies in the summer months. Autocorrelation in GEM5-NEMO decays faster than in observations for up to about 6-month lag, comparably to CanCM4i. For longer lags, both models show a slow decay or even a slight increase in autocorrelation, particularly for initial months in summer.

5. Evaluation of the hindcasts

Globally averaged correlation skill for seasonal VFSM in CanSIPSv2.1 single- and two-model hindcast ensemble means is shown in Fig. 5 for months 1–3 to months 6–8. Analogous results for the top soil layer are shown in Fig. 2 in the online supplemental material. The two-model values for the full soil column (Fig. 5a) are on average about the same or slightly less than those of GEM5-NEMO (Fig. 5e), and significantly larger than those for CanCM4i (Fig. 5c). Their seasonality mirrors that of GEM5-NEMO and is similar to the reconstruction runs (Fig. 1c), with values peaking typically during summer for months 1–3 to months 3–5. For months 1–3, globally averaged correlation skill ranges from over 0.4 during winter to just about 0.5 during summer, and decays with forecast month to below 0.25 at months 6–8.

Because the reconstruction runs were evaluated in section 4 and soil moisture anomalies are relatively persistent (discussed
in detail in section 6 below), we examine here CanSIPSv2.1 forecast performance for target seasons 1 month after initialization. Indeed, for months 2–4, globally averaged correlation skill is between 0.35 and 0.4 (Fig. 5). The hindcast performance is better over tropical land (Fig. 6) extending toward southeastern United States and Asia during winter-spring (Figs. 6a,b) and across South America, Africa, and eastern Australia during summer-fall (Figs. 6c,d). A notable exception is the western Amazon basin, where correlation skill can be negative in all seasons owing partly to the poor representation of soil moisture in CanCM4i initial conditions (Fig. 3c) and possibly to the dry and warm biases in freely running CanCM4 (Merryfield et al. 2013) impacting the forecast once the observational constraints are removed. Correlation skill is lower in higher latitudes, not unexpectedly given the relatively low correlation of GEM5-NEMO initial conditions during winter (Fig. 3d), although pockets of high correlation are seen in Eurasia and western North America depending on season.

Globally averaged RMSE in CanSIPSv2.1 ensemble mean hindcasts is larger in winter than in summer (Fig. 5b), much as for CanCM4i initial anomalies (Fig. 1d) and hindcasts (Fig. 5d), and increases with forecast month particularly in summer. Values are similar or larger than for GEM5-NEMO, most notably in winter and for shorter lead times. For CanCM4i, RMSE decreases with forecast month for winter as the target season. This unexpected behavior is linked to the deficiencies in the representation of soil moisture in CanCM4i reconstruction runs, particularly during winter and in regions such as the Himalayas, Siberia, and the Far East (Fig. 3e).

For months 2–4, RMSE tends to be larger in the mid- to high latitudes (Fig. 7), notably in Europe and eastern Asia during winter and spring, and mostly coincident with regions of relatively low correlation skill (Fig. 6). In the lower latitudes, areas with relatively large RMSE include central Africa, eastern South America, and northern Australia. Larger RMSE is seen in central United States extending northward to the Canadian prairie provinces, a region known for a dry and warm model bias during JJA (Merryfield et al. 2011). Actually, CanESM2, a previous version of the Canadian Earth System Models based on CanCM4 and contributing to CMIP5 exhibits one of the largest JJA temperature bias and negative precipitation biases in the region among the CMIP5 models (Fig. 2, Lin et al. 2017). Although climatological biases are removed for the evaluations shown, the resulting precipitation deficit combined with soil moisture-atmosphere feedback (Lin et al. 2017) is likely to impact the amplitude of the
VFSM forecast anomalies and its RMSE. In this region, the total variance of VFSM hindcasts is much lower than observed (not shown), which affects RMSE. For the top soil layer (Fig. 2 in the online supplemental material), which is more exposed to atmospheric variations and has higher interannual variability, CanSIPSv2 exhibits generally larger RMSE than for the total soil column.

For months 1–3 in CanCM4i, the potentially predictable VFSM variance ranges, on average, from about 50% of the total (i.e., predictable + noise) during summer and fall, to about 70% in winter (Fig. 8a). This variance fraction decays to below 20% for months 6–8, except during spring when it remains above 30% of the total. For GEM5-NEMO, the potentially predictable variance fraction is typically stronger in summer and winter (Fig. 8b), with predictable variance ranging from over 50% to almost 70% of the total variance for months 1–3, to about 20% or below for months 6–8. Interestingly, correlation skill in GEM5-NEMO (Fig. 5e) is generally higher during summer, implying that the model capitalizes on its predictable signal, whereas correlation skill is lower during winter, suggesting that some of the potentially predictable variance is the result of model biases and does not improve the forecast performance. This is consistent with issues noted for the GEM5-NEMO initial conditions in the mid- to high latitudes during winter (Figs. 3b,d). For the total variance, GEM5-NEMO is on average in good agreement with observations, with a slight overestimation during spring (Fig. 8d). CanCM4i on the other hand, exhibits a total variance that is on average 1.5–2 times as large as observed (Fig. 8c). As for the initial conditions (section 4), regions of relatively large VFSM variance in CanCM4i include the northern latitudes, alpine regions, and the Amazon basin, whereas larger than

FIG. 9. Variance and correlation skill decomposition of monthly soil moisture in CanSIPSv2.1 hindcasts as a function of forecast month in terms of contributions from persistence of the initial conditions and its residual, showing globally averaged (a),(b) contribution $r^2_{YlZ}$ in percent of persistence of the initial anomalies to ensemble mean hindcast variance; (c),(d) correlation skill decomposition $r_{XY} = P_R + R_{Yl}$; and (e),(f) contributions in percent of persistence of the initial anomalies to correlation skill for monthly soil moisture in (left) CanCM4i and (right) GEM5-NEMO. Evaluation metrics are computed as described in section 3b. Colors indicate the month for the reconstruction runs relative to which the various measures are computed; e.g., iJan for month-1 VFSM hindcasts refers to month-0 January soil moisture from the reconstruction runs and month-1 February soil moisture from the ensemble mean hindcasts. Solid and hatched bars in (c) and (d) respectively indicate contribution $P_R$ from persistence of the initial anomalies and its residual $R_{Yl}$ to correlation skill. Percent contribution of the residuals in (a), (b), (e), and (f) is the difference between 100 and the values indicated by the solid bars.
observed variance in GEM5-NEMO are notable in the Pacific Northwest and Alaska, the Great Lakes region and eastern Canada, eastern Europe and Siberia, and the Sahel, depending on the season (not shown).

Figure 8e shows globally averaged spread-over-error ratio for CanCM4i VFSM ensemble hindcasts for several forecast months. Underdispersion (SOE < 1) occurs for most seasons up to about months 5–7 when it reaches 1 during winter, indicating initially overconfident forecasts. For GEM5-NEMO (Fig. 8f), SOE is above 0.8 from months 2–4 and onward, except during winter, and slightly surpasses 1 during spring indicating a generally reliable forecast on average. For months 2–4, regions of pronounced underdispersion (SOE < 0.5) includes, for CanCM4i (Figs. 3a,c in the online supplemental material), tropical and subtropical regions of North America and Central Asia during summer, and most of the mid- to high latitudes during winter and spring, and for GEM5-NEMO (Figs. 4b,d in the online supplemental material), most of the mid- to high latitudes during winter, and subtropical areas during summer. By contrast, regions of pronounced overdispersion (SOE > 2) in CanCM4i (Figs. 3a,c in the online supplemental material) includes eastern Asia, Arctic regions, the Himalayas and Karakoram, and to some extent the Sahel during summer and fall, and for GEM5-NEMO (Figs. 4b,d in the online supplemental material), eastern United States and Europe, and more broadly in temperate zones during spring and summer.

6. Contributions from persistence and its residual to VFSM forecast performance

As discussed in section 4, CanSIPSv2.1 VFSM reconstruction runs and verifying data exhibit moderate to strong anomaly persistence as measured by autocorrelation. Figures 9a and 9b show that, on average, the contribution $r_{Y,0}^2$ of the initial anomalies to the variance of the ensemble mean hindcast [Eq. (7)] is typically over 50% and can reach up to 70% in month 1, decreasing with forecast month similar to the autocorrelation of the reconstruction runs (Fig. 4), thus indicating persistence of the initial anomalies with potential for contributions to correlation skill.

![Figure 10](image-url)
a. Contributions to VFSM forecast performance from persistence of the initial anomalies

The contribution $P_{ll}$ from persistence of the initial VFSM anomalies to forecast correlation skill can be measured from the decomposition in Eq. (8). Globally averaged values of this decomposition for CanCM4i and GEM5-NEMO are shown in Figs. 9c and 9d. For month 1, correlation skill in CanCM4i (Fig. 9c) ranges from 0.3 to 0.35 for all seasons, with averaged positive contributions from persistence accounting for over 60% to about 70% of the correlation skill (Fig. 9e). Larger $P_{ll}$ occurs during winter, and lower values in summer and fall, consistent with the contribution from persistence to the ensemble mean hindcast variance (Fig. 9a). Similarly, contributions to correlation skill in GEM5-NEMO (Fig. 9d) tend to mirror the seasonality of the contributions the hindcast ensemble mean variance (Fig. 9a). Similarly, contributions to correlation skill in GEM5-NEMO (Fig. 9d) tend to mirror the seasonality of the contributions the hindcast ensemble mean variance (Fig. 9a). Highly persistent forecasts of highly persistent climate variables should greatly benefit from initialization, since initially realistic forecast anomalies persisting in time should remain correlated with the observed anomalies for some period within the forecast decorrelation time. Figures 10 and 11 show for CanCM4i and GEM5-NEMO, respectively, that in locations where observed and hindcast VFSM autocorrelation is sufficiently high, the correlation skill of the reconstruction runs $r_{X_{0},Y_{0}}$ is a very good indicator of the correlation skill of the hindcast $r_{X_{1:7},Y_{1:7}}$ for up to at least month 7 in the forecast, depending on target month. The figures show that both models exhibit a strong linear relationship between $r_{X_{0},Z_{0}}$ and $r_{X_{1:7},Y_{1:7}}$ along the 1:1 correspondence, particularly for observed and modeled autocorrelation above 0.85. A slight deviation from this correspondence is seen for GEM5-NEMO at longer leads depending on the season due to performance degradation not linked to persistence. These results show that for regions and seasons of relatively high persistence, forecast correlation skill is largely determined by the accuracy of VFSM initial conditions.

Regions of high VFSM persistence in CanCM4i, matching that in the observations, encompass alpine and snow-covered areas, and equatorial West Africa, as seen for instance for the month 3 hindcast in March (Fig. 5 in the online supplemental material). For GEM5-NEMO, regions of high VFSM persistence...
matching the observations include India, the Sahel, and Siberia for March hindcasts, and semiarid regions in South America, South Africa, Central Asia, and Australia for September hindcasts, for up to at least month 3 (Fig. 6 in the online supplemental material). Except for Siberia, where frozen soil moisture and therefore persistence is expected, these are regions and seasons of dry conditions, which favor stronger soil moisture persistence and thus predictability [e.g., Nicolai-Shaw et al. (2016), and the references therein]. In the case of India and the Sahel, winter is the dry season, whereas summer brings rainfall that alters soil moisture persistence and its predictability. Forecast correlation skill in these regions, which is found to be significant for up to at least month 3, is largely determined by the accuracy of the initial conditions (Figs. 10 and 11) and is expected to improve significantly with improved initialization provided persistence is not affected by the initialization process.

b. Contributions to VFSM forecast performance not linked to persistence of the initial anomalies

The contribution $R_l$ to correlation skill $r_{X,Y}$ that is not linked to the persistence of the initial anomalies is expressed as the product of three quantities [Eq. (8)]; the correlation $r_{X,Y}$ of the residuals from the ensemble mean hindcasts and observations...
regressed against the reconstruction runs (i.e., the partial correlation of the ensemble mean hindcasts and observations after controlling for the reconstruction runs, which measures forecast performance that is not attributed to variability of the initial conditions), times the contributions $s_{X}'/s_X$ and $s_{Y}'/s_Y$ of residual standard deviations from observations and ensemble mean hindcasts, respectively. In contrast with the contribution $P_{ll}$ from persistence (Figs. 9c,d), globally averaged $R_{ll}$ for both CanCM4i and GEM5-NEMO accounts for about 30%–40% of the correlation skill at month 1 to about 60%–70% for month 6 (Figs. 9e,f). That is, the relative contribution $R_{ll}$ increases although the overall correlation skill $r_{XlYl}$ decreases with lead time. The seasonality of $R_{ll}$ is, on the other hand, different for both models and opposite to that of $P_{ll}$ by construction (section 6a).

To gain insight into the contributions to correlation skill not linked to persistence, Figs. 12 and 13 show maps of correlation skill $r_{XlYl}$, the correlation of residuals $r_{XlYl}'$, and the contribution $R_{ll}$ to correlation skill from these residuals, for June and December month 2 VFSM hindcasts, respectively. For June hindcasts (Fig. 12), CanCM4i correlation skill $r_{XlYl}$ is significant mostly in tropical landmasses, subtropical North America and temperate zones of Eurasia (Fig. 12a), whereas GEM5-NEMO shows significant correlation skill across the global land (Fig. 12b). Residual correlations $r_{XlYl}'$ in both CanCM4i and GEM5-NEMO are mostly confined to subregions of the above (Figs. 12c,d), implying that skillful residual variations contribute to the full correlation skill (Figs. 12c,d). A notable exception is the Amazon basin in CanCM4i (Figs. 12a,c,e), where correlation skill is small or negative in the western sector (Fig. 12a) even though residual correlation skill is high (Fig. 12c) and residual variations do contribute significantly to correlation skill (Fig. 12c). This suggests that the hindcasts are somewhat...
realistic over the Amazon but evolve from unrealistic and persistent initial states (Fig. 2c, along with Fig. 7a in the online supplemental material) leading to poor forecast performance, which illustrates the benefits of the correlation skill decomposition [Eq. (8)] as an assessment tool.

For December hindcasts (Fig. 13), CanCM4i exhibits higher correlation skill than for June hindcasts (Fig. 12) over the mid- to high latitudes, and significant skill also in the tropics (Fig. 13a). Residual correlation skill is generally reduced compared to the full correlations (Fig. 13c), and is significant in regions where residuals contribute to forecast skill (Fig. 13e). Contributions to correlation skill from persistence of the initial anomalies are, however, stronger and widespread (Fig. 7c in the online supplemental material), as noted in section 6a (Fig. 9c). For GEM5-NEMO, significant correlation skill encompasses most global land with notable exceptions in temperate zones of North America and Eurasia (Fig. 13b). Although reduced, the distribution of residual correlation skill (Fig. 13d) mirrors that of the full correlations (Fig. 13b), and residual variations largely contribute to correlation skill (Fig. 9). An exception is eastern Australia, where correlation is high, residual correlation is low, and most of the correlation skill is attributed to the persistence of the initial anomalies (Fig. 7d in the online supplemental material), which are well represented in the reconstruction runs (Fig. 3d).

c. Remote ocean influence on VFSM forecasts

Hindcast variations in CanSIPSv2.1 not linked to the persistence of VFSM initial anomalies were found to be skillful at different seasons and forecast months (Figs. 9c-f, 12c-f, and 13c-f), but sources of this residual variability remain to be determined. Climate modes such as ENSO have been shown to drive soil moisture variations regionally through their remote influence on precipitation and evapotranspiration (e.g., Nicolai-Shaw et al. 2016; Sospedra-Alfonso and Merryfield 2018), and are thus expected to contribute to VFSM forecast skill. Sospedra-Alfonso and Merryfield (2018) showed that for CanCM4 hindcasts, ENSO affects soil moisture in teleconnected regions several months into the forecasts, indicating the potential for long-range prediction.

To establish whether there is remote Ocean influence on CanSIPSv2.1 VFSM hindcasts, Fig. 14 shows the percentage of global land area with significant (p < 0.05) 1-month-lagged correlation between residual VFSM and residual climate index forecasts (hatched bars), with residuals taken relative to month-0 VFSM reconstruction runs, and percentage from this area with VFSM forecast skill (solid bars), for the (a),(b) Niño-3.4 index and the (c),(d) tripole index for the IPO in (left) CanCM4i and (right) GEM5-NEMO. Lagged correlation of residuals is computed as the correlation between the ensemble mean of month \( l = 2, \ldots, 6 \) VFSM and month \( l - 1 \) climate index forecasts regressed against the ensemble mean of the month-0 VFSM reconstruction runs. Correlation skill \( r_{X,Y} \) of residuals is computed as the correlation between the residuals of, respectively, month \( l = 2, \ldots, 6 \) ensemble mean hindcasts and observed VFSM regressed against month-0 VFSM reconstruction runs. Colors represent target month of VFSM hindcasts.

![Fig. 14](image-url)

**FIG. 14.** Percentage of global land area with significant (p < 0.05) 1-month-lagged correlation between residual VFSM and residual climate index forecasts (hatched bars), with residuals taken relative to month-0 VFSM reconstruction runs, and percentage from this area with VFSM forecast skill (solid bars), for the (a),(b) Niño-3.4 index and the (c),(d) tripole index for the IPO in (left) CanCM4i and (right) GEM5-NEMO. Lagged correlation of residuals is computed as the correlation between the ensemble mean of month \( l = 2, \ldots, 6 \) VFSM and month \( l - 1 \) climate index forecasts regressed against the ensemble mean of the month-0 VFSM reconstruction runs. Correlation skill \( r_{X,Y} \) of residuals is computed as the correlation between the residuals of, respectively, month \( l = 2, \ldots, 6 \) ensemble mean hindcasts and observed VFSM regressed against month-0 VFSM reconstruction runs. Colors represent target month of VFSM hindcasts.
For the Niño-3.4 index (Figs. 14a,b), the impact of SST on VFSM depends on target month and is statistically significant over 20%–35% of the global land area for month 2 forecasts, and can reach about 25%–55% in CanCM4i and 25%–45% in GEM5-NEMO for month 6. Seasonality is similar for both models, with higher lagged partial correlations (i.e., strongest SST influence) in winter–spring and lowest in summer. The impact of ENSO increases with lead time for all target months and is generally stronger for CanCM4i than GEM5-NEMO. However, the percentage of global land with significant partial correlation skill (shown as solid bars in the figure) is about the same for all lead times, and it is generally higher for GEM5-NEMO than CanCM4i forecasts, suggesting that GEM5-NEMO better capitalizes on ENSO-driven VFSM variability to attain correlation skill. For the tripole IPO index (Figs. 14c,d), the seasonality and behavior with lead time are similar to those for the Niño-3.4 index (Figs. 14a,b), except that the area of SST-influence on VFSM is larger, particularly during summer. Unlike the Niño-3.4 index, the tripole IPO index depends on SST in the extratropical Pacific (Fig. 1, Henley et al. 2015), which contributes to VFSM variability and possibly enhanced VFSM partial correlation skill, more notably for GEM5-NEMO hindcasts.

The maps of the 1-month lagged partial correlation between month 1 Niño-3.4 index and month 2 VFSM ensemble mean hindcasts is shown for target months June (Figs. 15a,c) and December (Figs. 15b,d). The maps indicate the regions of significant SST teleconnections, which are largely consistent for both models, although differences do exist. The CanCM4i maps, in particular, are consistent with known ENSO-teleconnected precipitation anomalies (Fig. 11, Sospedra-Alfonso and Merryfield 2018). A comparison of Fig. 15 and Figs. 12 and 13 reveals the regions where significant partial correlation skill can be attributed to tropical Pacific SST-driven VFSM variability. These include eastern South America and central Africa during December, most notably for GEM5-NEMO, and southeastern Asia, the Maritime Continent, and sectors of South America during June. Analogous plots for the teleconnected regions of the tripole IPO index are shown in Fig. 16, which presents a more encompassing influence than the Niño-3.4 index, particularly in the higher latitudes and most notably for June, as seen for instance in eastern Europe and Asia Minor, possibly contributing to the collocated VFSM hindcast correlation skill (Figs. 12 and 13). Overall, these results strongly suggest that climate variability linked to the Pacific Ocean contributes to regional forecast skill of soil moisture in CanSIPSv2.1 hindcasts.

7. Summary and conclusions

CanSIPSv2.1 provides ECCC’s operational seasonal forecasts up to 12-month lead over Canada (https://climate-scenarios.canada.ca/?page=cansips-prob; last accessed 20 January 2023).
and globally (https://climate-scenarios.canada.ca/?page=cansips-global; last accessed 20 January 2023). These include seasonal predictions of hydroclimatic variables such as precipitation and snow water equivalent. Soil moisture, however, is currently not included in the suite of available forecasts despite being a key variable of the climate system. This work provides a framework for comprehensive evaluation of soil moisture in seasonal hindcasts that will inform decision making on including soil moisture in ECCC’s operational forecasts and can similarly inform evaluations and decisions by other forecasting centers. Our approach was to assess the quality of the forecasts and quantify the contributions to prediction skill from sources beyond persistence. Consistent with findings using other prediction systems (e.g., Kanamitsu et al. 2003; Yao and Yuan 2018; Esit et al. 2021; Chikamoto et al. 2015), we showed that both anomaly persistence and forcing from ocean-driven meteorology are key to the performance of soil moisture forecasts in CanSIPSv2.1. Because soil moisture displays strong anomaly persistence (Fig. 4), VFSM forecasts can greatly benefit from a realistic representation of its initial conditions. For each CanSIPSv2.1 model, we evaluated the reconstruction runs that provide VFSM initial conditions in terms of their climatological bias and interannual variability. The accuracy of these runs is seasonally dependent, with stronger seasonality seen for GEM5-NEMO. On average, CanCM4i was found to have a smaller bias, whereas GEM5-NEMO exhibits a more realistic variability, typically with higher correlation and lower RMSE, particularly in summer (Fig. 1). Higher correlations are generally seen in the tropics, and for GEM5-NEMO, relatively low or negative values are seen in the mid- to high latitudes during winter (Figs. 1 and 3). In CanCM4i, the western portion of the Amazon basin stands out for the poor representation of soil moisture (Figs. 2 and 3), consistent with the dry conditions reported by Sospedra-Alfonso and Merryfield (2018) and the strong temperature and precipitation biases discussed by Merryfield et al. (2013).

Several measures were used to assess the performance of CanSIPSv2.1 VFSM hindcasts, both in terms of its potential predictability and in comparison with ERA5 reanalysis (Figs. 5–8). A novel approach was implemented to decompose the forecast correlation skill in terms of contributions from persistence of the initial anomalies and those not linked to persistence [Eq. (8)]. The contribution from persistence is determined by the accuracy of the initial conditions and the autocorrelation in forecasts and verifying observations [Eqs. (8) and (C3)]. For about four months into the forecasts, globally averaged correlation skill is largely attributable to persistence (>50%) depending on target season, with contributions not linked to persistence becoming relatively more important with lead time (Fig. 9). Contribution from persistence varies between the two models, with stronger impacts for CanCM4i forecasts initialized in winter and GEM5-NEMO forecasts initialized in summer–fall. Regions of strong anomaly persistence at least three months, and up to seven months (Figs. 10 and 11), into the forecast are characterized by dryer conditions that favor predictability. These include India and the Sahel during late winter, and semiarid regions of South America, South Africa, Central Asia, and Australia during late summer (Figs. 5 and 6 in the online supplemental material).
For short lead times (e.g., months 1–3 and months 2–4), the performance of VFSM ensemble mean hindcasts surpasses that of the single models during winter-spring and is comparable to that of the best performing model during summer-fall (Fig. 5). The ensemble mean hindcasts perform better during summer-fall than in winter-spring largely because GEM5-NEMO has more accurate initial anomalies during late spring to early fall (Figs. 2 and 3) and it has a stronger persistence of the initial anomalies during summer-fall (Fig. 9b). CanCM4i correlation skill is much lower during summer-fall than in winter-spring (Figs. 5c and 9a,c,e), consistent with the seasonality of its initial anomaly persistence (Fig. 4).

Although not generally the case, multimodel forecasts can underperform compared to single model forecasts (e.g., Mishra et al. 2019), as shown in this work for certain target seasons and initial times (Fig. 5). However, the use of multimodel forecasts remains a sensible choice as it avoids having to select the best performing model for specific climate variables, forecast period, and region (Delgado-Torres et al. 2022). The performance disparity between the two CanSIPSv2.1 models can be attributed in part to their different initialization scheme, as the ISBA land component in GEM5-NEMO is forced offline by ERA5 reanalysis, whereas CLASS in CanCM4i is forced by the data-constrained atmospheric fields (section 2a). Precipitation and radiation fields in these data-constrained runs can be significantly biased (Merryfield et al. 2013; Sospedra-Alfonso and Merryfield 2018), leading to issues in the soil moisture initial conditions (section 4). Unrealistic initial anomalies and persistence lead to unrealistic forecasts, indicating that CanCM4i forecasts would improve significantly with a direct initialization of its land component. CLASS and ISBA are, on the other hand, different land schemes with different components and parameterizations that impact on the representation of soil moisture and forecast performance. Our results show that ISBA, in particular, provides realistic soil moisture estimates over tropical land, or globally during warmer months, when driven by realistic atmospheric fields.

For lead times longer than 2–4 months, the performance of CanSIPSv2.1 VFSM ensemble mean hindcast is comparable to that of the best performing model for most seasons (Fig. 5). The two models have significant correlation skill not linked to the persistence of the initial anomalies, more notably GEM5-NEMO (Fig. 9). They also exhibit similar SST-driven VFSM teleconnection patterns sourced in the Pacific Ocean, although regional differences do exist (Figs. 15 and 16). The extent of Pacific Ocean influence on VFSM increases with lead time, more notably for CanCM4i (Fig. 14). However, this does not necessarily translate into enhanced forecast performance. GEM5-NEMO appears to better capitalize on this source of forecast skill whereas CanCM4i exhibits more regions of unrealistic SST-driven VFSM as lead time increases. The remote climate influence on VFSM variability and correlation skill is largely associated to ENSO, although notable influences from extratropical SST are also present. We conclude that, in general, there is quality in CanSIPSv2.1 soil moisture forecasts beyond persistence of the initial anomalies, due in part to the influence of the Pacific Ocean on soil moisture variability in teleconnected regions.

Acknowledgments. We are thankful to Marko Markovic for helpful comments on an earlier version of this paper. We are grateful also to three anonymous reviewers for their careful reading of the paper and constructive comments that helped improve the presentation of this work.


APPENDIX A

Ensemble Average, Variance, and Covariance

Using notation in section 3, the ensemble mean is

$$Y_{\mu} = \frac{1}{K} \sum_{k=1}^{K} Y_{k\mu},$$  \hspace{1cm} (A1)

the climatatological mean is

$$Y_{t} = \frac{1}{JK} \sum_{j=1}^{J} \sum_{k=1}^{K} Y_{kjl},$$  \hspace{1cm} (A2)

the ensemble mean climatatology is

$$Y_{t} = \frac{1}{JK} \sum_{j=1}^{J} \sum_{k=1}^{K} Y_{kjl},$$  \hspace{1cm} (A3)

the covariance is

$$X_j Y_j = \frac{1}{J-1} \sum_{j=1}^{J} (Y_{\mu} - Y_j)(X_{\mu} - X_j),$$  \hspace{1cm} (A4)

the ensemble mean variance is

$$\sigma_{Y_t} = \frac{1}{J-1} \sum_{j=1}^{J} (Y_{\mu} - Y_j)^2,$$  \hspace{1cm} (A5)

the total ensemble variance is

$$\sigma_{Y_t} = \frac{1}{K(J-1)} \sum_{j=1}^{J} \sum_{k=1}^{K} (Y_{kjl} - Y_{\mu})^2,$$  \hspace{1cm} (A6)

and the noise variance is

$$\sigma_{Y_n} = \frac{1}{J(K-1)} \sum_{j=1}^{J} \sum_{k=1}^{K} (Y_{kjl} - Y_{j\mu})^2.$$  \hspace{1cm} (A7)

Other quantities are straightforwardly computed from Eqs. (A1)–(A7), as indicated in section 3.
APPENDIX B

Correlation Skill Decomposition

Using notation in section 3, the ensemble mean hindcast and observation anomalies are written in terms of the reconstruction run anomalies as

\[ Y_{\mu} = \alpha_{0}\bar{Z}_{\mu} + Y_{\mu}^{*} \quad \text{and} \]
\[ X_{\mu} = \beta_{0}\bar{Z}_{\mu} + X_{\mu}^{*}, \tag{B1} \]

where \( \alpha_{0} \) and \( \beta_{0} \) are the coefficients of the linear regression of \( Y_{\mu} \) and \( X_{\mu} \) on \( Z_{\mu} \), and \( Y_{\mu}^{*} \) and \( X_{\mu}^{*} \) are the corresponding residuals. The regression coefficients are

\[ \alpha_{0} = \frac{Y_{\mu}^{2}}{\sigma_{z_{0}}^{2}} \quad \text{and} \quad \beta_{0} = \frac{X_{\mu}^{2}}{\sigma_{z_{0}}^{2}}. \tag{B2} \]

where \( \sigma_{z}^{2} \) denotes variance. By construction, \( Y_{\mu}^{*} \) and \( X_{\mu}^{*} \) are uncorrelated with \( Z_{\mu} \), and Eq. (B1) is an orthogonal decomposition of \( Y_{\mu} \) and \( X_{\mu} \).

From Eqs. (B1) and (B2), the variance of the ensemble mean hindcasts, and the covariance of the ensemble mean hindcasts and observations respectively satisfy

\[ \bar{Y}^{2} = \left( \frac{Y_{\mu}^{2}}{\sigma_{z_{0}}^{2}} \right) + \bar{Y}^{*2} \quad \text{and} \]
\[ \bar{X}^{2} = \left( \frac{X_{\mu}^{2}}{\sigma_{z_{0}}^{2}} \right) + \bar{X}^{*2}, \tag{B3} \]

leading to Eqs. (7) and (8) for the variance and correlation skill decomposition of the ensemble mean hindcasts:

\[ \sigma_{\hat{Y}}^{2} = r_{\hat{Y}Z_{0}}^{2} \sigma_{Z_{0}}^{2} + \sigma_{\hat{Y}}^{2} \quad \text{and} \]
\[ r_{X_{\mu}Y_{\mu}} = r_{X_{\mu}Z_{0}Y_{\mu}} + \frac{\sigma_{X_{\mu}} \sigma_{Y_{\mu}}}{\sigma_{X_{\mu}} \sigma_{Y_{\mu}}} r_{X_{\mu}Y_{\mu}} \tag{B4} \]

where \( \sigma_{\hat{Y}}^{2} \) denotes phase errors in the hindcast initial condition.

From Eqs. (C1) and (C2), and analogous to Eq. (B5) in appendix B, the lagged correlation of observations and reconstruction runs satisfies

\[ r_{X_{\mu}Z_{0}} = r_{X_{\mu}Z_{0}} r_{X_{\mu}Y_{\mu}} + \frac{\sigma_{X_{\mu}} \sigma_{Z_{0}}}{\sigma_{X_{\mu}} \sigma_{Z_{0}}} r_{X_{\mu}Y_{\mu}} \tag{C3} \]

leading to the decomposition of the persistence term in Eq. (8):

\[ P_{\mu} = r_{X_{\mu}Z_{0}} r_{Y_{\mu}Z_{0}} = r_{X_{\mu}Z_{0}} r_{Y_{\mu}Z_{0}} + r_{X_{\mu}Z_{0}} \frac{\sigma_{X_{\mu}} \sigma_{Z_{0}}}{\sigma_{X_{\mu}} \sigma_{Z_{0}}} r_{Y_{\mu}Z_{0}}. \]

In particular, if \( Z_{\mu} = X_{\mu} \) (i.e., perfect initialization), then the contribution from persistence of the initial anomalies to forecast correlation skill reduces to \( P_{\mu} = r_{X_{\mu}X_{\mu}Y_{\mu}Z_{0}} \).

REFERENCES


