The Canadian Land Data Assimilation System (CaLDAS): Description and Synthetic Evaluation Study

MARCO L. CARRERA, STÉPHANE BÉLAIR, AND BERNARD BILODEAU
Meteorological Research Division, Environment Canada, Dorval, Quebec, Canada

(Manuscript received 9 May 2014, in final form 3 February 2015)

ABSTRACT

The Canadian Land Data Assimilation System (CaLDAS) has been developed at the Meteorological Research Division of Environment Canada (EC) to better represent the land surface initial states in environmental prediction and assimilation systems. CaLDAS is built around an external land surface modeling system and uses the ensemble Kalman filter (EnKF) methodology. A unique feature of CaLDAS is the use of improved precipitation forcing through the assimilation of precipitation observations. An ensemble of precipitation analyses is generated by combining numerical weather prediction (NWP) model precipitation forecasts with precipitation observations. Spatial phasing errors to the NWP first-guess precipitation forecasts are more effective than perturbations to the precipitation observations in decreasing (increasing) the exceedance ratio (uncertainty ratio) scores and generating flatter, more reliable ranked histograms. CaLDAS has been configured to assimilate L-band microwave brightness temperature TB by coupling the land surface model with a microwave radiative transfer model. A continental-scale synthetic experiment assimilating passive L-band TBs for an entire warm season is performed over North America. Ensemble metric scores are used to quantify the impact of different atmospheric forcing uncertainties on soil moisture and TB ensemble spread. The use of an ensemble of precipitation analyses, generated by assimilating precipitation observations, as forcing combined with the assimilation of L-band TBs gave rise to the largest improvements in superficial soil moisture scores and to a more rapid reduction of the root-zone soil moisture errors. Innovation diagnostics show that the EnKF is able to maintain a sufficient forecast error spread through time, while soil moisture estimation error improvements with increasing ensemble size were limited.

1. Introduction

It has become increasingly clear that an accurate initialization of the land surface is important for skillful weather and seasonal climate predictions (e.g., Koster et al. 2004; Drusch 2007; Drusch and Viterbo 2007; Gao et al. 2008; Mahfouf 2010; Douville 2010). Soil moisture, snow characteristics, surface temperature, and vegetation properties of the land surface influence both the water and energy budgets, exerting important controls on land–atmosphere interactions.

Given the significant advances in the development of earth observation systems, real opportunities exist to improve the representation of the land surface, and soil moisture in particular, within numerical weather prediction (NWP) systems. Two separate satellite missions dedicated to the measurement of microwave radiation emitted from the soil surface in the highly sensitive L-band frequency have been developed. In November 2009, the Soil Moisture Ocean Salinity (SMOS) mission was launched by the European Space Agency (ESA) as the second Earth Explorer Opportunity mission as part of the ESA’s Living Planet Programme. An L-band radiometer on board the SMOS satellite measures the microwave emission from the soil, which provides information on the soil moisture content in the top 2–5 cm of the soil (Kerr et al. 2010).

The National Aeronautics and Space Administration (NASA) launched the Soil Moisture Active Passive (SMAP) mission in January 2015 (Entekhabi et al. 2010b). In addition to a passive L-band radiometer, the SMAP mission payload also includes an L-band active radar that
will enable higher-resolution soil moisture mapping. Additional passive and active sensors, which operate at shorter wavelengths (C and X band), are also providing information on soil moisture. These include the Metop-A Advanced Scatterometer (ASCAT; Scipal et al. 2008; Mahfouf 2010; Draper et al. 2012), and the Advanced Microwave Scanning Radiometer for Earth Observing System (AMSR-E; de Jeu et al. 2008; Draper et al. 2009, 2011, 2012).

At several meteorological centers, including the Canadian Meteorological Centre of Environment Canada (EC), soil moisture is inferred from short-range NWP forecast errors in screen-level temperature and humidity (Bélair et al. 2003a; Drusch and Viterbo 2007; Mahfouf et al. 2009). Soil moisture is used as a sink variable where errors in atmospheric forcing and the land surface model can accumulate over time (Seuffert et al. 2004; Drusch and Viterbo 2007). To better represent the land surface in environmental prediction and assimilation systems, the Canadian Land Data Assimilation System (CaLDAS) is being developed at EC’s Meteorological Research Division (MRD). It is planned that CaLDAS will provide the initial conditions, including soil moisture, surface temperature, snow cover, and vegetation properties to both EC’s deterministic and ensemble prediction systems.

Much of the early operational development of CaLDAS focused on the assimilation of screen-level variables for the analysis of the soil moisture state; however, this study will emphasize the research work done in the context of assimilating passive L-band microwave data for soil moisture initialization. Results from the assimilation of screen-level variables will be reported on in a forthcoming paper. The first part of this study will present and evaluate the method used within CaLDAS to generate an ensemble of precipitation analyses by combining an NWP model forecast with precipitation observations. In the second part of this study, results from a continental-scale synthetic data assimilation experiment are presented where passive microwave L-band brightness temperatures TBs are assimilated. The added skill in the analysis of the soil moisture state from the inclusion of these TB observations is quantified by means of domain-averaged performance metrics. Innovation diagnostics are presented to evaluate the assimilation scheme performance. The experiments in this study were performed as part of the development of CaLDAS and helped to determine the optimal configuration for a future NWP forecasting implementation.

2. CaLDAS

Figure 1 shows the main components of CaLDAS in blue. In the subsequent sections, each of these components of CaLDAS is described in greater detail.
a. External land surface model

CaLDAS is built around an external land surface modeling system that provides the temporally and spatially varying first guess of the land surface state. This external or offline approach is used in the North American Land Data Assimilation System (NLDAS; Mitchell et al. 2004) and the Global Land Data Assimilation System (GLDAS; Rodell et al. 2004), as well as the surface externalized (SURFEX) module at Météo-France (Masson et al. 2013).

The advantage of using the offline approach is that the land surface model can be integrated at much higher horizontal resolutions, when compared with the full 3D version of the atmospheric model, at a fraction of the computational costs (Cosgrove et al. 2003; Mailhot et al. 2010). The principal limitations on the land surface model resolution are mainly a function of the geophysical field forcing datasets. In the current framework, the atmospheric forcing data are taken from the lowest model level in EC’s operational NWP model, which is at a height of roughly 40 m. This is important to describe turbulent transfers in the surface boundary layer in order to allow for the assimilation of screen-level meteorological observations (Balsamo et al. 2007; Mahfouf et al. 2009).

The land surface model configured within CaLDAS is the Canadian implementation of the Interactions between Soil, Biosphere, and Atmosphere (ISBA) model (Noilhan and Planton 1989; Noilhan and Mahfouf 1996), which is described in detail in Béclair et al. (2003a,b). This land surface scheme has been used operationally at the Meteorological Service of Canada (MSC) since 2001. The Canadian implementation of ISBA consists of two layers, a superficial layer and a rooting depth layer, whose thickness depends on vegetation type and varies between 1 and 3 m, in which temperature and soil water contents evolve following the force–restore equations (Deardorff 1978). The superficial layer has an arbitrarily set depth of $d_1 = 10$ cm, where the depth of soil influenced by external atmospheric forcing varies as a function of soil moisture and soil texture. A single effective dynamical range of soil moisture variations is specified for the entire soil column, and soil texture. A single effective dynamical range of soil moisture with skill and are able to propagate soil moisture increments from the surface layer to the root zone.

b. Atmospheric forcing

The ISBA land surface model is forced with atmospheric data from short-range forecasts from EC’s NWP models. The atmospheric forcing variables required are short- and longwave radiation incident at the surface (ISW and ILW, respectively), air temperature, specific humidity, wind, surface pressure, and precipitation. Air temperature, specific humidity, and wind forcing are taken from the lowest vertical level in EC’s NWP model (~40 m), with the other variables representing surface values.

Precipitation is arguably the most important atmospheric forcing to the land surface model (Gottschalck et al. 2005; Liu et al. 2011; Maggioni et al. 2011, 2012a,b). The Canadian Precipitation Analysis (CaPA) product (Mahfouf et al. 2007) became operational at the MSC in April 2011, and this CaPA methodology is used to generate precipitation analyses to drive the land surface model predictions within CaLDAS. CaPA is constructed using the optimum interpolation (OI) procedure, which is a statistical interpolation technique combining information from a model first-guess precipitation field with precipitation gauge observations. The observational database available to CaPA consists of precipitation accumulations from the land surface synoptic reports (SYNOP) network, the aviation routine weather report (METAR), and the U.S. Standard Hydrometeorological Exchange Format (SHEF) surface observing network, along with various regional precipitation networks.

c. Geophysical land surface characteristics

Accurate specifications of the land surface characteristics such as topography, vegetation fraction, land-cover type, soil texture, and canopy density are important to provide an accurate model first guess. At EC’s MRD, considerable efforts have been made to obtain global high-resolution land surface ancillary databases from space-based remote sensing platforms to drive the external land surface model (e.g., Leroux et al. 2009; Mailhot et al. 2010). A new geophysical processor software has been developed to ingest these higher-resolution databases and incorporate revised calculations for subgrid-scale processes (e.g., roughness length and slope; Mailhot et al. 2010).

d. Observations

CaLDAS was developed with the goal to include observations from space-based remote sensing platforms
that offer near-real-time information with near-global coverage. These data sources would not replace the traditional surface-based observation networks but will be combined with them in an optimal way. At present, screen-level temperature and dewpoint temperatures are assimilated as proxy variables to correct the soil moisture state. These data offer the advantage of near-global coverage with a high temporal frequency. Surface snow depth observations are also used in CaLDAS for the analysis of snow depth.

For the specification of soil moisture, there are a number of remote sensing missions dedicated to collecting near-real-time data that have been discussed in the introduction. ISBA has been coupled to a surface microwave radiative transfer model for the assimilation of passive L-band TBs within CaLDAS. This study will present the results from data assimilation experiments in which synthetically derived L-band TBs are assimilated within CaLDAS for an improved specification of the soil moisture state.

e. Assimilation algorithm

As part of the early development of CaLDAS, a simplified two-dimensional variational assimilation scheme for the assimilation of passive microwave L-band data was developed and tested (Balsamo et al. 2006, 2007). For consistency with the model development directions at EC (Lavaysse et al. 2013; Houtekamer et al. 2014), the ensemble Kalman filter (EnKF) assimilation technique was chosen for CaLDAS. Numerous studies have demonstrated the positive impacts of experimental versions of the EnKF for selected case studies within the context of both real and idealized experiments (Reichle et al. 2002a,b; Margulis et al. 2002; Crow and Wood 2003; Crow 2003; Zhou et al. 2006; Kumar et al. 2008). A finite number of randomly generated model trajectories are used to approximate the necessary model error information within the EnKF (Reichle et al. 2002a,b; Reichle 2008).

The EnKF method works sequentially moving forward in time, alternating between a model forecast step and an update step when observations are present. The forecast step for a given ensemble member $k$ from time $t$ to $t + \Delta t$ can be written as

$$y_k^+(t + \Delta t) = F[y_k^+(t), u_k, w_k(t)],$$

where $y_k^+(t)$ is the updated model state at time $t$ for ensemble member $k$, $y_k^+(t + \Delta t)$ is the model forecast state at time $t + \Delta t$ for ensemble member $k$, $F$ represents the nonlinear ISBA land surface model, $u_k$ represents the time-invariant model parameters (e.g., soil texture and model topography), $u_k$ is the atmospheric forcing for ensemble member $k$, and $w_k(t)$ represents the model error. This model error term includes errors from model physics, poor specification of model parameters, and errors in model forcing data.

Throughout the forecast step, perturbations to the surface and atmospheric forcing variables are imposed on each ensemble member in order to generate an ensemble of model states at the next available observation time. At the time of observations, an update step is performed where the set of control variables are updated by optimally combining a prior estimate with an observation. The relationship between the observations and the state variables is given by the measurement operator and can be written as follows:

$$z = M[y(t)] + \omega,$$

where $z$ represents the observations, $M$ is the measurement operator, and $\omega$ represents the measurement errors. In this study, the observations are passive microwave L-band TBs, and the measurement operator consists of a microwave radiative transfer model that relates the soil moisture values to a top-of-the-atmosphere TB.

The update equation for the EnKF is based on the calculation of a Kalman gain matrix $K$, which determines the relative weights given to the prior estimate and the observations based on their respective uncertainties. The update (3) and Kalman gain (4) equations are as follows:

$$y_k^+(t + \Delta t) = y_k^+(t + \Delta t) + K_{t + \Delta t}(z + \omega - M[y_k^+(t + \Delta t)])$$

and

$$K_t = C_{zz}C_{zy}^{-1},$$

where $K_{t + \Delta t}$ is the Kalman gain matrix at time $t + \Delta t$, $C_{zz}$ is the cross covariance between the model state variables and the model first-guess TB observations, and $C_{zy}$ is the covariance of the model first-guess TB observations. The $\omega$ term represents a realization of the observation error that is added to the observations as they are treated as random variables (Burgers et al. 1998).

The so-called 1D-EnKF approach is used within CaLDAS, where the larger-scale assimilation is performed by means of a series of independent assimilation problems for each grid box. Horizontal error correlations between state variables from different grid boxes are neglected. This 1D-EnKF approach has the advantage of being straightforward to implement and computationally affordable (e.g., Reichle and Koster 2003; De Lannoy et al. 2010).
3. Synthetic truth observations

This section will describe the generation of truth L-band TBs and precipitation observations to be used in the synthetic data assimilation experiment. The ISBA model is integrated over time to produce a time series of truth land surface states, a so-called “nature run.” The domain is the North American landmass, and the ISBA model was integrated at 1-km resolution for the period from 31 March to 30 September 2009. The atmospheric forcing data used to drive the ISBA land surface model were derived from 6–18 h forecasts from the operational Regional Deterministic Prediction System (RDPS) model (Mailhot et al. 2006). A combination of different data sources was used to describe the geophysical characteristics of the land surface state that are summarized in Table 1. Throughout this study, atmospheric forcing fields are generated in 6-h periods and, as the RDPS was launched twice daily (0000 and 1200 UTC) in 2009, this necessitated combining 6–12- or 12–18-h forecasts to cover a 24-h period, ensuring that no atmospheric forcing variables are derived from forecast ranges longer than 18 h. A similar strategy was adopted for the 30–42 h forecast range (i.e., 30–36- or 36–42-h period) when discussing the data assimilation experiments. See Fig. 2a for more details.

Figure 2b is a schematic that describes the procedure used to generate the synthetic truth L-band TB observations to be assimilated. The Community Microwave Emission Modelling Platform (CMEM), developed at the European Centre for Medium-Range Weather Forecasts (ECMWF; de Rosnay et al. 2009; Drusch et al. 2009; Sabater et al. 2011; Parrens et al. 2014a), is used as a forward model to generate L-band TBs. Horizontal and vertical polarization L-band TBs are calculated for an incidence angle of 40° using the CMEM submodels shown in Table 2. Recent studies of CMEM-modeled TBs found the Kirdyashev et al. (1979) vegetation opacity model and the Wigneron et al. (2001) surface roughness model to outperform other models over a variety of surfaces in North and South America (Drusch et al. 2009) as well as over the African monsoon region (de Rosnay et al. 2009). De Rosnay et al. (2009) did note further the superior performance, at C band, of the Wang and Schmugge (1980) dielectric mixing model. These 1-km TBs are up-scaled to 40-km resolution, similar to the SMOS–SMAP passive radiometer resolution by taking a linear average of the TBs within each 40 km × 40 km pixel (Drusch et al. 1999; Reichle et al. 2001). This 40-km grid will serve as the EnKF integration grid for the subsequent data assimilation experiments and the 40-km L-band TBs will be referred to as the pixel-scale synthetic truth TBs.

Figures 3a and 3c illustrates this TB upscaling, showing TBs at the 1- and 40-km scales for the same time. An orbit simulator is then used to spatially and temporally distribute these TBs using orbital parameters from the SMAP mission. A 3-day satellite trajectory is calculated and repeated throughout the assimilation period. To increase the spatial coverage, swaths within 1 h of the central hour were assumed to be valid at the given central hour. The pixel-scale truth soil moisture values were calculated from the nature run in a similar manner as for the TBs, except that the values were weighted using the land-cover fraction in each 1-km grid cell. An illustration of this upscaling of the superficial soil moisture $w_f$ is provided in Figs. 3b and 3d. These soil moisture values are used to evaluate the skill of the assimilation and open-loop experiments described in section 5.

The upscaled TBs and soil moisture appear to be a smoother version of those at the 1-km scale, but there are differences, especially over regions with pronounced inland lakes. To avoid introducing systematic biases in the TB innovations, a land–water mask is applied in the EnKF to exclude pixels at the 40-km scale with greater than 10% water and is shown in Fig. 3e.

Six-hourly precipitation accumulations from the 6–18-h RDPS precipitation forecasts, used to force the nature run, were linearly interpolated to the location of the SYNOP and METAR stations for the given day and time in 2009. Typically, there are roughly 1000 such reports every 6 h. These 6-h “synthetic” precipitation observations are then assimilated using the CaPA OI methodology in the generation of the precipitation ensembles.

4. Precipitation ensemble sensitivity experiments

The generation of an ensemble of precipitation analysis to force an EnKF data assimilation system is
challenging (Clark and Slater 2006; Slater and Clark 2006; Pan and Wood 2009; Maggioni et al. 2011, 2012a,b) owing to the nonstationary and intermittent nature of precipitation. The precipitation analysis generated with the CaPA methodology is a function of several input parameters, including the precipitation gauge observations, the NWP first-guess precipitation, and the statistical parameters related to the OI technique. Within CaPA, the OI technique performs the analysis on precipitation innovations. To bring the probability distribution close to a normal distribution, a cubic root transformation is performed on the innovations and the OI analysis is performed in this transformed space. The observation error standard deviation $s_o$, model error standard deviation $s_b$, and the correlation length scale $L$ required by the OI procedure are calculated at each.

![Fig. 2. (a) Selection of 6-h RDPS atmospheric forcing periods for the nature run (above) and EnKF integrations (below). (b) Methodology for the generation of the synthetic truth TB observations to be assimilated in the synthetic experiments. Surface conditions refer to the ISBA prognostic land surface variables (Table 2; Bélair et al. 2003a).](image-url)

### Table 2. Physical parameterization submodels used within the CMEM.

<table>
<thead>
<tr>
<th>Soil</th>
<th>Dielectric mixing model</th>
<th>Wang and Schmugge (1980)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Effective temperature model</td>
<td>$T_{\text{eff}} = T_{\text{soil}}$</td>
</tr>
<tr>
<td></td>
<td>Smooth surface emissivity</td>
<td>Wigneron et al. (2001)</td>
</tr>
<tr>
<td></td>
<td>Surface roughness model</td>
<td>Kirdyashev et al. (1979)</td>
</tr>
<tr>
<td>Vegetation</td>
<td>Vegetation opacity model</td>
<td>Pulliainen et al. (1999)</td>
</tr>
<tr>
<td>Snow</td>
<td>Snow emission model</td>
<td>None</td>
</tr>
<tr>
<td>Atmosphere</td>
<td>Atmosphere optical thickness</td>
<td>None</td>
</tr>
</tbody>
</table>
analysis time based on a theoretical fit to the experimental semivariogram using a 30-day exponential time filter (Hollingsworth and Lönnberg 1986; Pannatier 1996).

Figure 4 illustrates the method used to generate the precipitation ensembles. For the precipitation ensembles, extended-range 30–42-h RDPS precipitation accumulations are used as a first guess, and the observations to be assimilated are the synthetic observations derived from interpolating the 6–18-h RDPS accumulations (see section 3 above). These longer-range first-guess precipitation accumulations (i.e., 30–42 h) are used to increase model uncertainties and allow for the assimilation of the 6–18-h synthetic precipitation observations. An ensemble size of 48 members is considered. Step one consists of

FIG. 3. Results of the L-band TB (K) and \( w_g \) (m\(^3\) m\(^{-2}\)) upsampling at 0000 UTC 4 Jul 2009. (a) TB at horizontal polarization and 40° incidence angle at 1 km and (b) \( w_g \) at 1 km. Upscaled (c) TB at 40 km and (d) \( w_g \) at 40 km. (e) The land-water assimilation mask.
generating a set of OI control parameters ($\sigma_p$, $\sigma_o$, and $L$) as a function of time, by running a control CaPA 6-h precipitation analysis with no perturbations to the first-guess precipitation or to the synthetic precipitation observations for the period January–September 2009. Commencing in January allows for enough time to generate a stable set of OI parameters.

Step two consists of generating the 48 precipitation ensemble members. Precipitation phasing errors are simulated by randomly perturbing the 30–42-h RDPS first-guess precipitation independently in both the latitude and longitude directions, assuming that the phasing error is constant throughout the accumulation period of six hours. A mean displacement error of 0 km with a variance of 75 km was used. These values were chosen after considerable testing and will vary depending on the particular application (i.e., grid resolution and forecast lead time). Errors in synthetic precipitation accumulations are simulated by adding Gaussian noise with zero mean and observation error standard deviation to the 6–18-h RDPS forecasts. These errors are added in transformed space. As shown in Fig. 4, each of the 48 ensemble members for a given time uses the same OI parameters ($\sigma_p$, $\sigma_o$, and $L$). The result is an ensemble of 6-h precipitation accumulations for the given time.

The precipitation ensembles are evaluated over the warm season, from June to August 2009, using a series of synoptic stations throughout North America. Table 3 gives the details on the four sensitivity experiments. In experiment PR1, first-guess precipitation phasing errors along with errors in precipitation observations are included. For experiment PR2, no errors are introduced to the precipitation observations, while the precipitation first-guess phasing errors remain as in PR1. Experiment PR3 ignores first-guess phasing errors, but retains the same perturbations to the precipitation observations as in PR1. Finally, in experiment ENRDPS, the 30–42-h RDPS first-guess precipitation is spatially perturbed as described above, but no synthetic precipitation observations are assimilated.

A set of 147 synoptic station locations were chosen for verification whose locations are shown in Fig. 5. These stations are separated by a distance of roughly 400 km to ensure a level of independence between stations (Hamill and Colucci 1997). For the verification, the CaPA leave-one-out analyses are used, where the precipitation at the given station is estimated without using the synthetic precipitation at that given station. Three summary metrics are shown in Table 3. To quantify the ability of the precipitation ensembles to capture the spatial precipitation gradient, the Spearman rank correlation between the observed total precipitation and the corresponding ensemble mean is shown, while the spread of the precipitation in the ensembles is summarized by the
uncertainty ratio (UR) and exceedance ratio (ER) scores (Hossain and Anagnostou 2005; Moradkhani et al. 2006; Maggioni et al. 2011).

The UR is defined as follows:

\[
\text{UR} = \left( \frac{(\sum P)_{75} - (\sum P)_{25}}{\sum P_{\text{obs}}} \right) \times 100\%,
\]

where \((\sum P)_{75}\) and \((\sum P)_{25}\) represent the 75th and 25th percentile of total accumulated precipitation from the precipitation ensemble and \(\sum P_{\text{obs}}\) refers to the observed total accumulated precipitation. This UR metric is a measure of the variability or dispersion of the ensemble compared to the magnitude of the observed variable. The average UR score calculated over all of the stations is shown in Table 3. The ER score is defined as the number of times the “observed” total accumulated precipitation falls outside of the range of the minimum and maximum values of the precipitation ensemble divided by the total number of stations. This ER score is bounded between 0 and 1. As the ER score approaches 0, there are fewer observed accumulated precipitation values that fall outside the ensemble envelope, which is desirable. The UR and ER scores tend to behave conversely (Moradkhani et al. 2006).

The inclusion of precipitation observations in experiments PR1, PR2, and PR3 resulted in superior Spearman rank correlations with observations (0.97 versus 0.92) when compared to only spatially perturbing the first-guess RDPS precipitation (ENRDPS). The UR score for PR3 is significantly lower than for the other experiments, indicating reduced ensemble spread and uncertainty. The largest UR score is found for ENRDPS, as there are no observations assimilated to constrain the precipitation analyses. For PR3 the ER score is high, very close to the maximum of 1, indicative that far too often the observed accumulated precipitation falls outside the ensemble envelope. It is interesting to note that perturbing the precipitation observations (PR1 versus PR2) has a relatively smaller influence on increasing (decreasing) the UR (ER). Perturbing the first-guess precipitation is shown to be more effective in increasing (decreasing) the UR (ER) scores (PR1 versus PR3).

Considering these same 147 station locations, ranked histograms (Hamill 2001) of 6-h precipitation accumulations for these same experiments are shown in Fig. 6.

---

Table 3. Configuration and summary scores of the precipitation ensemble experiments.

<table>
<thead>
<tr>
<th>Precipitation ensemble experiment</th>
<th>Configuration</th>
<th>Summary scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>First-guess precipitation perturbations</td>
<td>Synoptic precipitation observation perturbations</td>
</tr>
<tr>
<td>ENRDPS</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>PR1</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>PR2</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>PR3</td>
<td>No</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Fig. 5. Location of the 147 synoptic stations used for the precipitation ensemble verification.
For these ranked histograms, only those cases where the observed 6-h precipitation is greater than or equal to 0.2 mm are considered. Each of the precipitation ensembles displays the characteristic U shape indicative of underdispersion. Clearly evident is the underdispersive nature of experiment PR3, where more than half of the 6-h observed precipitation is found in either the top or bottom rank. For a reliable ensemble system, the ensemble members and the true verifying state should be considered as random samples from the same probability distribution (Hamill 2001; Candille and Talagrand 2005, 2008). This implies a flat histogram, where for an
an $n$-member ensemble, with the predicted values ranked from lowest to highest, the verifying observations are equally likely to occur in each of the $n + 1$ possible ranks.

Following Candille and Talagrand (2008), a quantitative measure of reliability can be calculated for a given ranked histogram based on the deviation from flatness. A diagnostic $\Delta$ for an $n$-ensemble member is calculated as follows:

$$\Delta = \sum_{i=1}^{n+1} \left( s_i - \frac{m}{n+1} \right)^2 .$$

In (6), $s_i$ refers to the number of realizations for the $i$th interval of the ranked histogram and $m$ denotes the total number of realizations over all $n + 1$ ranked intervals. If the ensemble system is reliable, the expected value of $\Delta$ is

$$\Delta_o = \frac{mn}{n+1} .$$

The ratio of the actual flatness $\Delta$ to its expected value $\Delta_o$, labeled $\delta$, can be considered a measure of the reliability of the ensemble system (Candille and Talagrand 2008). A value of $\delta$ significantly larger than 1 indicates an ensemble that lacks reliability. The value of $\delta$ is shown for each of the four precipitation ensembles in Fig. 6.

The $\delta$ values are all significantly larger than 1, confirming the underdispersive nature of the ensembles. Nonetheless, the inclusion of first-guess precipitation spatial perturbations acts to reduce the $\delta$ scores by a factor of roughly 10 (PR3 versus PR1; 2623.70 versus 263.39). Perturbing the precipitation observations reduces $\delta$ much less (379.32 versus 263.39; PR2 versus PR1) in comparison to perturbing the first-guess precipitation. Ensemble PR1 possesses the best (lowest) $\delta$ score of 263.39.

Two RDPS forecast ranges are used: the shorter 6–18-h range is used to generate the synthetic precipitation observations, but the 30–42-h RDPS precipitation is the same for each member, resulting in the largest population in the top rank. Observation errors can be added to the rank histogram, which could potentially improve the flatness (Hamill 2001; Candille and Talagrand 2008). This was not attempted in this study.

5. Synthetic EnKF experiment

Previous studies have examined the utility of assimilating passive L-band TBs for the analysis of soil moisture within the context of synthetic and real-data experiments (Reichle et al. 2002a; Crow and Wood 2003; Crow 2003; Seuffert et al. 2004; Dunne and Entekhabi 2005, 2006; Balsamo et al. 2006, 2007). Modeling and assimilation studies utilizing different versions of the two-layer force–restore version of ISBA to infer the root-zone soil moisture from surface soil moisture observations are reported on in previous studies (e.g., Calvet et al. 1998; Wigneron et al. 1999; Calvet and Noilhan 2000; Draper et al. 2009, 2011; Mahfouf 2010; Parrens et al. 2014b). These ISBA studies found that skillful increments to the root-zone soil moisture could be gained from the assimilation of surface soil moisture.

a. Experimental setup

A continental synthetic EnKF experiment is set up and the resulting analyzed soil moisture for the superficial $w_g$ and root-zone $w_r$ layers are compared against the synthetically derived “truth” soil moisture upscaled from the nature run (see section 3). The EnKF grid has a horizontal resolution of 40 km and the pixel-scale synthetic truth TBs at horizontal polarization and 40° incidence angle derived from the nature run are assimilated. Longer-range 30–42-h RDPS atmospheric forcing is used in the EnKF experiment (see Fig. 2a). The forecast step in the EnKF is 6 h; however, TB data are only available for the 0000 and 1200 UTC update times over North America.

Assimilation is restricted to grid points where the fraction of land exceeds 90%. If snow, frozen soil, or precipitation is present in either the nature run or in one of the EnKF members, then no assimilation is performed. Owing to the reduced information content of TBs over heavily vegetated and forested regions (Entekhabi et al. 2010b), the domain of evaluation consists of the short-grass and agriculture regions of central North America, which represents 28.7% of the unfrozen and snow-free area. The initial conditions were taken from a 240-h (10 day) forecast valid at 0000 UTC 1 June 2009 from EC’s Global Deterministic Prediction System (GDPS) model (Béclair et al. 2009).
Details on the perturbations applied to the atmospheric forcing variables along with the introduction of model errors to \( w_g \) and \( w_2 \) are given in Table 4. Air temperature, precipitation, and radiative forcing are perturbed in this study based on their direct and indirect impacts on soil moisture. The statistics of the differences between the 6–18- and the 30–42-h RDPS forecasts were used to specify the distribution parameters shown in Table 4. Statistics were calculated over land points for the period June–September 2009 for air temperature and ILW. For ISW, only daytime periods were considered.

Figure 7 shows histogram plots of the differences for air temperature (Fig. 7a) and ILW (Fig. 7b) along with the sample means and standard deviations. Errors in both air temperature forcing and ILW are essentially unbiased, and perturbations were modeled as additive Gaussian. For ISW the ratios of the 6–18- to 30–42-h values are shown in Fig. 7c, and these ratios are approximated with a lognormal distribution as in previous studies (e.g., Kumar et al. 2008, 2012; De Lannoy et al. 2010; Forman and Margulis 2010). Ratios indicate a bias in ISW with higher values for shorter-range RDPS forecasts. Perturbations to air temperature and radiative forcing are uncorrelated in space and time as this is a 1D-EnKF study.

To quantify the impact of the atmospheric forcing uncertainties on \( w_g \) and TB ensemble spreads, a series of data assimilation experiments are performed and evaluated against the pixel-scale truth soil moistures and TBs. Along with the UR and ER scores, the normalized innovations are examined by means of the reduced centered random variable (RCRV), defined as

\[
\text{RCRV} = \frac{Z - M(Y_k)}{\sqrt{C_{zz} + C_{v}}}. \tag{8}
\]

In (8), the TB innovations are normalized by the forecast error variance (sum of model and observation errors). An analysis of the statistics of the innovations provides important information on the correct behavior and optimality of the EnKF (Reichle and Koster 2002; Reichle et al. 2002b; Crow 2003; Kumar et al. 2008, 2012). Calculated over all realizations of the ensemble system, a perfectly reliable system should possess an RCRV mean and standard deviation of 0 and 1, respectively. The mean RCRV provides information on the bias while an RCRV standard deviation greater (less) than 1 is indicative of too small (large) a forecast error variance (sum of model and observation errors) and underdispersion (overdispersion; Candille et al. 2007; Candille 2009).

All ensemble members start with the same initial conditions and use the PR1 precipitation ensemble methodology. The only source of ensemble spread is derived from the atmospheric forcing perturbations, and the TB observation error standard deviation is set to 3 K. Scores are calculated over the short-grass and agricultural regions at times and locations when TB observations are available. Note that the UR scores differ from that in section 4 and are calculated using the upper and lower bounds of the ensemble as follows:

\[
\text{UR} = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{x^i_{\text{upper}} - x^i_{\text{lower}}}{x^i_{\text{true}}} \right) \times 100\%. \tag{9}
\]

In (9), \( N \) is the total number of times and \( x^i \) denotes the truth.

Examining the experiments with 48 members, precipitation perturbations alone generate a \( w_g \) UR score of greater than 25% and an ER score of just under 70% (Table 5). The inclusion of ISW perturbations is the most effective way to increase (decrease) the UR (ER) scores for \( w_g \). With the ISW perturbations, the ensemble has wider uncertainty limits, accounting on average for over 31% of the magnitude of the truth \( w_g \). The decrease in ER implies that the ensemble envelope better

<table>
<thead>
<tr>
<th>Forcing variable</th>
<th>Perturbation type</th>
<th>Distribution parameters</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air temperature at 40 m</td>
<td>Additive</td>
<td>( \mu = 0 ) K</td>
<td>30–42-h RDPS</td>
</tr>
<tr>
<td></td>
<td>Gaussian</td>
<td>( \sigma = 1 ) K</td>
<td>30–42-h RDPS</td>
</tr>
<tr>
<td>ISW</td>
<td>Multiplicative</td>
<td>( \mu = 1 )</td>
<td>30–42-h RDPS</td>
</tr>
<tr>
<td></td>
<td>Lognormal</td>
<td>( \sigma = 0.5 )</td>
<td>30–42-h RDPS</td>
</tr>
<tr>
<td>ILW</td>
<td>Additive</td>
<td>( \mu = 0 ) W m(^{-2})</td>
<td>30–42-h RDPS</td>
</tr>
<tr>
<td></td>
<td>Gaussian</td>
<td>( \sigma = 5 ) W m(^{-2})</td>
<td>30–42-h RDPS</td>
</tr>
<tr>
<td>Precipitation</td>
<td>Spatial phasing</td>
<td>Ensemble PR1 (see section 4)</td>
<td>6–18-h RDPS (observations);</td>
</tr>
<tr>
<td></td>
<td>plus observation</td>
<td></td>
<td>30–42-h RDPS (first guess)</td>
</tr>
<tr>
<td>Model error</td>
<td>( w_g = w_g(1 + \alpha) )</td>
<td>( \alpha ) (Gaussian); ( \mu = 0, \sigma = 0.016 )</td>
<td></td>
</tr>
<tr>
<td></td>
<td>( w_2 = w_2(1 + \omega) )</td>
<td>( \omega ) (Gaussian); ( \mu = 0, \sigma = 0.008 )</td>
<td></td>
</tr>
<tr>
<td>TB</td>
<td></td>
<td>( \sigma_o = 3 ) K</td>
<td></td>
</tr>
</tbody>
</table>
encapsulates the truth \( w_g \). Uncertainties in ILW and air temperature forcing were found to have very small influences on the \( w_g \) UR and ER scores.

Precipitation perturbations alone generate an RCRV standard deviation of 1.20 indicative of underdispersion. The addition of ISW perturbations brings the RCRV standard deviation to 1.13 and, including all atmospheric perturbations, results in a value of 1.11. All experiments possess a positive mean RCRV indicative of a wet bias, and the addition of atmospheric forcing uncertainties has a negligible impact on reducing this bias. The atmospheric uncertainties are unable to generate sufficient spread in \( w_g \), and the initial wet bias (Fig. 8) is not reduced significantly by the forcing uncertainties alone. In Table 4 the standard deviation for ILW perturbations was set to 5 W m\(^{-2}\), which is lower than shown in Fig. 7b. An additional experiment including all atmospheric perturbations and setting the ILW perturbation standard deviation to 20 W m\(^{-2}\) shows the impacts on the UR, ER, and

**Fig. 7.** Histograms comparing the 6–18- and 30–42-h RDPS atmospheric forcing for the period June–September 2009. (a) Differences in air temperature forcing \((6–18 h) - (30–42 h)\), (b) differences in ILW \((6–18 h) - (30–42 h)\), and (c) ratio of ISW \((6–18 h)/(30–42 h)\). The number of data points considered along with the mean and std dev are shown.
RCRV scores to be very small, and the value of $5 W m^{-2}$ is kept.

Several studies have noted an attenuation of the soil moisture variability from the rainfall to soil moisture transition limiting the added skill from increasing ensemble size (Hossain and Anagnostou 2005; Maggioni et al. 2011, 2012a,b). Four additional experiments were performed with ensemble sizes of 6, 12, 24, and 36, with the only source of uncertainty being precipitation (see Table 5). Beyond 24 members very modest improvements are seen in the UR, ER, and RCRV scores, consistent with previous studies (e.g., Maggioni et al. 2012a).

Owing to the underdispersive ensemble with only atmospheric forcing uncertainties, model errors are added directly to $w_g$ and $w_2$ through multiplicative Gaussian noise (Table 4). These errors are added at the beginning of each 6-h period, and an additional assimilation experiment is performed including all atmospheric uncertainties and model errors, labeled EPRTB48. The standard deviation of the RCRV is reduced to 0.99 while the mean RCRV is reduced by roughly 50% to 0.20.

### b. Soil moisture results

Along with experiment EPRTB48, two additional experiments, EPR48 and an open-loop (OL) experiment, are performed. The OL experiment consists of a single run using 30–42-h RDPS atmospheric and surface forcing and is used as a reference to quantify the error reductions, while experiment EPR48 is identical to experiment EPRTB48, except that no TB information is assimilated and no model errors are added to the soil moisture reservoirs after the initial time. Figure 8 shows the domain-averaged bias and standard deviation of the errors (unbiased RMSE; Entekhabi et al. 2010a) for both $w_g$ and $w_2$ for all three experiments. These spatial scores are calculated across all pixels at one instant in time. Note that these assimilation experiments extend from June to the end of September 2009.

The time-mean $w_g$ standard deviation (Fig. 8b) in EPRTB48 is reduced by roughly 40% when compared to OL. The initial wet bias converges toward a stable value near 0 (Fig. 8a), an overall reduction of 69.2% in the time-mean $w_g$ bias when compared to the OL. The results from EPR48 indicate that much of the improvements in spatial standard deviation scores are derived from the use of the PR1 ensemble precipitation forcing as results from EPRTB48 and EPR48 are very close to each other. Quantitatively, the mean $w_g$ standard deviation is reduced by 5.4% with the inclusion of the TB observations (EPRTB48 versus EPR48), when compared to the OL. The overall reduction in the $w_g$ bias is larger (47.3%) with the inclusion of TB observations.

A slower reduction of the spatial metric scores is seen for the longer time scale root-zone layer. The $w_2$ standard deviation (Fig. 8d) scores at the end of the 4-month period are reduced by 49.7% in EPRTB48, when compared to the OL, while the $w_2$ bias is essentially eliminated (Fig. 8c). The assimilation of TBs is more effective in correcting the initial errors in $w_2$ when compared to the use of the precipitation ensembles as forcing (EPRTB48 versus EPR48). Quantitatively, the inclusion of TB observations (EPRTB48) acts to reduce the $w_2$ standard deviation scores at the end of the 4-month period by 35.9%, when compared to EPR48.

Considering the soil moisture analyses at times and locations where observations are assimilated, so-called EnKF update times, the error reduction for both $w_g$ and $w_2$ in EPRTB48 is greater when compared to EPR48. The mean $w_g$ error reductions in EPRTB48 when

### Table 5. Values of UR, ER, and RCRV for the various EnKF data assimilation experiments during June–August 2009. TEMP refers to air temperature forcing perturbations. UR and ER are for $w_g$ (%), while the RCRV score is for TB (dimensionless).

<table>
<thead>
<tr>
<th>Assimilation experiment</th>
<th>No. of members</th>
<th>UR ($w_g$) (%)</th>
<th>ER ($w_g$) (%)</th>
<th>Mean</th>
<th>Std dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR1</td>
<td>48</td>
<td>26.3</td>
<td>68.9</td>
<td>0.39</td>
<td>1.20</td>
</tr>
<tr>
<td>PR1 + TEMP</td>
<td>48</td>
<td>26.8</td>
<td>67.7</td>
<td>0.41</td>
<td>1.19</td>
</tr>
<tr>
<td>PR1 + ILW (σ = 5 W m⁻²)</td>
<td>48</td>
<td>26.3</td>
<td>68.9</td>
<td>0.39</td>
<td>1.20</td>
</tr>
<tr>
<td>PR1 + ISW</td>
<td>48</td>
<td>31.7</td>
<td>55.7</td>
<td>0.37</td>
<td>1.13</td>
</tr>
<tr>
<td>PR1 + TEMP + ISW + ILW</td>
<td>48</td>
<td>32.0</td>
<td>55.6</td>
<td>0.39</td>
<td>1.12</td>
</tr>
<tr>
<td>PR1 + TEMP + ISW + ILW</td>
<td>48</td>
<td>32.1</td>
<td>55.5</td>
<td>0.38</td>
<td>1.11</td>
</tr>
<tr>
<td>PR1</td>
<td>6</td>
<td>15.2</td>
<td>80.9</td>
<td>0.39</td>
<td>1.31</td>
</tr>
<tr>
<td>PR1</td>
<td>12</td>
<td>19.3</td>
<td>76.2</td>
<td>0.39</td>
<td>1.26</td>
</tr>
<tr>
<td>PR1</td>
<td>24</td>
<td>23.0</td>
<td>72.1</td>
<td>0.39</td>
<td>1.22</td>
</tr>
<tr>
<td>PR1</td>
<td>36</td>
<td>25.0</td>
<td>70.2</td>
<td>0.39</td>
<td>1.21</td>
</tr>
</tbody>
</table>
compared to EPR48 for the bias and standard deviation are 53.8% and 22.3% at the EnKF update times. These scores can be compared with the improvements in bias and standard deviation at all times of 47.3% and 5.4%, respectively. The inclusion of TB observations is more effective in reducing $w_g$ errors at the update times, but owing to the short memory of $w_g$, some of this update information is lost for the subsequent nonupdate times. The greater reduction of $w_g$ error scores between EPR48 and EPRTB48 considering all times points to the importance of the higher temporal frequency nature of the precipitation observations, when compared with TBs, over North America, in reducing $w_g$ errors (Crow 2003; Liu et al. 2011).

For $w_2$ the improvement in scores at the EnKF update times is not as pronounced as for $w_g$; the mean $w_2$ bias and standard deviation improvements over EPR48 are 42.8% and 29.7%, which can be compared with those at all times of 41.7% and 24.3%. The skillful $w_2$ corrections at the EnKF update times, while smaller, are sustained through the subsequent nonupdate times, as the root-zone layer is less sensitive to atmospheric forcing errors.
as compared to $w_g$. Histograms of the temporal correlation for $w_g$ between the given experiment and the upscaled truth $w_g$ for the period June–September 2009 are shown in Fig. 9. A significant improvement in domain-averaged correlations is seen for both EPR48 and EPRTB48 when compared with the OL. Mean correlations increase from 0.64 to 0.83 with the inclusion of the PR1 ensemble precipitation forcing, with a further
increase to 0.85 with the inclusion of TB observations (Fig. 9c). Both of these increases in mean correlation are statistically significant at the 99% level.

Figure 10 shows the time evolutions of the RCRV mean and standard deviation calculated over the prescribed 3-day repeat trajectory of the simulated satellite orbit (see section 3). The RCRV mean and standard deviation values of 0.18 and 0.96 deviate from the optimum values of 0 and 1, respectively. These scores differ from those in section 5a as they include the month of September. The initial wet bias (Figs. 8a,c) is gradually reduced over time to values that fluctuate within ±0.2 around 0. The RCRV standard deviations depict values that are above 1 for the first month, but these values are reduced throughout the assimilation period to a situation where the ensemble is slightly overdispersive in September. RCRV standard deviations are consistently above 0.7, indicating that the EnKF maintains an adequate forecast state spread through time.

c. Sensitivity to ensemble size

Experiments EPRTB6, EPRTB12, EPRTB24, and EPRTB36 are identical to EPRTB48, except that the ensemble sizes are reduced to 6, 12, 24, and 36 members, respectively, and Fig. 11 shows the time-mean $w_g$ and $w_2$ bias and standard deviation scores at the EnKF update times. Both scores show a decreasing estimation error with increasing ensemble size. The mean $w_g$ bias and standard deviation decrease by 10.3% and 5.1%, respectively, when increasing the number of members from 6 to 48, compared to the OL. For $w_2$, the mean bias and standard deviation decrease by 14.8% and 10%,
respectively, with respect to the OL. The $w_2$ root-mean-square error (RMSE) reduction (not shown) is delayed by roughly 2 weeks when reducing the numbers of members from 48 to 6.

The added benefits from increasing the number of ensemble members is smaller compared to the impacts of assimilating TBs (EPR48 versus EPRTB48) or the use of precipitation ensembles as forcing (OL versus EPR48). Using only six members, but assimilating TBs (EPRTB6), when compared to EPR48, shows mean $w_g$ ($w_2$) bias and standard deviation reductions of 43.5% (28%) and 17.2% (19.7%), respectively, at the EnKF update times. The modest improvements with increasing ensemble size beyond 12 members are consistent with previous 1D-EnKF studies (Reichle and Koster 2003; Kumar et al. 2008), and larger gains would be expected if horizontal error correlations in the background field were included (Reichle and Koster 2003).

6. Summary

This study has outlined the development of CaLDAS, which represents important advances from the current operational land surface data assimilation system at Environment Canada. CaLDAS is structured around an offline land surface model, capable of running at very high resolution, which is forced by EC’s best estimates of atmospheric and surface forcing variables along with
more accurate geophysical databases. The more sophisticated and flexible EnKF assimilation strategy has been implemented to replace the OI-based system.

A series of precipitation ensemble sensitivity experiments has shown the importance of perturbing both the NWP first-guess precipitation and the precipitation observations in generating a precipitation ensemble with higher (lower) uncertainty ratio (exceedance ratio) scores. Perturbing the first-guess precipitation was most effective in reducing the ranked histogram flatness parameter δ and giving rise to a more reliable ensemble. Adding perturbations only to the precipitation observations generates an ensemble that lacks sufficient spread and is highly underdispersive.

The use of precipitation ensembles, generated by assimilating precipitation observations, as forcing combined with the assimilation of brightness temperatures TBs resulted in more rapid and greater overall reductions in the spatial superficial soil moisture \( w_8 \) and root-zone soil moisture \( w_2 \) bias and standard deviation scores when compared to an open-loop experiment. The significant error reductions in \( w_2 \) with the assimilation of TB observations is a very encouraging result, as the initialization of the root-zone soil moisture is of particular importance for NWP (e.g., Draper et al. 2009; Mahfouf 2010). The combination of the L-band TBs and the use of precipitation ensembles was the most effective in correcting \( w_2 \) estimation errors (Crow 2003; Liu et al. 2011).

Innovation diagnostics show that the EnKF is able to rectifying 2 estimation errors (Crow 2003; Liu et al. 2011). Innovation diagnostics show that the EnKF is able to maintain a sufficient forecast error spread through time. The combination of the L-band TBs and the use of TB observations has shown the importance of perturbing both the NWP first-guess precipitation and the precipitation observations in generating a precipitation ensemble with more accurate geophysical databases. The more sophisticated and flexible EnKF assimilation strategy has been implemented to replace the OI-based system.

**Acknowledgments.** The authors would like to acknowledge the financial support of the Canadian Space Agency through the Government Related Initiatives Program. Discussions with Dr. Houtekamer and Dr. Fortin were helpful throughout this study. The comments and suggestions from two anonymous reviewers helped to improve the manuscript.

**REFERENCES**


—, D. Entekhabi, and D. B. McLaughlin, 2001: Downscaling of radiative brightness measurements for soil moisture estimation: A four-dimensional variational data assimilation


