Advances in Land Data Assimilation at Environment Canada

Stéphane Bélair, Marco L. Carrera, Bernard Bilodeau,
Sheena J. Solomon, Natacha B. Bernier, and Sylvie Leroyer
Meteorological Research Division, Environment Canada

Linying Tong
Canadian Meteorological Centre, Environment Canada

Chris Derksen, Libo Wang, and Douglas Chan
Climate Research Division, Environment Canada

1. Introduction

Recognizing the increasing importance of land surface processes in numerical weather prediction (NWP), and more generally in numerical environmental prediction (NEP), Environment Canada (EC) has been mobilizing to develop its new generation of land surface modeling and assimilation systems. In a collaboration that involves the Meteorological Research Division (MRD), the Climate Research Division (CRD), and the Canadian Meteorological Centre (CMC), these systems are being prepared for operational implementation at CMC.

As part of this effort, a new Canadian Land Data Assimilation System (CaLDAS) has been coded and is currently being tested. In contrast with the land surface data assimilation systems currently used operationally at CMC (mostly based on the assimilation of surface data with an optimal interpolation technique), CaLDAS is built around an external modeling system, with special emphasis on the assimilation of space-based remote sensing data for soil moisture, terrestrial snow, and vegetation (although surface data still plays an important role in CaLDAS). More sophisticated assimilation methods are applied in CaLDAS for the optimal combination of observations and first guess, based on a simple variational technique developed by Balsamo *et al*. (2006, 2007) and an Ensemble Kalman Filter (EnKF, e.g., Reichle *et al*. 2002).

In the context of this development work, several research projects are currently underway in search of the best possible CaLDAS configuration for NWP and other environmental applications (e.g., agricultural risk management, hydrological forecasting):

- High-resolution land surface modeling (first guess modeling);
- Modeling of first guess uncertainty;
- Assimilation of soil moisture with screen-level data and space-based remote sensing data;
- Optimization of the use of screen-level observations (or analyses) in land data assimilation;

- Comparison of several techniques for land surface data assimilation (simple variational, Extended Kalman Filter, or EnKF);
- Assimilation of terrestrial snow with high-resolution optical remote sensing data (for snow coverage area) and with microwave passive data (for snow water equivalent – SWE);
- Improved characteristics of vegetation (coverage fraction and leaf area index) using modeling outputs from the community ecosystem modeling system Biome-BGC;
- Evaluation of the impact of improved land surface initial conditions and of land surface modeling on NWP and more generally on NEP.

For brevity's sake, the following discussion will focus upon the projects which relate to high-resolution land surface modeling and the assimilation of soil moisture.

2. High-resolution land surface modeling

Rightly so, a lot of emphasis is put in land data assimilation systems on the optimal combination of first guess and observations. Arguably, the quality of land surface analyses (for soil moisture, snow coverage area, snow water equivalent, …) depends on this combination. But it also depends (in an even more direct manner) on the absolute values of the errors associated with both the first guess and the observations. For instance, in regions where observation errors are large, information for the land data analyses will mostly come from the land surface modeling system (i.e., first guess), with its associated errors.

In an effort to reduce first guess errors, several projects have been completed at EC to refine land surface processes using an external land surface modeling system. Driven by EC's current operational modeling and analysis outputs, this external system is based on the same physics present in EC's atmospheric model GEM (Global Environmental Multiscale), but is integrated at much higher horizontal resolution.

This system has been applied for urban meteorology (Lemonsu *et al*. 2010; Leroyer *et al*. 2010) and for hydrological modeling (Carrera *et al*. 2010). A first experimental implementation of the external land surface system has been completed with the goal to improve and refine surface and near-surface numerical predictions for the 2010 Vancouver Winter Olympic and Paralympic Games (Bernier *et al*. 2010a, 2010b).

Typically, the external land surface system is integrated with grid sizes on the order of 100m to 1km. Horizontal refinement of land surface predictions is obtained from detailed information on land surface characteristics (land use / land cover, surface albedo, land / water fractional coverage, …) and from downscaling of the atmospheric forcing (air temperature and humidity, downwelling atmospheric radiation, precipitation phase, …). For example, relatively small-scale urban heat islands are represented with this system because of the high-resolution information available over cities (fraction of built-up areas, urban geometry, thermal characteristics, surfaces albedo), in spite of the low-resolution atmospheric forcing (not shown).

For the Olympic Games application, the increase in horizontal resolution mostly comes from downscaling of the atmospheric forcing, simply obtained in this case by adjusting air temperature at a reference level approximatively 40m above ground level (forcing level) using hydrostatic corrections that are function of the height differences between the low-resolution forcing model and the high-resolution land surface external system. These air temperature corrections are also used to adjust the phase of precipitation at the surface.

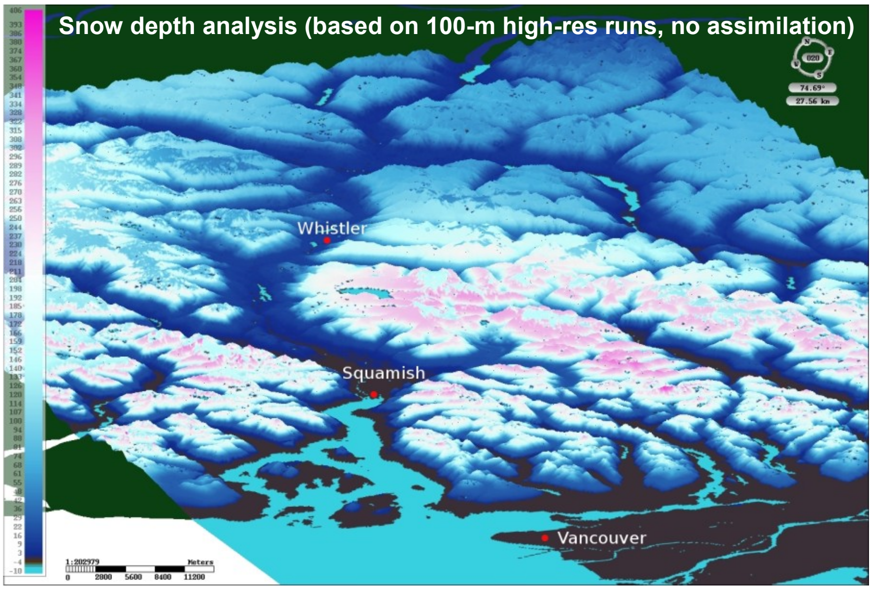


Figure 1: Snow depth analysis produced by an open-loop integration of a 100-m external land surface system. The integration was done over the Vancouver Olympic alpine region (north of the city). The color bar on the left indicate the values in cm of snow depth.

An example of snow depth fields produced following this approach is shown in Fig. 1. For this case, a 100-m version of the external system was integrated in an open-loop manner (i.e., without assimilation of any surface data) for a complete winter. Atmospheric forcing were provided by short-range forecasts of EC's regional 15-km operational system. Results show that the external system is able to reproduce realistic spatial variability of the snow fields (in spite of the relatively low-resolution forcing). Objective evaluation against observational surface networks (not shown) indicate important improvements in the numerical representation of snow depth and of near-surface air characteristics.

3. Assimilation of soil moisture

A first version of CaLDAS has recently been completed and is currently being tested with synthetic experimental configurations. This version, coded with the Supervisor Monitor Scheduler (SMS), includes both simple variational and EnKF approaches.

This version of the new land data assimilation system also includes new and original approaches for the modeling of the first guess error covariance matrix **B**. As can be found in the literature, there are several techniques that can be used to assimilate land data, or combine them with background information or a first guess. In the end, the results obtained with these assimilation approaches may very well depend as much on details of their implementation as on the specific approaches themselves. For instance, the specification of the error covariances that are assigned or calculated for both the first guess and for the observations is of crucial importance because it determines the weights for the linear combination of these two components of the analysis solution.

Uncertainty in the background information comes from several sources, including initial conditions for the land surface state variables, land surface characteristics provided by ancillary data (e.g., albedo, soil texture, vegetation characteristics), atmospheric forcing (e.g., precipitation, downwelling atmospheric radiation, air temperature, humidity, winds), and land surface modeling. The approach used in the EnKF version of CaLDAS is to perturb initial conditions, the atmospheric forcing, and land surface characteristics, in order to obtain an ensemble of first guess runs with a spread that is realistic enough to accurately represent the uncertainty in background information used for the assimilation.

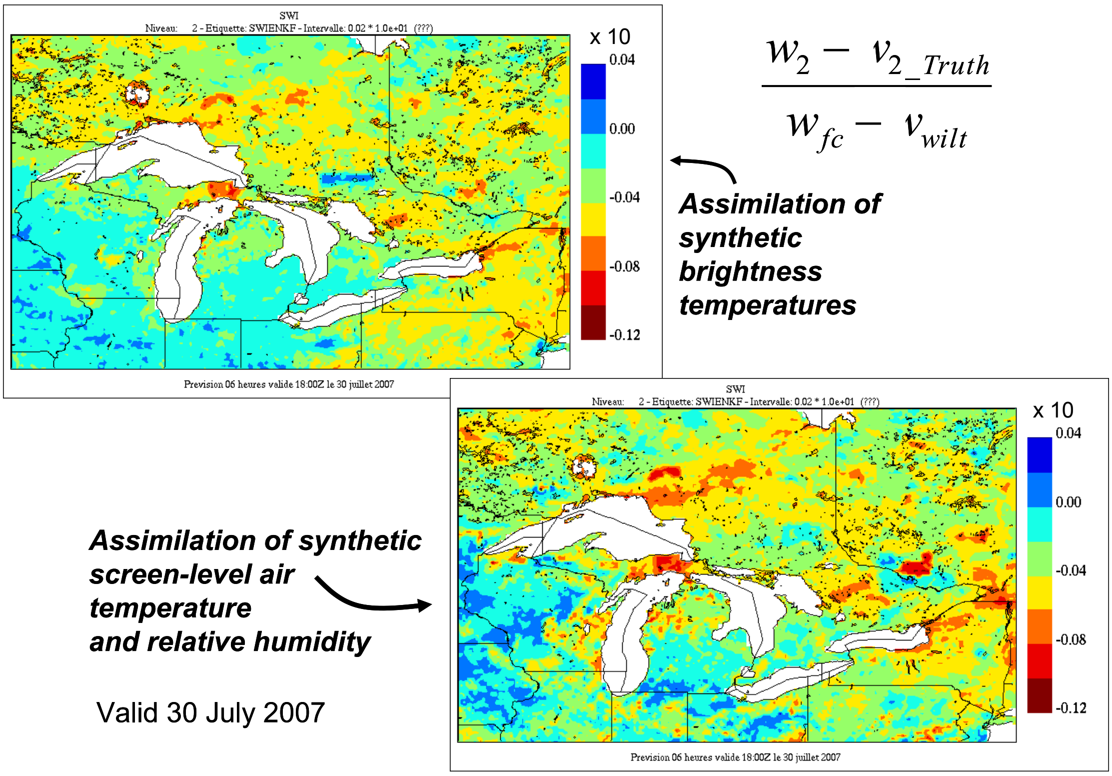
Being part of a national weather centre, and having a certain control and influence on the generation of the products used as atmospheric forcing in the land data assimilation system, it is possible to perturb at the source the first-guess fields and observations that are used as atmospheric forcing. This has been done for the precipitation analysis, which comes from the Canadian Precipitation Analysis (CaPA, Mahfouf *et al*. 2007), which in this situation produces an ensemble of precipitation based on perturbed first-guess fields and observations. This technique could also be done for screen-level forcing (if used as forcing, but this has not been done or tested yet).

For land surface characteristics, spatially and temporally correlated perturbations are generated globally with Spherical Harmonics development in which the coefficients are specified using simple Markov chains (Li *et al*. 2008). This technique is currently being tested in the context of assimilation of terrestrial snow, but has not been applied in the tests described below.

The first tests with this system were performed in the context of a relatively simple idealized experiment, in which synthetic observations were generated from an open loop integration. The main objective of these tests is to demonstrate that the mechanics of both the simple variational and the EnKF versions of CaLDAS works properly. Integrations were done over a 10-km grid-size domain (with 200x120 points) covering the Great Lakes region. This domain was selected to facilitate interactions with a project of the International Joint Commission that regulates the hydrology of the Great Lakes basin. Synthetic observations of L-band brightness temperatures (H and L polarizations) and of screen-level air temperature and relative humidity were produced every 6h with a one-month open loop integration on the same grid, initialized on 1 July 2007 with very wet conditions (both reservoirs of the land surface scheme ISBA set at field capacity). Perturbations of these observations were assimilated in the EnKF version of CaLDAS (with a 6 hour assimilation step), starting on 1 July 2007 with very dry conditions (set at the wilting point). Ten members are used for the EnKF, with the same surface characteristics and perturbed realizations of the same forcing as those used in the production of the truth. Several experiments were conducted following this protocol. Two of them are discussed here: one in which only L-band brightness temperatures (H and L polarizations) are assimilated, and another in which only screen-level observations are assimilated.

The convergence of the two assimilation experiments was examined using an index based on the difference between the assimilated root-zone soil moisture volumetric content and the so-called truth, obtained from the open loop experiment. This difference is normalized by the dynamical range of soil moisture, estimated as the difference between volumetric contents at field capacity and at wilting point.

The spatial variability of this convergence index, shown in Fig. 2, indicates that CaLDAS EnKF was able to recover within a month from the very dry values set at the initial day of the assimilation cycle. It is interesting to note that the assimilation of either L-band brightness temperatures or screen-level parameters leads to similar convergence, with even similar spatial patterns. In general, convergence is better for the southern portion of the domain. Reasons for this are not known at this time, and investigation is currently under way to relate these patterns to the Kalman gain for both experiments.



$$\frac{w_2 - v_{2\_Truth}}{w_{fc} - v_{wilt}}$$

Assimilation of synthetic brightness temperatures

Assimilation of synthetic screen-level air temperature and relative humidity

Valid 30 July 2007

Figure 2: Convergenge index for root-zone soil moisture (i.e., difference between the assimilated root-zone soil moisture and that generated in the open loop, normalized by the dynamical range of soil moisture, see equation in the figure) after 30 days of assimilation of synthetic brightness temperatures only (top panel) and synthetic screen-level air temperature and relative humidity only (bottom panel).

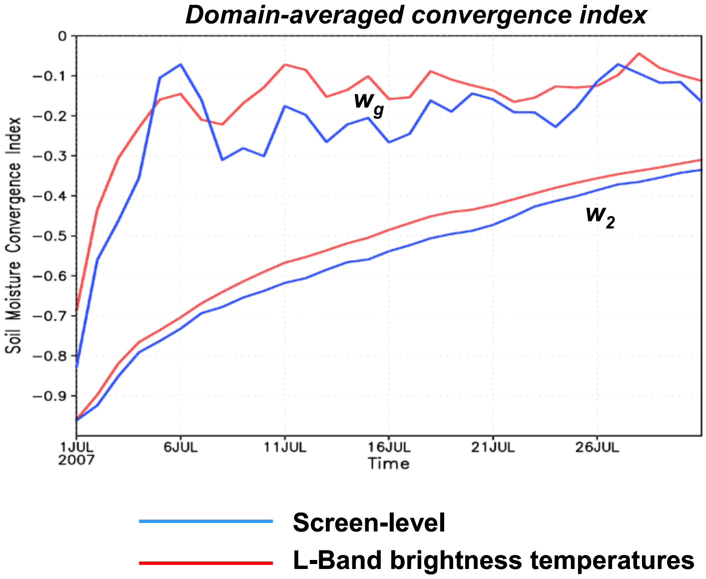


Figure 3: Convergence index for both near-surface (wg) and root-zone (w2) volumetric soil moisture contents as a function of days of assimilation, for the assimilation of L-band brightness temperatures only (red lines) and screen-level parameters only (blue lines).

The time evolution of domain-averaged root-zone soil moisture convergence, shown in Fig. 3, indicates that reasonable convergence is achieved for both experiments in this idealized context. It could be argued that more rapid convergence is achieved with the assimilation of brightness temperatures, but this conclusion would be preliminary considering the synthetic nature of the experiment. It should also be recognized that in reality screen-level parameters are available more frequently than the 6h assimilation step used here, whereas the opposite situation will exist for the L-band brightness temperatures (from SMOS or SMAP). There results show, however, that the ENKF algorithm coded in CaLDAS seem to behave properly (in fact, other tests not discussed here have shown that the simple variational approach also seems to work fine).

A set of more complex synthetic experiments, with a truth reference done at much higher horizontal resolution, with different atmospheric forcing and land surface characteristics than those used in the assimilation cycle, is currently being prepared.

4. Upcoming activities

It is expected that an EnKF version of CaLDAS, with assimilation of screen-level observations, will be tested in 2010 for possible operational implementation in CMC’s global medium-range and regional short-range NWP systems. Work should also start in 2010 to assure proper connections exist with both the 4DVAR upper-air assimilation system currently used operationally at CMC for the deterministic global system, and with the EnKF system operationally used for CMC’s Global Ensemble Prediction System.

Other research and development projects are also expected to start within the next year, focusing on the following aspects of CaLDAS :

- Modeling of observations (in control variable space) uncertainty
- Assimilation with high-resolution first guess
- Impact on deterministic and probabilistic NWP
- Evolution / improvement of land surface modeling (with ISBA or with CLASS).

5. References

Balsamo, G., J.-F. Mahfouf, S. Bélair, and G. Deblonde, 2007: A land data assimilation system for soil moisture and temperature: An information content study. *J. Hydromet.*, **8**, 1225-1242.

Balsamo, G., J.-F. Mahfouf, S. Bélair, and G. Deblonde, 2006: A global root-zone soil moisture analysis using simulated L-band brightness temperature in preparation for the Hydros satellite mission. *J. Hydromet*., **7**, 1126-1146.

Carrera, M.L., S. Bélair, V. Fortin, B. Bilodeau, D. Charpentier, and I. Doré, 2010: Land surface modeling of Canadian Rockies snowpack. Submitted for publication in *J. Hydromet*.

Lemonsu, A., S. Bélair, and J. Mailhot, 2010: The new Canadian urban modeling system: Evaluation for two cases from the Joint Urban 2003 Oklahoma City experiment. Accepted for publication in *Bound. Layer Meteor*.

Leroyer, S., J. Mailhot, S. Bélair, and I.B. Strachan, 2010: Numerical weather prediction over Montreal with the Canadian urbanized external land surface modeling system. In preparation, to be submitted to *J. Appl. Meteor. Climatol*.

Li, X., M. Charron, L. Spacek, and G. Candille, 2008: A regional ensemble prediction system based on moist targeted singular vectors and stochastic parameter perturbations. *Mon. Wea. Rev.*, **136**, 443-463.

Mahfouf, J.-F., B. Brasnett, and S. Gagnon, 2007: A Canadian Precipitation Analysis (CaPA) project: Description and preliminary results. *Atmos.-Ocean*, **45**, 1-17.

Reichle, R.H., D.B. McLaughlin, and D. Entekhabi, 2002: Hydrologic data assimilation with the Ensemble Kalman Filter. *Mon. Wea. Rev.*, **130**, 103-114.