Impact of Weak Coupling between Land and Atmosphere Data Assimilation Systems on Environment and Climate Change Canada’s Global Deterministic Prediction System

MAZIAR BANI SHAHABADI, STÉPHANE BÉLAIR, BERNARD BILODEAU, MARCO L. CARRERA, AND LOUIS GARAND

Meteorological Research Division, Environment and Climate Change Canada, Dorval, Québec, Canada

(Manuscript received 5 April 2019, in final form 29 August 2019)

ABSTRACT

A new ensemble-based land surface data assimilation (DA) system is coupled with the atmospheric four-dimensional ensemble-variational data assimilation (4D-EnVar) system with the goal of improving the analyses within Environment and Climate Change Canada’s Global Deterministic Prediction System. Since 2001, the sequential assimilation of surface variables is used to generate the initial conditions to launch the Global Environmental Multiscale (GEM) coupled forecast model. The work presented here is to replace the sequential DA with an independent surface DA system, the Canadian Land Data Assimilation System (CaLDAS) assimilating screen-level observations, and to compare assimilation experiments with CaLDAS run in uncoupled and weakly coupled modes. In the uncoupled mode, CaLDAS is used to initialize the forecast without interacting with the 4D-EnVar system. In the coupled mode, the analyses generated from CaLDAS and 4D-EnVar are used to initialize the forecast model. The analyses and forecasts from uncoupled and coupled runs are evaluated against surface and radiosonde observations over different subdomains to conclude the impact of coupling CaLDAS with 4D-EnVar. Results indicate a statistically significant reduction in bias and standard deviation at the surface for screen-level temperature and dewpoint temperature on the order of 0.1 K, and in the lower troposphere between 1000 and 500 hPa on the order of 0.1 dam for geopotential height and 0.1 K for air temperature and dewpoint depression in the coupled DA runs. The positive impact persists up to 5 days over some subdomains. It is concluded that the coupled DA approach generally performs better than the uncoupled version.

1. Introduction

At operational numerical weather prediction (NWP) centers there are ongoing efforts to improve the skill of forecasts by properly initializing not only the atmospheric variables but also other components of the Earth system such as ocean, sea ice, and land variables. Coupling between different data assimilation (DA) systems should help reducing the imbalance between different components that are currently updated independently.

Coupled DA can be categorized into two types: weakly coupled and strongly coupled. In weakly coupled DA the coupling occurs during the time integration of the coupled forecast model. Different sets of control variables are analyzed independently for each DA component, assimilating different sets of observations. In strongly coupled DA, cross-domain error covariances are utilized so that the entire Earth system is analyzed as if it were a single system. In this approach, all available observations at a given time have instantaneous impact across all domains when forming the analysis (Penny and Hamill 2017).

Many climate and subseasonal to seasonal prediction systems have achieved coupling between different components to represent the complete Earth system (Robertson et al. 2015; Guemas et al. 2016). At NWP time scales, studies report improvement in forecast skill when weak coupling is used. Mulholland et al. (2015) found that forecasts initialized by the weakly coupled atmosphere–ocean DA systems have smaller initialization shocks in the lower atmosphere, relative to separate atmosphere and ocean analyses. Sluka et al. (2016) indicated significant improvement in a perfect model scenario by using a weakly coupled DA scheme between...
the atmosphere and ocean in an observing system simulation experiment (OSSE). At the European Centre for Medium-Range Weather Forecasts (ECMWF), the Integrated Forecasting System (IFS) for atmosphere, land, and waves are coupled to the Nucleus for European Modeling of the Ocean (NEMO) model and to the prognostic thermodynamic–dynamic sea ice model, LIM2 (Laloyaux et al. 2016).

At ECMWF, the Land Data Assimilation System (LDAS) uses two-dimensional optimal interpolation (OI) for screen-level parameters analysis and the snow analysis. Analyzed screen-level temperature, relative humidity, and Advanced Scatterometer (ASCAT) surface-wetness data are assimilated in a simplified extended Kalman filter (EKF) for soil moisture analysis (De Rosnay et al. 2013). Screen-level analysis temperature increments are used in a one-dimensional OI for soil and snow temperature analysis. The weak coupling mechanism between the land and atmosphere data assimilation systems at ECMWF is explained in ECMWF IFS documentation for the 2018 operational implementation (ECMWF 2018). The Met Office has been focusing up to now on weakly coupled DA in an operational environment. Their global DA system is based on a coupled model consisting of a 0.25° ocean/sea ice model and a 60-km resolution atmosphere/land model, each with its own DA system (Lea et al. 2015). At the Met Office, the atmospheric analyses are generated by a 4D-Var DA system. A separate 3D-Var analysis of screen temperature and humidity is used to correct the soil moisture analyses via the scheme of Best and Maisey (2002). The soil moisture analyses are further improved by assimilating ASCAT observations, using a nudging scheme (Dharssi et al. 2011). Santanello et al. (2016) examined the impact of assimilating Advanced Microwave Scanning Radiometer for EOS (AMSR-E) soil moisture retrievals on coupled WRF Model forecasts in the southern Great Plains. The results showed the potential use of satellite soil moisture to improve land surface model spin-up and coupled land–atmosphere forecast skills. In a later study, Santanello et al. (2019) reports the recent studies that demonstrated the critical role of the land surface, and in particular soil moisture, in terms of impacts on precipitation, temperature, and humidity, and the planetary boundary layer state in the context of coupled land–atmosphere DA in NWP. Draper and Reichle (2019) used a weakly coupled land and atmosphere data assimilation system for NASA’s Global Earth Observing System model to demonstrate the advantage of assimilating satellite soil moisture into an atmospheric reanalysis. They reported a reduction of RMSE against independent station observations for various variables, including daily maximum 2-m temperature and 2-m specific humidity, when Advanced Scatterometer and soil moisture ocean salinity retrievals are assimilated into the system. Penny and Hamill (2017) summarize the progress in coupled DA made at operational centers and highlight developments in the research community that are helping to advance the field.

At Environment and Climate Change Canada (ECCC), atmosphere and ocean DA systems are uncoupled: the atmospheric analysis is generated with the four-dimensional ensemble variational (4D-EnVar) DA system, which receives the sea surface temperature and sea ice analysis from independent sources. The ocean DA system provides sea surface temperature and sea ice analyses, which together with the atmospheric 4D-EnVar analyses, provide the initial condition to launch the coupled atmosphere–ocean model providing medium-range forecasts. Since 2001, sequential assimilation of surface variables is used as the initial condition to launch the coupled atmosphere–land forecast model (Bélair et al. 2003a,b). The work presented here is to replace the surface sequential DA with an independent surface DA system, the Canadian Land Data Assimilation System (CaLDAS), and compare the results when CaLDAS is run in an uncoupled and weakly coupled mode. CaLDAS has been developed at ECCC. The need for accurate initial conditions has been met by the use of separately generated atmosphere and land state estimates. CaLDAS is defined in more detail in section 2b(1). CaLDAS provides consistent land surface initial conditions for coupled atmosphere–land forecasts. The objective of this work is to examine whether such initial conditions will provide improved forecasts for the various components of the system. In this paper, the experiments are designed for CaLDAS surface analyses to be used in uncoupled and coupled modes to the 4D-EnVar atmospheric analyses before launching the forecasts. In section 2, the atmospheric and surface models, their corresponding DA systems, and the observations used to validate the forecast are described. In section 3, the results of the two seasons coupled and uncoupled DA systems are evaluated. Section 4 concludes the article with a discussion on findings.

2. System description

Coupled and uncoupled assimilation experiments were conducted for two periods of two and half months, in boreal summer (15 June 2016–31 August 2016) and boreal winter (15 December 2016–28 February 2017). The first 15 days of each assimilation experiment are considered as a spinup period and are not used for evaluation of the forecasts. All experiments have the same configuration for the atmospheric and land surface models.
The same sets of observations are assimilated in each experiment, very similar to the Global Deterministic Prediction System (GDPS) implemented operationally at ECCC in November 2017. The assimilation experiments were conducted in two modes: uncoupled and coupled. Figure 1 shows the schematic diagram to demonstrate the interaction between the surface and upper-air DA systems for two consecutive assimilation cycles for the two experiments considered in this study.

In the uncoupled mode (E16OFF for boreal summer and H17OFF for boreal winter) 4D-EnVar provides the upper-air analysis, while the sequential assimilation technique is used to generate the surface analyses (Bélair et al. 2003a). The 4D-EnVar and sequential assimilation analyses initialize the coupled forecast model for generating 0–9-h (short range) forecasts, used during the next assimilation cycle. CaLDAS runs offline to generate another set of surface analyses. The 4D-EnVar atmospheric and CaLDAS-generated surface analyses are then used to initialize the coupled forecast model for the medium- and long-range forecasts. There is no feedback from CaLDAS to the next assimilation cycle in the uncoupled mode, and CaLDAS analyses only affect the medium- and long-range forecasts.

In the coupled mode (E16CPL for boreal summer and H17CPL for boreal winter), CaLDAS uses the 4D-EnVar lowest level analyses to generate the surface analyses. The 4D-EnVar atmospheric and CaLDAS surface analyses are used as initial conditions to drive the coupled forecast model [introduced in section 2a(2)] and generate short-, medium-, and long-range forecasts. In the coupled mode, CaLDAS feedbacks to the upper-air analyses of the next cycle through its influence on the short-range forecasts that propagate information from one cycle to the next. In this study, the global medium- and long-range forecasts in GDPS system are generated up to 240 h.

a. Atmospheric component

1) ATMOSPHERIC ASSIMILATION SYSTEM

The global assimilation is a 4D-EnVar system (Buehner et al. 2015) that provides the initial condition for global forecasts. In this system the background error covariances
are calculated from the combination of flow-dependent estimates defined from ensemble Kalman filter (EnKF) forecasts and static climatological estimates. Below 40 hPa, the background error covariance is the average of dynamic and static covariances. Above 10 hPa, only the static covariances are used. Between 10 and 40 hPa, there is a linear transition. The length of the assimilation window is 6 h, and it is discretized in 15-min time slots for assimilation of the observations. The operationally assimilated observations include those from radiosondes, aircrafts, wind profilers, land stations, ships and buoys, scatterometers, atmospheric motion vectors, satellite-based radio occultation, and microwave and infrared satellite sounders and imagers. The analyzed variables are the horizontal winds, temperature, humidity, surface pressure, and sea surface temperature. About 13 million observations per day are assimilated in the operational global system.

2) FORECAST MODEL

The Global Environmental Multiscale (GEM) atmospheric model is the operational forecast model used at ECCC (Côté et al. 1998a,b). The GEM model runs on a global yin–yang grid with a horizontal grid spacing of 0.225° × 0.225° in the meridional and zonal directions of the computational grids. The model contains 85 hybrid levels, extending from the surface to 0.1 hPa. Compared to the operational version, as described by Bélaire and Roch (2009), there are significant modifications made to improve the representation of atmospheric physics in the GEM model. These modifications focused on two main goals: increasing the vertical resolution and improving the hydrological cycle in the model. The details for these changes are further described by McTaggart-Cowan et al. (2019a, b). The results of this study are generated with this modified version of Canadian GEM model, scheduled for operational implementation in 2019.

b. Land surface components

1) LAND SURFACE ASSIMILATION SYSTEM

The ensemble-based CaLDAS has been developed at ECCC to generate land surface analyses (Carrera et al. 2015) and has been implemented at ECCC’s operations for short-range numerical weather prediction in 2014 (Milbrandt et al. 2016). In CaLDAS, the first guess solution is from an external land surface model and an ensemble of 24 members is used to produce the background error covariances matrix. The screen-level observed variables, namely the 2-m temperature (TT2m) and dewpoint temperature (TD2m), and snow depth are assimilated every 6 h at the beginning of the assimilation window. In CaLDAS, an ensemble of screen-level analyses is first produced based on a sequential DA approach (for each member) using perturbed screen-level observations combined with the ensemble of first guesses. This ensemble of screen-level analyses is then used based on an ensemble Kalman filter approach to produce an ensemble of land surface temperature, soil moisture, and snow depth analyses. The ensemble mean is then used to initialize the atmospheric model.

2) LAND SURFACE MODEL

The operational land surface model at ECCC has been the Interactions between Soil, Biosphere, and Atmosphere (ISBA) scheme since 2001 and is described in detail in Noilhan and Planton (1989) and Bélaire et al. (2003a, b). Recently, a more advanced land surface model, Soil, Vegetation, and Snow (SVS) (Alavi et al. 2016; Husain et al. 2016), has been developed to eventually replace ISBA. SVS is the land surface model used in all experiments in this work.

SVS is forced with the lowest level (20 m for thermodynamic and 40 m for dynamic variables) atmospheric data from short-range forecasts from ECCC NWP model output. The forcing variables are the incident short- and longwave radiation, surface pressure, precipitation at the surface, air temperature, specific humidity, and winds. As discussed in Carrera et al. (2015), the precipitation ensembles are generated by making use of the operational Canadian Precipitation Analysis (CaPA; Mahfouf et al. 2007; Lespinas and Fortin 2015) methodology. CaPA uses the OI algorithm to combine information from a model first-guess precipitation field with precipitation gauge observations, which are from land surface synoptic reports (SYNOP) network and aviation routine weather reports (METAR) surface observing network. The CaPA product has been validated extensively over the North America region (Fortin et al. 2015). Therefore, precipitation forecast scores are only discussed over North America in this paper, even though the forecasts are global in nature.

3) EVALUATION OF SOIL MOISTURE ANALYSES

The Level 3 NASA Soil Moisture Active Passive (SMAP; Entekhabi et al. 2010) retrievals are used for the validation of the soil moisture analysis. The L-band radiometer such as that on SMAP has a nominal penetration depth of 5 cm, although it can be much shallower when soil water content is high (Escorihuela et al. 2010). SMAP estimates the soil moisture in the top 5 cm of soil with an unbiased root mean squared error no greater than 0.04 cm$^3$ cm$^{-3}$ (Entekhabi et al. 2014). Only the top 5-cm soil moisture analyses of CaLDAS are evaluated against SMAP observations in this study.
The Level 3 SMAP retrievals (Chan et al. 2018) are distributed on 9-km EASE-grid 2.0 whereas the CaLDAS soil moisture analyses are on a 25-km grid. In this study, a nearest-neighbor approach was used to re-map the SMAP retrievals to the CaLDAS 25-km grid. Prior to the evaluation, only the SMAP soil moisture retrievals with recommended quality flag are used. Then the data were additionally filtered to avoid using retrievals with less than 36-km distance to nearby significant water bodies. Furthermore, to avoid the known problem of obtaining untrustworthy soil moisture estimates when soil freezing occurs (Brocca et al. 2011), a surface temperature threshold of 3°C is used. Calvet and Noilhan (2000) emphasized the need to normalize the soil moisture because of an existing bias in the simulated soil moisture observations. Also, there are biases attributed to the soil moisture retrievals from satellite observations (Reichle and Koster 2004) and differences in soil moisture values from the satellite and the model that neither agrees better with ground data (Reichle et al. 2004). To reduce the negative impact of these discrepancies, similar to Sabater et al. (2007), both the analyzed and retrieved soil moisture data were normalized, at each grid point, using their corresponding maximum and minimum values during the verification period according to Eq. (1):

$$w_N = \frac{w_g - w_{min}}{w_{max} - w_{min}},$$

(1)

where $w_g$ is the original soil moisture value at a grid point, $w_{min}$ and $w_{max}$ are the minimum and maximum values of soil moisture during the verification period at the same grid point, and $w_N$ is the normalized soil moisture value. The retrieved and analyzed soil moisture data were compared in a normalized space ranging from 0 to 1. The absolute bias, standard deviation of error (stddev), and correlation of the retrieved versus analyzed soil moisture are calculated at each grid point and the results are compared between the two uncoupled and coupled experiments during two-month period of 1 July 2016–31 August 2016 and 1 January 2017–28 February 2017.

3. Results

a. Verification and comparison of global soil moisture analyses

To compare the soil moisture analyses between the uncoupled and coupled experiments, CaLDAS is rerun in the uncoupled experiment to ensure the differences in soil moisture values between the two experiments are only due to the coupling mechanism. The comparison of the two-month mean of the top layer (5 cm) soil moisture analyses between the coupled and uncoupled experiments during summer and winter is shown in Fig. 2. The coverage is not complete over the globe, and the surface analyses are available only every 3 h. During summer, the E16CPL is drier than E16OFF over east and north Canada, parts of South America, and northern Asia, by at most 0.125 m$^3$ m$^{-3}$. However, E16CPL is wetter compared to E16OFF in central Asia and southern Europe during summertime by a similar quantity. During winter compared to H17OFF, the H17CPL experiment is drier over Central America by at most 0.125 m$^3$ m$^{-3}$. These regional differences in soil moisture analysis in the coupled and uncoupled experiments will contribute to the spatially averaged differences between the two experiments over different regions. This will be shown and discussed later, when the skill of forecasts for screen-level variables are computed and compared for a specific region, using independent surface observations. The geographical domains used for the evaluation of the forecasts are marked by black boxes in Fig. 2a.

Equation (1) is used to calculate the normalized soil moisture analysis. For each experiment, the absolute bias, stddev, and correlation of this quantity is computed, using the SMAP normalized retrieved soil moisture as reference. Figure 3 shows the absolute bias, stddev, and correlation between the normalized soil moisture analyses and the SMAP retrievals of the coupled experiment for two seasons (summer in the top panel, and winter in the bottom panel), respectively. The statistics are calculated only for the grid points with more than 30 SMAP overpasses during the two-month verification period in the coupled experiment. The difference statistics between the normalized soil moisture analysis for absolute biases and stddev of the uncoupled minus coupled experiments are also shown in Fig. 4 for the two seasons. Similarly, the coupled minus uncoupled experiment difference of the correlation between the normalized soil moisture analysis and SMAP retrievals are shown in Fig. 4. The difference statistics are calculated for the grid points with more than 30 SMAP overpasses in either the coupled or uncoupled experiments during the same period. Otherwise, gray shading is used to indicate when the sample size is smaller than 30 or no observation data are available. The SMAP retrieval data from the ascending and descending nodes are combined to increase the sample size. At each grid point for absolute bias and stddev (correlation) difference statistics, blue indicates smaller (larger) value in the coupled, compared to the uncoupled experiment.

In North America during summertime, the bias and stddev of the normalized soil moisture analysis, using the reference SMAP normalized retrieved soil moisture,
is smaller in the coupled run by at most 0.225 m$^3$ m$^{-3}$ for bias and 0.175 m$^3$ m$^{-3}$ for the stddev, compared to the uncoupled experiment (Figs. 4a,c). Similarly, the temporal correlation of the soil moisture analysis during the summer season over North America is larger in the coupled run by at most 0.45 as shown in Fig. 4e. A comparable behavior is seen during the winter season (Figs. 4b,d,f), with fewer available SMAP retrieved data for validation due to the threshold used to eliminate the possibility of frozen soil. The reduction in the coverage of SMAP retrieved soil moisture data during the cold season is more noticeable in regions at higher altitudes (e.g., Asia) where there are fewer data passing the quality control and available to perform a statistical analysis (more gray shading). Similar improvements in bias, stddev, and correlations are noticeable over other parts of the globe during the warm and cold seasons.

As shown in Fig. 4, when enough data are available to perform statistical analysis, the one to one comparison with SMAP retrieved soil moisture reveal smaller biases and stddev and larger correlations in the coupled experiments, indicated by the blue color on the map. The soil moisture analyses from the coupled runs are better fitted with the SMAP data, compared to the uncoupled runs. This is due to better quality of the short-term forecasts, as a result of coupling with surface, which leads to the improvement in soil moisture analyses.

b. Verification and comparison of global upper-air analysis

The upper-air analyses of the coupled and uncoupled experiments are compared for selected levels in the lower atmosphere where the impact from different initial conditions is more pronounced. The comparisons between the two-month averaged temperature and specific humidity analyses of the coupled and uncoupled experiments at 850 and 925 hPa levels for summer and winter are shown in Fig. 5. The upper-air comparison between the coupled and uncoupled run analyses will be linked to the respective behavior of each run when the...
surface and upper-air forecast skills over smaller domains are assessed further in this study.

Over western Asia during the summer season, the E16CPL is warmer (by at most 2 K) and drier (by at most 2.7 g kg\(^{-1}\)) than the E16OFF run at the 925-hPa level. Similar but smaller differences are seen at the 850-hPa level. Over Africa, the E16CPL is colder by at most 1.5 K and wetter by at most 2 g kg\(^{-1}\) than the E16OFF run at the 925-hPa levels (see Figs. 5e and 5f), with similar but smaller differences happening at 850 hPa. During winter, the magnitude of the differences between the two experiments is larger than for the summer period. At 925 hPa, the coupled experiment is colder by at most 2.5 K and wetter by at most 4.3 g kg\(^{-1}\) over central Africa, compared to the uncoupled experiment (Figs. 5g,h), with similar but smaller differences at 850 hPa. Over the North America domain during winter season, the coupled run is colder by at most 1 K and wetter by at most 1.5 g kg\(^{-1}\)
than the uncoupled run at 925 hPa (Figs. 5g,h). During the summer season, the coupled experiment is warmer by 1 K than the uncoupled run at the 925-hPa level, especially over the western U.S. coastal region (Fig. 5e).

When averaged over the entire global domain, the verification of the upper-air analyses using the global radiosondes does not indicate any statistically significant difference between the coupled and uncoupled experiments (not shown). The verifications of the upper-air analyses and forecasts for a few subdomains (shown by black boxes in Fig. 2a) are presented in the following section.

c. Upper-air and surface forecast verification over different subdomains

The NWP impact of the weak coupling between the atmosphere and land surface assimilation systems are
evaluated using independent upper-air and surface observations. The forecasts are evaluated over different subdomains separately: North America, Europe, and Asia (black boxes in Fig. 2a). Results presented in sections 3a and 3b are referenced in forecast verifications against in situ observations at the analysis time.

For upper-air forecast verification, forecasts launched at 0000 and 1200 UTC are combined and evaluated at 0- (analysis time), 48-, and 120-h forecast lead times against radiosonde observations. Geopotential height, air temperature, and dewpoint depression are the set of quantities chosen for the evaluation of the analysis and forecast skills. Since the scores are averaged quantities over each subdomain, the differences between the two experiments might look small. The color circles, which show which experiment performs better with a statistical significance above 90% at each level, are superimposed on the statistic curves to facilitate the evaluation.
At the surface, the TT2m and TD2m of the 0000 UTC runs are compared against SYNOP, METAR, and Surface Weather Observation XML (SWOB) surface observations for the 0–240-h forecast lead time. The screen-level observations are used for validation only if the elevation difference between the station and the model topography is less than 100 m.

1) NORTH AMERICA

The comparisons of the bias and stddev of TT2m and TD2m of the H17OFF and H17CPL experiments over North America are shown in Fig. 6. Compared to the H17OFF, the cold daytime TT2m bias is reduced by 0.1–0.2 K in H17CPL (Fig. 6a), while the stddev of TT2m is similar between the two experiments (Fig. 6b). H17CPL has a smaller two-month mean soil moisture analysis over the east and south United States (Fig. 2b), when compared against the H17OFF experiment (H17CPL is drier). However, Fig. 6c at the analysis time (run hour 0) indicates that the H17CPL experiment has a positive TD2m bias, compared to the negative TD2m bias in H17OFF experiment. This translates into larger mean soil moisture analysis in H17CPL, compared to the H17OFF. This discrepancy could be due to the fact that there are few observation stations over the south of North America region, where Fig. 2b shows the largest soil moisture differences between the two experiments. Overall, there is a reduction of dewpoint temperature bias by at most 0.5 K and daytime reduction of TD2m stddev by at most 0.2 K in the H17CPL experiment with 90% statistical significance up to 3-day forecasts.

The verifications of the upper-air analysis and 48- and 120-h forecasts over North America against radiosonde observations are shown in Fig. 7. For bias in geopotential height, neither experiment H17OFF nor H17CPL demonstrates a clear superior performance. In Fig. 5g it was shown that H17CPL is colder than H17OFF, but the cold 925-hPa temperature analyses in H17CPL is comparable to H17OFF experiment and the difference is not statistically significant (full line in Fig. 7b). Similarly, the two experiments produce comparable dewpoint temperature bias at 925 hPa (Fig. 7c, full line), even though H17CPL analyses is wetter than H17OFF at that level (Fig. 5h).

For the verifications forecasts, compared to H17OFF, there is reduction of the cold bias by 0.1–0.2 K and...
reduction of moist bias by 0.2–0.5 K in H17CPL in the boundary layer up to 850 hPa in 48- and 120-h forecasts. The stddev values of geopotential height, temperature, and dewpoint depression in the boundary layer are also smaller in the H17CPL coupled run in 48- and 120-h forecasts with statistical significance. The surface and upper-air verifications of the 48- and 120-h forecasts indicate that the coupled experiment performs better over North America.

The upper-air impact of the coupling between the atmosphere and surface is further evaluated during the summer season and shown in Fig. 8. E16CPL has a smaller geopotential height bias at the analysis, 48-, and 120-h forecasts with statistical significance. The surface and upper-air verifications of the 48- and 120-h forecasts indicate that the coupled experiment performs better over North America.

The upper-air impact of the coupling between the atmosphere and surface is further evaluated during the summer season and shown in Fig. 8. E16CPL has a smaller geopotential height bias at the analysis, 48-, and 120-h forecasts with statistical significance. The surface and upper-air verifications of the 48- and 120-h forecasts indicate that the coupled experiment performs better over North America.

Over Europe during wintertime, the coupled and uncoupled runs perform similarly at the surface and
upper air (results not shown). The comparisons of the surface forecast scores for the two experiments during summertime are shown in Fig. 9. The two experiments have similar cold bias up to 24 h, while the coupled run (E16CPL) has a smaller daytime cold bias by 0.1 K and a slightly larger nighttime warm bias between 24 and 240 h (Fig. 9a). The E16CPL experiment has a smaller moist bias of the order 0.1 K from 0 to 240 h, as shown in Fig. 9c. The coupling between atmosphere and surface DA systems in the E16CPL experiment has a small positive impact on the TT2m and TD2m biases over Europe. In the coupled run, there is also statistically significant reduction of the stddev of TT2m and TD2m during daytime by 0.2 K, that extends up to the 120-h forecast (Figs. 9b,d).

The upper-air verification of uncoupled and coupled run forecasts against radiosondes is shown for geopotential height, air temperature, and dewpoint depression in the boundary layer at 48- and 120-h forecasts (Fig. 10). There is a statistically significant reduction of geopotential height biases in the E16CPL experiment at 120-h forecast at the 850- and 700-hPa levels by 0.1–0.2 dam, shown in Fig. 10a. The temperature biases at 925 and 850 hPa at 48-h forecast in the E16CPL run are similar to the E16OFF run. However, the E16OFF has slightly lower but statistically significant temperature biases in the boundary layer at 120-h forecasts at 925, 850, and 700 hPa (Fig. 10b). The changes in dewpoint depression biases between the experiments are not statistically significant, except at 925 hPa where E16CPL has lower biases by 0.1 K. The differences between the geopotential height stddev of the two experiments are not statistically significant. At the 925- and 850-hPa levels, the E16CPL run has a slightly lower stddev of temperature at 48 h by 0.1 K (Fig. 10e) and lower stddev of dewpoint depression at 48- and 120-h forecasts by 0.1 K (Fig. 10f). In general over Europe, the coupling of the atmosphere and surface improves
the 48-h forecast biases and stddev of temperature and 48- and 120-h forecast stddev of dewpoint depression in the boundary layer.

3) Asia

The summertime surface forecast scores over Asia are shown in Fig. 11. Wintertime results show similar signatures and are not shown for brevity. Similar to Europe, the coupled run E16CPL has slightly smaller cold bias after 120 h, although the improvement is not statistically significant. Figure 11c shows the moist TD2m bias is smaller in the coupled run by 0.1 K for days 3–5, compared to the uncoupled experiment. The stddev values of TT2m and TD2m are smaller in the coupled experiment by maximum of 0.2 K (Figs. 11b,d) and statistically significant for 0–108 h.

The upper-air verification of uncoupled and coupled run forecasts against radiosondes over Asia is shown in Fig. 12. The geopotential height biases in 48- and 120-h forecasts are reduced in E16CPL by 0.1–0.3 dam in the boundary layer. The E16CPL experiment shows a statistically significant reduction in 700- and 850-hPa temperature biases at 48- and 120-h forecasts by 0.1–0.2 K (Fig. 12b). Over Asia in the E16CPL run, the two-month mean specific humidity analyses at 850- and 925-hPa levels are slightly smaller than the E16OFF experiment (Figs. 5b,f). This is also confirmed by a 0.1-K smaller negative dewpoint depression bias at the analysis time in E16CPL run at the 850- and 925-hPa levels (Fig. 12c, solid lines), when compared against radiosondes. The E16OFF has smaller dewpoint depression bias at 850 hPa in 48- and 120-h forecasts by 0.1 K, while the two runs at other levels and forecast hours show similar or statistically insignificant performance in terms of dewpoint depression bias. The two experiments have comparable stddev of geopotential height (Fig. 12d). At the 925-, 850-, and 700-hPa levels, the E16CPL run has a lower stddev of temperature by 0.1 K at 48 h (Fig. 12c) and lower stddev of dewpoint depression at 48 h by 0.1–0.2 K (Fig. 12f), limited to the 925- and 850-hPa levels at 120-h forecasts.

4) Precipitation Verification over North America

As mentioned in the introduction, the precipitation scores are only discussed over the North America domain, where the CaPA product is used for validation of the
precipitation. Several metrics are used, namely, the frequency bias index (FBI), the false alarm ratio (FAR), the equitable thread score (ETS), and the probability of detection (POD). To define these metrics, “yes” refers to “a precipitation event will happen” and “no” is when “the precipitation event will not happen.” FBI is the ratio of the forecast frequency of “yes” events compared to the observed frequency of “yes” events. FBI indicates whether the forecast system has a tendency to underestimate (FBI < 1) or overestimate (FBI > 1) events. FAR is the fraction of the forecast “yes” events that actually did not occur. FAR ranges from 0 to 1 with FAR = 0 a perfect score. ETS is the measure of the forecast “yes” events corresponding to the observed “yes” events. The perfect ETS score is 1. POD is the fraction of the observed “yes” events that were correctly forecasted. The perfect POD score is 1.

Figure 13 shows the four metrics for the 24-h accumulation of precipitation, from 1200 UTC (12 h; day 1) to 1200 UTC (36 h; day 2) over North America as function of accumulation threshold value in millimeters (mm). At longer ranges, the precipitation forecast skills are very similar between the two experiments (not shown). The impact of coupling on the precipitation is larger during wintertime and hence only winter experiments are compared. The reduction of FBI indicates smaller overforecasting of rain frequency in the coupled run compared to the uncoupled experiment (Fig. 13a). The FBI values decrease by 5%–7% for accumulations greater than 1 and 2 mm. The improvements in FBI become insignificant at larger precipitation thresholds. For accumulations greater than 1, 2, 5 and 10 mm, there is a 2%–3% reduction in the FAR values in the coupled experiment (Fig. 13b). Similar to the FBI score, the FAR scores become statistically insignificant at larger accumulation quantities owing to a smaller sample size of larger precipitation events. The ETS of the coupled run shows 1%–2% larger fraction of observed and/or
forecast events that were correctly predicted, compared to the uncoupled run (Fig. 13c), for accumulations greater than 1, 2, 5, and 10 mm. The POD values from the two runs are roughly similar except that the uncoupled run predicts smaller rain events slightly better by 1%. Overall, the results show the coupled experiment has slightly better precipitation scores over North America, compared to the uncoupled run.

4. Discussion and conclusions

The work presented here is the comparison of CaLDAS performances in an uncoupled versus weakly coupled atmosphere–land DA in ECCC’s Global Deterministic Prediction System (GDPS). Coupled and uncoupled assimilation cycles were conducted for two periods of two and half months, in boreal summer and boreal winter. In the uncoupled assimilation experiments, CaLDAS runs offline, does not feedback to the assimilation cycle, and is only used to generate surface analyses to initialize the medium- and long-range forecasts. In the coupled assimilation experiments, the 4D-EnVar analyses provide the upper-air forcings to run CaLDAS to generate the land surface analyses. CaLDAS feedbacks to the 4D-EnVar upper-air analysis of the next cycle through its influence on the short-range forecasts that propagates information from one cycle to the next. The two assimilation experiments have the same configuration for the atmospheric and land surface models. The same sets of observations were assimilated in each experiment, very similar to the operational GDPS. The first 15 days of each assimilation experiment were considered as the spinup period and not used for evaluation of the forecasts.

The soil moisture analyses from the coupled and uncoupled runs were normalized and validated against normalized SMAP soil moisture retrievals. The difference statistics of the normalized soil moisture absolute bias, standard deviation of error, and correlation of the retrieved versus analyzed soil moisture were calculated for each experiment for the period of two months (summer: 1 July 2016–31 August 2016; winter: 1 January 2017–28 February 2017). Compared to the uncoupled runs during each season, the global maps of the normalized soil moisture, in reference to normalized SMAP soil moisture retrievals, show a smaller bias and smaller standard deviation of error and larger correlation in the coupled experiments. The validation of upper-air analyses using radiosonde observations over the global domain does not indicate any statistically significant differences between the coupled and uncoupled experiments at the analysis time.
For upper-air forecast verifications, the 0000 and 1200 UTC forecasts are combined and validated at 0- (analysis time), 48-, and 120-h forecast lead time against radiosonde observations for geopotential height, air temperature, and dewpoint depression. At surface, the TT2m and TD2m of the 0000 UTC runs are compared with the combination of SYNOP, METAR, and SWOB surface observations for 0–240-h forecasts. The forecasts are evaluated over different subdomains such as North America, Europe, and Asia.

For each subdomain, a statistically significant reduction in bias and stddev for TT2m and TD2m, on the order of 0.1 K, was found in the coupled DA runs, compared to the uncoupled runs. The magnitude of the positive impact differs from one subdomain to another, but in general the positive impact extends to 120 h. Similarly for the verifications of the upper-air forecasts, there is a statistically significant reduction in bias and stddev of geopotential height on the order of 0.1 dam and air temperature and dewpoint depression on the order of 0.1 K in the lower atmosphere between 1000 and 500 hPa in different subdomains. In conclusion, the coupled DA run generally performs better than the uncoupled version. Improved 0–9-h forecasts, as a result of weak coupling between surface and atmospheric DA systems, lead to better soil moisture analyses, which in turn lead to better upper-air forecasts, especially in the first 48 h.

Based on these findings, the operational implementation of the new CaLDAS surface DA system should be in weakly coupled mode with the atmospheric DA component. Toward this goal, the first step would be to evaluate the impact of replacing the current operational
sequential surface DA with CaLDAS. This will be the subject of future studies. We also plan to investigate the impact of strongly coupled land–atmosphere DA systems in future studies.

Acknowledgments. The authors wish to thank Dr. Ron McTaggart-Cowan and Dr. Ayrton Zadra for their help to clarify the changes made to the forecast model. We acknowledge financial support from the Government Related Initiatives Program (GRIP) from Canadian Space Agency.

REFERENCES


FIG. 13. Different verification metrics of 24-h precipitation accumulation in H17CPL (red) vs H17OFF (blue) experiments as function of accumulation threshold (mm) for 59 individual forecasts launched at 0000 UTC. (a) Frequency bias index (FBI), (b) false alarm ratio (FAR), (c) equitable threat score (ETS), and (d) probability of detection (POD). The accumulations are from $T + 12$ to $T + 36$ over North America. Precipitation forecasts are validated against the CaPA precipitation analysis. At the bottom of each panel the difference between the two experiments is shown with gray shading representing the 90% confidence level based on a bootstrapping technique.