

Version 3.0

# **A Land Data Assimilation System for soil moisture and temperature: An information content study**

G. Balsamo<sup>1</sup>, J.-F. Mahfouf<sup>1</sup>, S. Bélair<sup>1</sup>, and G. Deblonde<sup>2</sup>

Atmospheric Science and Technology Directorate

Environment Canada

<sup>1</sup>Recherche en prévision numérique

<sup>2</sup>Data Assimilation and Satellite Meteorology Section

Dorval, Québec, Canada

*Submitted to Journal of Hydrometeorology*

29 May 2006

Corresponding author present address:

Dr. Gianpaolo Balsamo

ECMWF

Shinfield Park, RG2 9AX, Reading, Berks., UK

email: gianpaolo.balsamo@ecmwf.int

## **Abstract**

A Canadian Land Data Assimilation System (CaLDAS) for the analysis of land surface prognostic variables is designed and implemented at the Meteorological Service of Canada for the initialization of numerical weather and climate prediction models.

The assimilation of different data sources for the production of daily soil moisture and temperature analyses is investigated in a set of Observing System Simulated Experiments over North America. A simplified variational technique is adapted to accommodate different observations at their appropriate time in a 24-hour time-window. The screen-level observations of temperature and relative humidity, from conventional SYNOP/METAR/SA networks, are considered together with presently available satellite observations provided by the AQUA satellite (microwave C-band), GOES (infrared, IR), and observations available in the future by SMOS satellite mission (microwave L-band). The aim of these experiments is to assess the information content brought by each observation type in the land surface analysis. The observation systems are simulated according to their spatial coverage, temporal availability, and nominal or expected errors. The results show that the observable with the largest dynamical response to perturbations of the control variable carries the greatest information content into the analysis. The observational error and the observation frequency counterbalance this feature in the analysis.

If one consider a single observation both for soil moisture and soil temperature analysis, the satellite measurements (L-band, C-band and IR in decreasing order of importance) are the primary source of information. When observation availability is considered and the highest temporal frequency of assimilated screen-level observations is used (one hour), a large amount of information is extracted from SYNOP-like networks.

The screen-level observations are shown to provide valuable soil moisture information mainly during the day-time while during night-time these observations (and particularly screen-level temperature) are mostly useful for the soil temperature analysis. The results are presented with perspectives for future operational developments and preliminary assimilation experiments are performed with hourly screen-level observations.

## **1. Introduction**

The land surface water and energy budgets have been the focus of several international research efforts from the point of view of modeling, data assimilation, and observation. An improved knowledge of land surface and its interaction with the atmosphere is known to be of crucial importance for the Numerical Weather Prediction (NWP) and Climate community. A realistic land surface description is likely to extend nowadays prediction capabilities both in terms of temporal range and of spatial resolution, and it appears a necessary development for global mesoscale NWP (Viterbo and Courtier 1995, Koster et al. 2004, Ferranti and Viterbo, 2006).

In this paper, the development of a Canadian Land Data Assimilation System (CaLDAS) prototype for the land surface analysis in the NWP context at the Meteorological Service of Canada (MSC) is presented. The focus is put on soil moisture and temperature as principal components driving the water (i.e. through evapotranspiration) and energy (i.e. via nocturnal radiative cooling) exchange at the surface, which are particularly effective on the evolution of the lower troposphere. Moreover, improvements of soil moisture and temperature predictive skills might be of interest for coupling NWP forecasts with hydrological and agricultural applications, such as flooding and drought risks assessment,

crop yield estimates, fire risk, etc. Improvements in soil temperature should allow a more realistic evolution of cold land processes in particular related to snow pack evolution and soil ice melting-freezing.

A widely used approach for the analysis of soil moisture and temperature in NWP context consists in the assimilation of indirect observations from the ground-based SYNOP-like network (at ECMWF, HIRLAM, MF, and MSC).

In particular, the assimilation of screen-level temperature and relative humidity observations with the optimum interpolation (OI) technique proposed by Mahfouf (1991), has proven to be reliable to retrieve the soil water content and soil temperature under specific meteorological conditions, and is widely used in operations in different NWP centers (Bouttier et al. 1993 a, b, and Giard and Bazile 2000, Mahfouf et al. 2000 and Douville et al. 2000, Bélair et al. 2003 a, b, Rodríguez et al., 2003).

In recent years, several studies have focused on variational and Kalman filter techniques, particularly for the analysis of soil moisture via the assimilation of screen-level observations, showing the reliability for operational NWP: Callies et al. (1998), Rhodin et al. (1999), Bouyssel et al. (2000), Hess (2001), Seuffert et al. (2004), and Balsamo et al. (2004a).

In parallel, the research advances in observing soil moisture have indicated that a successful use of remote sensing data (microwave and infrared radiances) may be specifically adapted for routinely measuring the soil moisture on global to continental scales (Kerr et al. 2001).

Ground-based soil moisture measurements are difficult to gather and only few observation networks exists (i.e. Global Soil moisture Data Bank, Robock et al., 2000).

This data is currently used mostly for validation purposes, and it often lacks representativity due to the low spatial correlation of soil moisture associated to the land surface heterogeneity. Besides the data transmission latency would be a major obstacle for operational data assimilation. Contrary to ground-based, satellite radiometric measurements are appropriate to capture timely area-averaged soil moisture signals, useful for the initialization of mesoscale NWP models (10-50 Km).

The microwave radiometric data for sensing superficial soil water content have been investigated in the past decade (Jackson 1997, Jackson et al. 1999, Drusch et al. 2001, Jackson et al. 2002, Njoku and Entekhabi, 2003) and seem to be promising at providing a global coverage.

The low frequency channels of the Special Sensor Microwave Imager (SSM/I) and the Advanced Microwave Scanner Radiometer (AMSR-E) are greatly influenced by the atmosphere and vegetation, and the use of a single polarization retrieval algorithm (Drusch et al., 2001) has shown some limits when compared to field data of the Southern Great Plains (SGP97) hydrology experiment. Nevertheless valuable information can be extracted (Jackson et al. 2002) as supported by airborne radiometric studies. Beside the presently available platforms, L-band microwave observations from the Soil Moisture and Ocean Salinity (SMOS) mission by the European Space Agency (ESA; Kerr et al. (2001), will be available from September 2007.

As far as surface temperature is concerned, model accuracy so far has not allowed the assimilation of satellite derived skin temperature, although this is certainly the most frequently observed land surface field with a large coverage. The infrared (IR) sensor onboard the geosynchronous satellites provides at a frequency of 1 hour or better (i.e. 15

min. with Meteosat Second Generation, MSG) a surface skin temperature in cloud free areas. Because of its large coverage and high resolution, the use of such a data has been proposed first by Wetzal et al. (1984) and investigated for soil state retrieval in several studies (McNider et. al. 1994, Bastiaanssen 1995, Jones et. al. 1998a, b, van den Hurk et al. 1997, van den Hurk and The 2002). The main obstacles preventing the assimilation of IR data are the sensitivity to surface roughness length and the still large underestimation of the diurnal cycle as highlighted by Trigo (2004). At mid-latitudes the cloud cover conditions reduce sensibly the number of available observations.

The synergetic assimilation of satellite and ground-based observations represents a real challenge for the next years in terms of improving the land surface description. In the present work, we explore the potential of the diverse observational platforms for the analysis of the root-zone soil water content and the soil temperature in the operational NWP context at the MSC.

A set of simulated observations is generated over North America and assimilated with a bi-dimensional simplified variational analysis (2D-VAR) which is based on a linear estimate of the observation operator. The information carried by each observation type is investigated in a set of Observing System Simulated Experiments (OSSEs). The influence matrix diagnostics (Cardinali et al., 2003) provide the information content of the observations in the analysis, therefore deriving the importance of each source of observation.

Section 2 describes the CaLDAS prototype used in the simulations experiments, which consists of three components: the NWP model providing the atmospheric forcing, the land surface scheme, and the analysis scheme. Section 3 is dedicated to the

observations, including the observation operators for microwave and infrared data, the satellite orbit, and the ground-based network simulations. In section 4, the setup for the land surface analysis is presented. Section 5 is dedicated to the information content study and discusses the results of the OSSEs performed. The results of an assimilation cycle over North America, which assimilates hourly screen-level observations, are shown in section 6 together with consideration of the effectiveness of the assimilated observations. The conclusions of the study, discussed within the context of the simulation hypotheses, are presented in section 7. Finally, the application of the system to real observations and the perspectives are presented.

## **2. The Land Data Assimilation System**

### *a. The land surface scheme (LSS)*

The land surface scheme ISBA (Interaction Soil Biosphere Atmosphere) (Noilhan and Planton 1989, Noilhan and Mahfouf 1996) comprises two soil layers with associated variables that describe the evolution of temperature and water content and is based on a force-restore mechanism.

The operational implementation of the ISBA scheme with a surface analysis, described in Giard and Bazile (2000), has required some further refinements, like the inclusion of soil freezing that allows a better description of the diurnal cycle during winter.

The land surface scheme ISBA has been implemented operationally at the MSC together with the OI land surface analysis scheme (Mahfouf, 1991) as described in Bélair et al.

(2003 I, II) and is currently used in the regional version of the Global Environmental Multiscale (GEM) model (Mailhot et al., 2005, Côté et al., 1998).

The operational version of ISBA describes the evolution of six prognostic variables:  $T_s$  the surface temperature,  $T_p$  the daily mean surface temperature,  $W_s$  is the superficial water contents,  $W_p$  and  $W_{pi}$  the total water contents (liquid and ice phase respectively), and  $W_l$  the vegetation intercepted water content. Four more variables are dedicated to the evolution of the snow pack (not listed). The variables  $T_s$ ,  $T_p$ ,  $W_s$ , and  $W_p$  are operationally analyzed on the global scale by the OI method. Each grid box over land is divided into bare ground and vegetated areas. A single heat budget is considered for the ground and vegetation, and if appropriate, snow cover. Within the hydrological soil reservoir, with a variable depth  $d_p$ , a thin top layer ( $d_s=10^{-2}$  m) is considered. The rainfall is the main input to the land-surface hydrology and drives the local water storage. The fraction of rainfall reaching the bare ground enters directly the soil reservoir, while the water intercepted by the leaves drips to the surface or re-evaporates. Snow and frozen water in the soil may hold an important amount of water during winter preventing the decrease of soil temperature.

In the parameterization of the evaporation processes,  $W_p$ , which contains the root-zone, largely determines the evaporation from bare ground (through dynamical constraint on the superficial soil water reservoir,  $W_s$ ) and vegetation (via the resistance to transpiration). In a similar way and in absence of strong radiative forcing, the mean surface temperature,  $T_p$  (slow thermal component) constrains the superficial temperature  $T_s$  (fast component). Although  $T_p$  does not refer to a precise depth, since this is not contemplated in the force-restore method, this variable is acting as a deep soil temperature in a multi-layer scheme,

here associated to a mean soil temperature at a depth below 1 cm (coherently with the assumption made for the total soil moisture). A land surface analysis, aiming at the initialization of NWP forecasts, has to concentrate on total (plant available) soil moisture and mean soil temperature, since the atmospheric impact near the surface lasts for several days.

*b. The atmospheric coupling*

The land surface state evolution and the coupling to the atmosphere can be modeled considering a 1-way or so called off-line mode, where the atmospheric state is a prescribed input, or alternatively, a 2-way or atmospheric-coupled mode where the land surface interacts with the atmosphere during the integration. In NWP, the land surface scheme is normally fully coupled to the atmosphere (Mahfouf et al. 2000, Bélair et al. 2003) and the full GEM model (Mailhot et al., 2005, Côté et al., 1998) is used in the 2-way type experiments.

The off-line approach has been investigated for the setup of a Land Data Assimilation System (LDAS) in the U.S. with the North-American and Global LDAS initiatives (NLDAS, Mitchell et al. 2004, and GLDAS, Rodell et al. 2004). Furthermore, this approach is widely used in LSSs intercomparison projects (i.e. Project for Intercomparison of Land surface Parameterization Schemes - PILPS, Global Soil Wetness Project - GSWP, Rhone Aggregation project - RhoneAgg). The atmospheric coupled approach has been considered in the frame of the European LDAS (van den Hurk, 2002, Balsamo et al. 2004b).

The applicability of both the off-line and the atmospheric coupled approach, is based on the assumption of a truncated control variable space, and considers that the physical and dynamical processes near the surface can be treated separately from the upper atmosphere due to the time and space scales involved. The atmospheric state provides the near surface forcing for radiation, precipitation, temperature, pressure, and wind, which drive the land surface evolution but without allowing for atmospheric retroaction from the land surface.

In the off-line mode setup considered here, the atmospheric forcing is provided at the lowest vertical level of the GEM model, which has an elevation of about 50 m above the surface. This allows for an interactive evolution of the surface boundary layer (SBL), without having to run the complete NWP model. The vertical interpolation is done according to Delage (1997) and provides screen-level temperature and humidity values. The atmospheric forcing frequency is 3-hourly whereas the integration time-step of the off-line system is 30 minutes (a linear temporal interpolation is performed).

In one experiment, the off-line and the atmospheric coupled approaches have been applied to the same set of simulated observations to provide an estimate of the approximation involved in using this off-line system and to investigate the truncated control variable assumptions of land surface analyses.

*c. The analysis scheme*

The land surface analysis scheme, described in Mahfouf (1991), optimally combines the information provided by screen-level observations and a model forecast

(the background term). In the variational method, the optimal weighting of these two components is achieved by the minimization of a cost function  $J(\mathbf{x})$ , which is of the form

$$J(\mathbf{x}) = \frac{1}{2} (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + \frac{1}{2} (H(\mathbf{x}) - \mathbf{y})^T \mathbf{R}^{-1} (H(\mathbf{x}) - \mathbf{y}) \quad (1)$$

where  $\mathbf{x}$  is the control variable vector,  $\mathbf{x}_b$  is the background (i.e. a short-range forecast),  $\mathbf{y}$  is the observation vector, and  $\mathbf{B}$  and  $\mathbf{R}$  are respectively, the background and the observation error covariance matrices.  $H$  is the observation operator that maps the model state vector  $\mathbf{x}$  into observation space and includes spatial interpolation and integration in time. Under the Tangent Linear (TL) hypothesis, the  $H$  operator is expressed by its first order Taylor expansion:

$$H(\mathbf{x} + \delta\mathbf{x}) = H(\mathbf{x}) + \mathbf{H} \delta\mathbf{x} \quad (2)$$

where  $\mathbf{H}$  is the TL observation operator matrix and the cost function is then quadratic. The minimization of  $J(\mathbf{x})$  is generally an iterative process that makes use of the tangent linear and adjoint models to evaluate the cost function and its gradient. However, for low dimension problems and under the linearity assumption, the minimum of the cost function imposed by  $\nabla J(\mathbf{x}) = 0$  can be directly obtained. The  $\mathbf{H}$  matrix is evaluated with a finite difference approach, and the analyzed state  $\mathbf{x}_a$  is given by the expression

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K} (\mathbf{y} - H(\mathbf{x}_b)) \quad (3)$$

where the terms:

$$\mathbf{K} = (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} \quad (4)$$

$$(\mathbf{y} - H(\mathbf{x}_b)) \quad (5)$$

are respectively named the gain matrix (Eq. 4) and the innovation vector (difference between the observation and the model forecast in observation space) (Eq. 5). The analysis error covariance matrix expresses the analysis uncertainties and is provided by

$$\mathbf{A} = (\mathbf{I} - \mathbf{KH}) \mathbf{B} \quad (6)$$

The linear hypothesis, expressed by (2), is best satisfied in the vicinity of  $\mathbf{x}_b$  when corrections to the first guess are small. This important simplification in the 2D-Var (hence called *simplified*) makes it similar to the OI technique. Both methods in fact obtain the analysis increments directly from (4), but in the simplified 2D-VAR, a dynamical estimate of the gain matrix replaces a statistical evaluation. In the 2D-VAR method, the atmospheric forcing and the characteristics for soil and vegetation are implicitly taken into account providing analysis increments adapted to the specific time and grid point location without any empirical regression and not limited to the use of observations at a single time.

*d. The information content of the observations in the analysis*

The diagnostic  $Tr(\mathbf{HK})$  as described in Cardinali et al. (2003), can be used to infer the weight of an observation type in the analysis. If the observation operator  $\mathbf{H}$  is applied on both sides of the analysis equation (3) then:

$$\mathbf{H}\mathbf{x}_a = \mathbf{H}\mathbf{x}_b + \mathbf{HK} (y - H(\mathbf{x}_b)) \quad (7)$$

Differentiating the above equation with respect to the observation  $y_0$  and taking the trace of the resulting matrix gives the sensitivity of the analysis to a given observation type:

$$Tr(\partial\mathbf{H}\mathbf{x}_a / \partial y) = Tr(\mathbf{HK}) \quad (8)$$

which when  $H$  is linear corresponds to the analysis error reduction, so that

$$Tr(\mathbf{HK}) = \mathbf{K}^T \mathbf{H}^T = Tr((\mathbf{B} - \mathbf{A}) \mathbf{B}^{-1}) \quad (9)$$

where  $\mathbf{A}$  and  $\mathbf{B}$  are the error covariance matrices for the analysis and the background, previously defined. This diagnostic estimates the relative information content of each observation type in the soil moisture analysis.

### 3. The Observation System Simulation

To assimilate a new type of observation, a first step consists in implementing the corresponding observation operator. Such an operator is composed of a set of parameterizations allowing the mapping of the land surface state in observation space. In

the present system, a satellite brightness temperature (in the considered radiometric frequency) or a screen-level temperature and humidity have to be associated with a particular land surface state.

The experiments are realized in a "perfect" observation assumption. The guiding requirements for a realistic simulation are the distribution of the observations and the reproduction of the observation dynamical range. These two features rule the observation availability in the analysis and the sensitivity for perturbations applied to the control variable which reflects on the values of the Jacobians in the analysis. The description of the parameterizations used is provided below.

*a. Microwave L-band and C-band*

Microwave radiative transfer models for land surface already exist and have been tested in a number of experiments. In this study, we have considered the following models: the Land Surface Microwave Emission Model (LSMEM), developed by Drusch et al. (1999, 2001), the LMEB (L-band Microwave Emission of Biosphere) developed by Pellarin et al. (2003), and a simple approximation for the infra-red model. The LSMEM has been tested recently on several case studies (Seuffert et al. 2004, Gao et al. 2004, Drusch et al. 2004) and is coupled to the ISBA scheme at MSC (Balsamo et al., 2006).

The relationship between soil moisture and microwave brightness temperature follows from the influence that soil moisture has on the dielectric properties of the soil, and therefore on the emissivity. The solution of the radiative transfer equation is calculated as in Kerr and Njoku (1990). The vegetation contribution is treated according

to the effective medium theory of Kirdyashev et al. (1979). The vegetation opacity  $\tau$  is linearly dependent on the vegetation water content VWC and an experimental parameter  $b$  accounts for the vegetation structure as follows:

$$\tau = b \cdot VWC \quad (12)$$

The values of vegetation water content VWC, single scattering albedo  $\omega$ , and vegetation structure parameter  $b$  for the L and C bands are reported in Table 1.

The simulated observations have to be spatially and temporally distributed according to the prescribed satellite coverage. Table 2 reports a summary of orbital information to simulate the satellites overpasses and their views.

For L-band observations a polar orbit is defined as a Keplerian orbit (Larson and Wertz, 1992) which satisfies a prescribed satellite inclination, altitude and revisit time of Hydros (Entekhabi et. al., 2004). It can be noted that the orbit geometry of other polar orbiting satellites is similar and results in comparable revisit time and data coverage. Therefore the procedure used in Balsamo et al. (2006) is transferred to application for SMOS. For C-band observations, the AQUA polar orbit is taken from an observed satellite track. For the polar orbiting satellites, a circular swath with given amplitude (Table 2) around the nadir location is considered to reproduce the expected satellite spatial coverage. The simulated observations are stored in hourly transects coverage (along the trajectory for polar orbiting satellites) between time  $t_i=t-30 \text{ min.}$  and  $t_f=t+30 \text{ min.}$

*b. Infrared (derived) surface temperature*

To simulate clear-sky infrared skin temperature, the ISBA soil surface temperature (which accounts for a mixed soil-canopy temperature) is taken over clear-sky areas with total cloud cover fraction less than 0.1. The approximation  $T_{SIR} \approx T_s$  is equivalent to assuming a uniform infrared emissivity of 1.0 while in reality this value ranges between 0.95 and almost 1.0, depending on land surface physiography and on the considered infrared channel. Following Garand (2003), the  $T_{SIR}$  is estimated by inverting the infrared radiative transfer model or with a 1D-Var. It is then divided by the emissivity, so that it corresponds to a physical skin temperature. Errors of the emissivity maps, the radiative transfer model, and the satellite observation have to be accounted for in the derivation of the observations.

The geostationary satellite coverage is simulated considering an elevation of 36000 km and a limit off-nadir angle of  $8.0^\circ$  to obtain the perimeter coverage. The infrared surface skin temperature is generated during the daylight and only for morning hours when the relationship with soil moisture is more direct as in van den Hurk and The (2002). Each simulated observation is assumed to match perfectly the model grid box and no sub-grid information is considered.

*c. Screen-level temperature and relative humidity*

The simulation of screen-level temperature and relative humidity is rather straightforward since those fields are diagnostic outputs of the GEM NWP model. However special care is given to realistically evaluate the observation errors and to limit the observation availability to regions where the distribution of SYNOPs, METARs and SAs stations is

dense. The operational OI screen-level analysis method (Bélair et al., 2003a) is applied to interpolate station points to model grid-points, considering both the station observation and a model background value.

To avoid over-weighting the screen-level observations in the assimilation, for instance over regions where the station density is poor, the real distribution of data is taken into consideration and the observation error is set that reflects this. (Table 4).

A quality control is applied based on the standard deviation error for the observation departure at the station point. For the screen-level analysis configuration, the errors prescribed are reported in Table 3. The GEM background error statistics for the screen-level temperature and relative humidity (as RMSE, BIAS and number of observations) are plotted for a selection of 3 days (1 May, 15 July, and 1 October 2005) in Fig. 1 while an example of the spatial patterns of the screen-level analysis error is provided in Fig. 2. The screen-level analysis error maps allow setting the observation errors in the 2D-VAR in a less arbitrary way, since these fields actually reflect the observation density and the representativity for these variables.

#### **4. A multiple observation-type 2D-VAR for $W_p$ and $T_p$ analyses**

The set up for the simplified 2D-VAR makes use of the full NWP model or the simplified SBL version to infer the forecast sensitivity to a perturbation of soil moisture or soil temperature and provides a linear estimate of the observation operator ( $\mathbf{H}$ ) from which the analysis gain matrix ( $\mathbf{K}$ ) is calculated. In the CaLDAS set up, the simplified 2D-VAR method makes use of the off-line driven land surface scheme coupled with the

observation operator models (i.e. radiative transfer) to obtain the sensitivity of the observable  $y$  to the analysis control variable (the total soil moisture  $W_p$  or the mean soil temperature  $T_p$ ). The liberalized observation operator  $\mathbf{H}$  is calculated as  $\Delta y(t=t_i)/\Delta W_p(t=t_0)$ , according to the perturbation method described in Balsamo et al. (2004a) for the following observations: the L-band brightness temperatures  $Tb_H (L)$  and  $Tb_V (L)$  (horizontal and vertical polarization at 1.4 GHz), the C-band brightness temperatures  $Tb_H (C)$  and  $Tb_V (C)$  (6.9 GHz), the clear-sky infrared derived skin temperature  $T_{sIR}$ , and the screen-level temperature  $T_{2m}$  and relative humidity  $RH_{2m}$ .

Let us consider the model guess  $G$  and a perturbed state  $G'$  where  $X$  (being  $W_p$  or  $T_p$ ) has been modified by a small quantity  $\Delta X$ , at time  $t_0$ . The model integrations starting from the two initial states,  $G$  and  $G'$ , provide the forecast sensitivity evaluated at time  $t_1$  (at which the observations are available).

Assuming that all the simulated observations (one per type) are available at different times in the assimilation time-window, the analysis for  $X$  is obtained as follow

$$X^a = X^b + \begin{pmatrix} k_1 & k_2 & k_3 & k_4 & k_5 & k_6 & k_7 \end{pmatrix} \cdot \begin{pmatrix} \Delta Tb_H (L) \\ \Delta Tb_V (L) \\ \Delta Tb_H (C) \\ \Delta Tb_V (C) \\ \Delta T_s (IR) \\ \Delta T_{2m} \\ \Delta RH_{2m} \end{pmatrix} \quad (7)$$

where the  $k_i$  terms are the elements of the analysis gain matrix  $\mathbf{K}$  obtained from (4).

The observation error covariance matrix  $\mathbf{R}$  is a diagonal matrix composed of  $(\sigma_y)^2$  where the cross-covariances are neglected (the observation errors, reported in Table 4, are supposed uncorrelated).

With a 24-h assimilation time-window, the soil moisture and temperature analyses are performed at 00 UTC. For satellite observations (i.e. the microwave brightness temperature at given  $P$  polarization,  $Tb_P$  and  $T_{SIR}$ ), the assimilation system has to be capable of ingesting observations at any time of day. The number of observations may vary according to satellite overpass, the meteorological conditions (in case of  $T_{SIR}$ ), and the land surface state. The synoptic observations (2m temperature and relative humidity) have been considered to be available at four synoptic times (06, 12, 18, 00(day+1) UTC) or every hour (24 observations per grid point per day).

This increase in observation intake in the analysis step is made possible due to the more flexible methodology presented in this paper compared to that of the OI technique, originally designed to assimilate  $T_{2m}$  and  $RH_{2m}$  every 6 hours (Bouttier et al. 1993a) or to consider only observations at 18 UTC in the implementation at MSC (Bélaïr et al. 2003a).

The simplified 2D-VAR method relies on the linear hypothesis of the observation operator, on the horizontal decoupling hypothesis of the surface points, and on the control variable decoupling assumption, allowing the consideration of the analysis of soil moisture and temperature separately from each other and from other model variables. These hypotheses allow this 2D formalism to be performed at each grid point separately. An extensive treatment and validation of those hypotheses are given in Balsamo et al. (2004a) in the case of screen-level observations.

The optimal set up of the simplified 2D-VAR considers a double estimate of the observation operator using two opposite phase chess-type initial perturbations. The  $\mathbf{H}$  matrix is given by the average of the two estimates. The perturbation amplitude is set to ten percent of the active range ( $w_{fc}-w_{wl}$ ) for soil moisture and 3 K for soil temperature.

## 5. Observation information content in the analysis

### a. An OSSE experiment for $W_p$

An Observing System Simulation Experiment (OSSE) is set up to test the performance of the analysis and investigate the impact of each assimilated observation type in correcting a given soil moisture initial error. A simplified 2D-VAR assimilation experiment with a 24-h time-window is performed over North America on 1 May, 15 July, and 1 October 2005. Those dates are chosen to extract samples of the warm season over North America. The initial state for the land surface water (soil moisture, ice content and snow water equivalent) is provided for the three dates in Fig. 3. A set of simulated observations is generated according to the procedure described in section 3. The objective of the test is to assess the impact of the simulated observations (an example is shown in Fig. 4) on the soil moisture analysis. The results of an assimilation cycle are analyzed to produce the  $Tr(\mathbf{HK})$  diagnostic which is summarized in Fig. 5. Although the observations are simulated and therefore the innovation contains no useful information, the Jacobians are realistically providing the gain which be used to assimilate real observations.

When assimilating 6-hourly screen-level observations and all types of satellite data, the analysis is mostly influenced by the L-band data, followed by those in C-band observations, the screen-level observations, and finally the infrared observations. The infrared surface temperature has a larger contribution (per observation) than screen-level temperature or relative humidity but this value is likely to change from one day to another depending on cloud cover conditions. A similar comment on information content variability can be made for the screen-level observations since the soil moisture sensitivity is bounded between the soil wilting point and the field capacity, and it is reduced under specific meteorological and land surface conditions and is modulated by the diurnal cycle. When considering the screen level analysis hourly and again, assimilating all types of satellite observations, the screen level observations are dominant in the analysis. This result has important consequences in the design of the LDAS system that are discussed in section 7.

*b. An OSSE experiment for  $T_p$*

In case of soil temperature, the atmospheric feedback is mostly effective during night-time when radiative cooling occurs at the surface. The reduced number of processes involved in this interaction (w.r.t. soil moisture) determines that a soil temperature analysis is somehow simpler to set. Practical implementations of the OI method, following Mahfouf (1991), have adopted a correlation of 1 between screen-level and soil temperature errors and ignored screen-level relative humidity (i.e. Giard and Bazile, 2000). The experiments performed for the analysis of the mean soil temperature,

presented in Fig. 6, support this choice and provide indication on the opportunity of including satellite observations.

The screen-level temperature is important in the analysis with high information content even when considering only a 6-hour observation frequency, and it becomes dominant if assimilated hourly. The relative humidity contains information as well but to a lesser extent ( $\sim 20\%$  of  $T_{2m}$ ), since physically related with temperature.

L-band is, not surprising, quite informative also for soil temperature, as confirmed in sensitivity tests with the radiative transfer model (i.e. Balsamo et al., 2006) which indicate that after soil moisture, soil temperature is the most effective parameter determining the surface brightness temperature. An interesting feature for the dual polarization microwave frequencies is the absence of a dominant component as for the soil moisture case (where the H component always contains more information). Infrared skin temperature has little information content following from the fact that only clear-sky morning observations are considered, therefore  $T_s(IR)$  is mostly responsive to solar radiation.

*c. The normalized information content*

The experiments performed so far have provided an estimate of the information content of a given observation network source considering the observation availability and its temporal and spatial coverage. To obtain an estimate of the potential of each observation, the previous results are normalized by the number of observations, and presented in Fig. 7. Both for soil moisture and temperature, the importance of L-band is

evident due to its high dynamical range of the response to soil moisture and temperature changes. The C-band and the clear-sky morning IR skin temperature (more effective for soil moisture) are less important than L-band. For the IR observations, one has to stress that the domain considered penalizes this observation with an extent that does not exceed 60 degrees north. The screen-level observations have the smallest information content if one considers a single observation, but they have the highest frequency and a good spatial coverage on the domain.

*d. On the equivalence of Off-line and Atmospheric coupled approach*

The information content diagnostics are used here to investigate the equivalence of the off-line CaLDAS setup with a fully coupled atmospheric system. The experiment is performed for the  $W_p$  analysis as previously described in section 5 a. The GEM model is used as observation operator to estimate the linearized Jacobians. This test allows evaluating the degree of approximation made when using an off-line configuration (where the atmospheric forcing is imposed at a given height of about 50m) which is computationally much less expensive than using the fully coupled mode. The test is performed for 5 July 2004 and the information content is extracted for 89223 land-points, for soil moisture only. In Fig. 7, the relative information content is displayed. The information content is preserved for each observation source and the spatial patterns of the Jacobians results are very similar (not shown). The screen-level information content is slightly reduced in the off-line system (in favor of L/C-band) as one can expect from having the model lowest atmospheric layer prescribed and not interactive. Nevertheless,

this feature is not undesirable and does not prevent the assimilation of screen-level observations since the gain is only slightly reduced but conserves signature and spatial patterns. The truncated control variable assumption for the land surface analysis is therefore acceptable for the considered time-window (1-day) although it is probably not satisfactory for longer (climate) periods of time.

## **6. Assimilation experiments**

### *a. A pre-operational configuration assimilating hourly screen-level observations for $T_p$ and $W_p$ analyses*

The screen-level observations, including airport data, are available hourly over a large portion of North America. These observations are widely used and verified and constitute a data source which could be readily assimilated in the operational version of CaLDAS. A set of assimilation experiments has been performed assimilating  $T_{2m}$  and  $RH_{2m}$  analyses. The simplified 2D-VAR is applied sequentially with a 24-h time-window. The screen-level observations are distributed on the model grid by the OI analysis (adapted from Brasnett, 1999) and are assimilated every hour where available. The OI analysis error (Fig. 2), calculated according to Rutherford (1972), is used to prescribe the observation error in the simplified 2D-VAR on each grid-point separately. As a result, the screen-level observations are more effective in constraining soil moisture and temperature in the regions where the SYNOP-like network is dense, while in areas with scarce observations the soil moisture analysis increments is negligible and the background value is kept unchanged.

The assimilation experiments are performed for the 1 May 2005. The analysis is performed sequentially for soil temperature and soil moisture (in the order). The analyzed land surface state is then integrated with the off-line system for 24 hours and the RMSE and BIAS of the model screen-level are calculated against around 2000 SYNOP-like observations as shown in Fig. 8 for the control and the analyzed model fields. This provides an evaluation of the so-called analysis residuals which has to be smaller than the innovations as the land surface analysis has minimized the screen-level observation departures.

The new land surface state obtained has a better match of the near-surface diurnal cycle both in temperature and relative humidity. Although the number of observations considered (~2000) supports the robustness of the results, further experiments need to be run to evaluate the impact of the land surface state in longer assimilation cycles. A set of atmospheric forecasts issued from the analyzed fields will be performed in the near future.

*b. Diurnal cycle of the screen-level information content*

The simplified 2D-VAR is applied sequentially with a 24-h time-window. Now, one single time slot (at varying time along the time-window) is considered in the assimilation period. The purpose of this test is to improve the understanding of the mechanisms that relate screen-level observations to soil moisture and temperature. The information content diagnostics are extracted to assess at which time during the day the observations have a greater chance to correct soil moisture and temperature in the analysis. The

physical processes involved for the two control variables are quite different even if one could argue that soil temperature and soil moisture are correlated with each other. This correlation becomes quite large if melting/freezing occurs, but to avoid this, a soil moisture analysis is not performed if the soil is frozen, and the soil temperature analysis is performed only in a temperature range that is far from melting/freezing phase (i.e. no analysis is performed for soil temperature in  $\pm 3$  K around 273.15 K).

In Fig. 10, the information content of hourly screen-level observations assimilated at a single hour time-slot in a 24-h time window is evaluated. The soil moisture and soil temperature analyses are performed each time using the given single set of observations at time  $t$  (with  $t$  going from 1- to 24-hour).

For the soil moisture analysis, the observations are likely to be more effective during the day and present a smaller contribution over night. This reflects the dependency of the soil moisture-evapotranspiration feedback from strong solar radiative forcing and explains the greater variability of the signal during the day (particularly evident in the relative humidity).

For the soil temperature analysis, the screen-level observations, and particularly temperature, are dominant during the night with a rather smooth bell-shape centered on midnight local time (for the center of the domain), whereas it is almost negligible during the day. Screen-level temperature is more important by a factor five than relative humidity in constraining soil temperature. As a result, the two curves have an opposite phase which results from the different processes relating the observable and the control variable: the long-wave radiative cooling control at night of the soil temperature at

screen-level and the control on evapotranspiration exercised by the soil moisture during day-time.

## **7. Discussion and conclusions**

A Canadian Land Data Assimilation System (CaLDAS) is developed at the Meteorological Service of Canada. The system is designed to ingest different source of observations obtained from either ground-based or satellite instruments. The observations are realistically simulated but are assumed to be bias-free since the purpose of this work is to assess the observation information content in a NWP framework where bias and quality control procedures are previously applied.

The ISBA land surface scheme integrations are performed in an off-line mode, where the atmospheric forcing is provided by the lowest NWP model level which is at 50 m above the ground and thus preserves an interactive surface boundary layer. This setup is shown to provide a good approximation of the full NWP system for estimating the Jacobians in the analysis. The use of an off-line system presents several practical advantages. Its low computation cost makes it affordable even for high resolution (i.e. 1-km global). The introduction of analyzed forcing such as those provided by the Canadian Precipitation Analysis (CaPA project, Mahfouf et al., 2005) can be investigated to reduce the error of the background soil moisture field. The simplified variational (2D-VAR) technique (Balsamo et al. 2004a), adapted to accommodate the observations at various times in a 24-h time-window, provides a flexible data assimilation scheme. The analysis equation is differentiated in observation space to provide the influence matrix diagnostic (Cardinali

et al., 2003). This tool allows for the study of the relative contribution of each observation type that is assimilated in the analysis. The screen-level observations (temperature and humidity) are simulated together with currently available satellite observations, as for instance those provided by AQUA satellite (microwave C-band), GOES (infrared). Observations available in the future by the SMOS satellite mission (microwave L-band) are also considered.

From these results, the influence of the observations in the analysis appears to be proportional to the ratio given by the dynamical range of the observable  $y$  and the observation error  $\sigma^o$ . A large contribution of microwave brightness temperatures and in particular of the L-band observations is obtained as is theoretically expected. These observations have the advantage of a reduced atmospheric absorption and a large dynamical range. The associated information content is likely to remain stable.

Infrared skin temperature observations represent a significant source of information for soil moisture, especially considering the observation frequency, but the temporal and spatial availability (limited by morning clear-sky window) is highly variable and shows a latitude dependence (van den Hurk and The, 2002).

The inclusion of real satellite data will have to face model biases and inaccuracies in the surface description as highlighted by Seuffert et al. (2004) and Drusch et al. (2006).

The screen-level observations, informative on the evapotranspiration processes and nocturnal radiative cooling of the surface, are readily available in today's land surface analyses. Their information content may vary depending on meteorological conditions and location. Although it is not the dominant data source if one considers the weight per observation, these measurements are available with high temporal frequency (hourly over

North America) and benefit from operational quality control. The link of screen-level observations to land surface fluxes make the assimilation of these variables desirable in a NWP assimilation system since this can translate into low-level atmospheric forecast improvements. Satellite observations when available should be added on top of this data source to reduce the land surface uncertainties and to provide more direct information on soil water content and temperature. Preliminary results assimilating real screen-level observations show that the analyzed land surface fields provide a better match to the near surface atmospheric state. Extensive forecast verification and the inclusion of real satellite data will be considered in the near future as extension to this work. The information content diagnostic is regarded as a useful tool for evaluating a land data assimilation system once in operations.

## **Acknowledgements**

The authors wish to thank Bruce Brasnett for his precious assistance with screen-level analyses. We are grateful for contributions made by Dr. Yves Delage to the off-line system developments and we all miss him very much. We are thankful to Pierre Pellerin for helping with the model setup. Matthias Drusch and Thierry Pellarin are acknowledged for providing the L-band radiative transfer models and for helpful comments. The financial support for a visiting fellowship in a Canadian government laboratory has been provided by the Earth Observation Government Related Initiatives Program (GRIP) of the Canadian Space Agency.

## References

\_\_\_\_\_, Bouyssel F., Noilhan J., 2004a: A simplified bi-dimensional variational analysis of soil moisture from screen-level observations in a mesoscale numerical weather prediction model. *Quart. J. Roy. Meteor. Soc.*, **130A**, 895-916.

\_\_\_\_\_, Bouyssel F., Noilhan J., Mahfouf J.-F., Bélair S., Deblonde G., 2004b: A simplified variational analysis scheme for soil moisture: Developments at Meteo-France and MSC. *Proc. ECMWF/ELDAS workshop on land surface assimilation. Reading, 8-11 November 2004.*

\_\_\_\_\_, J.-F. Mahfouf, S. Bélair, G. Deblonde, 2006: A global root-zone soil moisture analysis using simulated L-band brightness temperature in preparation for the Hydrosatellite mission", *J. Hydromet.* (accepted).

Bastiaanssen, W.G.M., 1995: Regionalization of surface flux densities and moisture indicators in composite terrain - A remote sensing approach under clear skies in Mediterranean climates. Agricultural University, Wageningen, The Netherlands

Bélair S., Brown R., Mailhot J., Bilodeau B. and Delage Y., 2003a: Operational Implementation of the ISBA land surface scheme in the Canadian regional weather forecast model. Part I: Warm season results. *J. Hydromet.*, **4**, 352-370

\_\_\_\_, Brown R., Mailhot J., Bilodeau B. and Crevier L.-P. 2003b: Operational Implementation of the ISBA land surface scheme in the Canadian regional weather forecast model. Part II: Cold season results. *J. Hydromet.* **4**, 371-386.

Bouyssel F., Casse V. and Pailleux J., 2000: Variational surface analysis from screen-level atmospheric parameters. *Tellus*, **51A**, 453-468.

Bouttier F., Mahfouf J.F, Noilhan J., 1993a: Sequential assimilation of soil moisture from atmospheric low-level parameters. PART I: Sensitivity and calibration studies. *J. Appl. Meteorol.*, **32**, 1335-1351.

Bouttier F., Mahfouf J.F, Noilhan J., 1993b: Sequential assimilation of soil moisture from atmospheric low-level parameters. PART II : Implementation in a mesoscale model. *J. Appl. Meteorol.*, **33**, 1352-1364.

Brasnett, B. 1999: A Global Analysis of Snow Depth for Numerical Weather Prediction *J. of Appl. Meteorol.*, **38**, 726-740.

Callies U. and Rhodin A. and Eppel D.P., 1998: A case study on variational soil moisture analysis from atmospheric observations. *J. Hydrol.*, **212-213**, 95-108.

Cardinali C., Pezzulli S. and Andersson E. 2003: Influence matrix diagnostic of a data assimilation system. *ECMWF Seminar on Recent developments in data assimilation for atmospheric and ocean*, 8-13 September, Reading, UK.

Côté, J., S. Gravel, A. Méthot, A. Patoine, M. Roch, and A. Staniforth, 1998: The operational CMC-MRB global environmental multiscale (GEM) model. Part I: Design considerations and formulation. *Mon. Wea. Rev.*, 126, 1373-1395.

Delage Y., 1997: Parametrizing sub-grid scale vertical transport in atmospheric models under statically stable conditions. *Bound. Layer Meteorol.*, **82**, 23-48.

Douville, H., P. Viterbo, J.-F. Mahfouf, A. C. M. Beljaars, 2000 : Evaluation of the Optimum Interpolation and Nudging Techniques for Soil Moisture Analysis Using FIFE Data, *Mon. Wea. Rev.*, **128**: 1733-1756.

Drusch M., Wood E. F., and Simmer C., 1999: Up-scaling effects in passive microwave remote sensing: ESTAR 1.4 GHz measurements during SGP97. *Geophys. Res. Lett.*, **26**, 879-882.

Drusch M., Wood E. F., Jackson T., 2001: Vegetative and atmospheric corrections for soil moisture retrieval from passive microwave remote sensing data Results from the Southern Great Plains hydrologic experiment 1997. *J. Hydromet.*, **2**, 181-192.

Drusch M., Wood E., Gao H., and Thiele A., 2004: Soil moisture retrieval during the Southern Great Plains Hydrology Experiment 1999: A comparison between experimental remote sensing data and operational products. *Water Resour. Res.*, **40**, W02504, doi:10.1029/2003WR002441.

Drusch, M., 2006: Initializing numerical weather prediction models with satellite derived surface soil moisture: Data assimilation experiments with ECMWF's Integrated Forecast System and the TMI soil moisture data set, *J Geophys. Res.* (submitted)

Entekhabi D., Njoku E., Houser P., Spencer M., Doiron T., Smith T., Girard R., B'elair S., Crow W., Jackson T., Kerr Y., Kimball J., Koster R., McDonald K., O'Neill P., Pultz T., Running S., Shi J.C., Wood E., van Zyl J., 2004: The Hydrosphere State (HYDROS) Mission Concept: An Earth System Pathfinder for Global Mapping of Soil Moisture and Land Freeze/Thaw. IGARS special issue

Ferranti, L. and Viterbo, P., The European summer of 2003: sensitivity to soil water initial conditions. *J. Climate*, (accepted).

Gao H., Wood E. F., Drusch M., Crow W. and Jackson T., 2004: Using a microwave emission model to estimate soil moisture from ESTAR observations during SGP99. *J. Hydromet.*, **5**, 49-63.

- Garand L., 2003: Towards an integrated land-ocean surface skin temperature analysis from the variational assimilation of infrared radiances. *J. Appl. Meteorol.*, **42**, 570-583.
- Giard D. and Bazile E., 2000: Implementation of a new assimilation scheme for soil and surface variables in a global NWP model., *Mon. Wea. Rev.*, **128**, 997-1015.
- Hess R., 2001: Assimilation of screen-level observations by variational soil moisture analysis. *Meteorol. Atmos. Phys.*, **77**, 145-154.
- van den Hurk B., 2002: European LDAS established. *GEWEX Newsletter*, **12**, 9.
- van den Hurk B., Bastiaanssen W., Pelgrum H., Van Meijgaard A., 1997: A new methodology for assimilation of initial soil moisture fields in weather prediction models using Meteosat and NOAA data. *J. Appl. Meteorol.*, **36**, 1271-1283.
- van den Hurk, B. and The H., 2002: Assimilation of satellite derived surface heating rates in a Numerical Weather Prediction Model. KNMI Scientific Report WR 2002-04.
- Jackson T., Hsu A., O'Neill P., 2002: Surface soil moisture retrieval and mapping using high-frequency microwave satellite observations in the Southern Great Plains., *J. Hydromet.*, **3**, 688-699.

Jackson T., Le Vine D., Hsu A., Oldak A., Starks P., Swift C., Isham J., and Haken M., 1999: Soil moisture mapping at regional scales using microwave radiometry: The Southern Great Plains hydrology experiment. *IEEE Trans. Geosci. Remote Sens.*, **37**, 2136-2151.

Jackson T., 1997: Soil moisture estimation using special satellite microwave/imager satellite data over a grassland region. *Water Resour. Res.*, **33**, 1475-1484.

Jones A., Guch I., Vonder Haar T., 1998: Data assimilation of satellite-derived heating rates as proxy surface wetness data into a regional atmospheric mesoscale model. Part I: Methodology. *Mon. Wea. Rev.*, **126**, 634-645.

Jones A., Guch I., Vonder Haar T., 1998: Data assimilation of satellite-derived heating rates as proxy surface wetness data into a regional atmospheric mesoscale model. Part II: A case study. *Mon. Wea. Rev.*, **126**, 646-667.

Kerr, Y.H., Waldteufel P., Wigneron J.-P., Font J., Berger M., 2001: Soil moisture retrieval from space: The Soil Moisture and Ocean Salinity (SMOS) mission. *IEEE Trans. Geosci. Remote Sens.*, **39**, 1729-1735.

Kerr Y. and Njoku E., 1990: A semiempirical model for interpreting microwave emission from semiarid land surfaces as seen from space. *IEEE Trans. Geosc. Remote Sens.*, **28**, 384-393.

Kirdyashev K., Chukhlantsev A., Shutko A. 1979: Microwave radiation of the Earth's surface in the presence of a vegetation cover. *Radio Eng. Electron. Phys.*, **24**, 37-44.

Koster R., Dirmeyer P., and the GLACE team, 2004: Regions of strong coupling between soil moisture and precipitation. *Science*, **305**, 1138-1140.

Larson, W. J. and J. R. Wertz , 1992: Space Mission Analysis and Design. Kluwer Ed., The Netherlands.

Mahfouf J.-F., 1991: Analysis of soil moisture from near-surface parameters: A feasibility study. *J. Appl. Meteorol*, **30**, 1534-1547.

Mahfouf J.-F., Viterbo P., Douville H., Beljaars A.C.M. and Saarinen S., 2000: A revised land-surface analysis scheme in the Integrated Forecasting System. *ECMWF Newsletter*, **88**, 8-13.

Mahfouf J.-F., Brasnett B., Gagnon S., 2004: Description of the Canadian Precipitation Analysis (CaPA) project, available from MSC, MRB/RPN Canada.

Mailhot J., S. Bélair , L. Lefaiivre, B. Bilodeau, M. Desgagne, C. Girard, A. Glazer, Anne-Marie Leduc, A. Methot, A. Patoine, A. Plante, A. Rahill, T. Robinson, D. Talbot, A. Tremblay, P. Vaillancourt, and A. Zadra, 2005: The 15-km version of the Canadian regional forecast system. *Atmosphere-Ocean*.

McNider R., Song A., Casey D., Wetzal P., Crosson W., Carlson T., 1994: Towards a dynamic-thermodynamic assimilation of satellite surface temperature in numerical atmospheric models., *Mon. Wea. Rev.*, **122**, 2784-2803.

Mitchell, K. E., Lohmann D., Houser, P. R., Wood, E. F., Schaake, J. C., Robock, A., Cosgrove, B., Sheffield, J., Duan, Q., Luo, L., Higgins, W. R., Pinker, R. T., Tarpley, J. D., Lettenmaier, D. P., Marshall, C. H., Entin, J. K., Pan, M., Shi, W., Koren, V., Meng, J., Ramsay, B. H., and Bailey, A. A., 2004: The Multi-institution North American Land Data Assimilation system (NLDAS): Utilization of multiple GCIP products and partners in a continental distributed hydrological modeling system. *J Geophys. Res.* **109**:DOI:10.1029/2003JD003823.

Noilhan J. and Planton S., 1989: A simple parameterization of land surface processes for meteorological models. *Mon. Wea. Rev.*, **117**, 536-549.

Noilhan J., Mahfouf J.-F., 1996: The ISBA land surface parameterization scheme. *Global and Planetary Change*, **16**, 145-159.

- Njoku E. and Entekhabi D., 1996: Passive microwave sensing of soil moisture. *J. Hydrol.*, **184**, 101-129.
- Pellarin T., Calvet J.-C., Wigneron J.-P., 2003: Surface soil moisture retrieval from L-band radiometry: a global regression study. *IEEE Trans. Geosc. Remote Sens.* **41**, 2135-2139.
- Rhodin A., Kucharski F., Callies U., Eppel D.P., 1999: Variational soil moisture analysis from screen-level atmospheric parameters: Application to a short-range weather forecast model. *Q. J. R. Met. Soc.*, **125**, 2427-2448.
- Robock A., Vinnikov K., Srinivasan G., Entin J., Hollinger S., Speranskaya N., Liu S. and Namkhai A., 2000: The Global Soil Moisture Data Bank., *Bull. Amer. Met. Soc.*, **81**, 1281.1299.
- Rodell, M., P. R. Houser, U. Jambor, J. Gottschalck, K. Mitchell, C.-J. Meng, K. Arsenault, B. Cosgrove, J. Radakovich, M. Bosilovich, J. K. Entin, J. P. Walker, D. Lohmann, and D. Toll, 2004: The Global Land Data Assimilation System. *Bull. Amer. Meteor. Soc.* **85**, 381-394.
- Rodríguez, E., B. Navascués, J. J. Ayuso, S. Järvenoja, 2003: Analysis of surface variables and parameterization of surface processes in HILRAM. Part I: Approach and verification by parallel runs. *HIRLAM Tech. Rep.*, **58**, 1-58.

- Rutherford, I. D., 1972: Data assimilation by statistical interpolation of forecast error fields, *J. Atmos. Sci.*, **29**, 809-815.
- Seuffert, G., Wilker, H., Viterbo, P., Drusch, M., Mahfouf, J.-F., 2004: The usage of screen-level parameters and microwave brightness temperature for soil moisture analysis. *J. Hydromet.*, **5**, 3, 516-531.
- Trigo, I. 2004: The use of remote sensing data to improve the modeling of skin temperature. *Proc. ECMWF/ELDAS workshop on land surface assimilation. Reading, 8-11 November 2004.*
- Viterbo P. and Courtier P., 1995: The importance of soil water for medium-range weather forecasting. Implications for data assimilation. Workshop on Imbalance of Slowly Varying Components of Predictable Atmospheric Motions, W.M.O., Beijing, China, March.
- Wetzel P., Atlas D., Woodward R., 1984: Determining soil moisture from geosynchronous satellite infrared data: A feasibility study. *J. Appl. Meteor.*, **23**, 375-391.

## List of Tables

Table 1: Microwave radiative transfer vegetation parameters: single scattering albedo ( $\omega$ ), vegetation water content (VWC), and structure coefficient ( $b$ ) for L-band and C-band..

Table 2: Satellite orbital parameters: orbit description (*e.q.t.* is the equator crossing time), altitude  $H$ , inclination (or maximum off-nadir viewing angle for geosynchronous)  $\phi$ , swath width  $\Delta X$ .

Table 3: Screen-level errors for the observations and the model background prescribed in the OI analyses for temperature and relative humidity.

Table 4: Prescribed observation errors (standard deviation) in the soil moisture and soil temperature analyses.

## List of Figures

Figure 1: Hourly BIAS (*dashed line*) and RMSE (*full line*) for GEM screen-level temperature (*a, c, e*) and relative humidity (*b, d, f*) for three days : 1 May (*a, b*), 15 July (*c, d*) and 1 October (*e, f*) 2005. Bars show the number of observations (right axis).

Figure 2: Observation error patterns (standard deviation) for screen-level temperature (*a*,  $T_{2m}$  [K]) and relative humidity (*b*,  $RH_{2m}$  [%]) from OI analyses on 1 May 2005 at 18 UTC.

Figure 3: Land surface water state (total soil moisture, ice and snow water equivalent [ $\text{kg}\cdot\text{m}^{-2}$ ]) on 1 May (*a*), 15 July (*b*) and 1 October (*c*) 2005.

Figure 4: Examples of simulated observations for 1 May 2005: screen-level temperature (*a*), relative humidity (*b*), infrared skin temperature (*c*), microwave brightness temperature (H component) for L-band (*d*), and C-band (*e*).

Figure 5: Information content ( $\mathbf{K}^T\mathbf{H}^T$ ) of the observations in a daily soil moisture analysis assimilating 6-hourly (*a, c, e*) and hourly (*b, d, f*) screen-level observations and satellite observations for three days: 1 May (*a, b*), 15 July (*c, d*) and 1 October (*e, f*) 2005. Bars show the number of assimilated observations (log axis).

Figure 6: Information content ( $\mathbf{K}^T\mathbf{H}^T$ ) of the observations in a daily soil temperature analysis assimilating 6-hourly (*a, c, e*) and hourly (*b, d, f*) screen-level observations and satellite observations for three days: 1 May (*a, b*), 15 July (*c, d*) and 1 October (*e, f*) 2005. Bars show the number of assimilated observations (log axis).

Figure 7: Normalized Information content (% weight in the analysis) of a single observation in a daily soil moisture (gray) and temperature (black) analyses

assimilating hourly screen-level observations and satellite observations (average over three days).

Figure 8: Relative Information content (% weight in the analysis) of the observations in a daily soil moisture analysis assimilating 6-hourly screen-level observations and satellite observations using the atmospheric coupled system (GEM 15 km) and the off-line system (MEC 15 km) for 5 July 2004 (Number of considered points =89223).

Figure 9: RMSE and BIAS for screen-level temperature (*a*,  $T_{2m}$  [K]) and relative humidity (*b*,  $RH_{2m}$  [%]) along a 24-hour off-line integration (before and after a land surface analysis) of  $T_p$  and  $W_p$ . Bars show the number of observations (right axis).

Figure 10: Information content of hourly  $T_{2m}$  (*a*) and  $RH_{2m}$  (*b*) when assimilating one time-slot in a 24-h time window for the daily soil moisture ( $W_p$ , in gray) and soil temperature ( $T_p$ , in black) analyses. Number of assimilated observations each hour ( $\sim 10^5$ ).

Table 1: Microwave radiative transfer vegetation parameters: single scattering albedo ( $\omega$ ), vegetation water content (VWC), and structure coefficient ( $b$ ) for L-band and C-band..

Class of vegetation	$\omega$ [-]	VWC[ $kg\ m^2$ ]	$b(L)$ [ $m^2kg^{-1}$ ]	$b(C)$ [ $m^2kg^{-1}$ ]
grassland	<i>0.04</i>	<i>0.5 LAI</i>	<i>0.1</i>	<i>0.5</i>
crops	<i>0.10</i>	<i>0.5 LAI</i>	<i>0.2</i>	<i>0.2</i>
coniferous	<i>0.10</i>	<i>3</i>	<i>0.1</i>	<i>0.5</i>
broadleaf	<i>0.10</i>	<i>4</i>	<i>0.1</i>	<i>0.5</i>
tropical	<i>0.10</i>	<i>6</i>	<i>0.1</i>	<i>0.5</i>

Table 2: Satellite orbital parameters: orbit description (*e.q.t.* is the equator crossing time), altitude  $H$ , inclination (or maximum off-nadir viewing angle for geosynchronous)  $\phi$ , swath width  $\Delta X$ .

Satellite	<i>Orbit type</i>	$H$ (km)	$\phi$ [°]	$\Delta X$ [km]
SMOS	sun-synchronous (6 am,6 pm e.q.t.)	670	98.0	1000
Aqua	sun-synchronous (1:30 pm e.q.t.)	705	98.2	1445
Goes E/W	Geostationary	36000	8.0 ( <i>max</i> )	-

Table 3: Screen-level errors for the observations and the model background prescribed in the OI analyses for temperature and relative humidity.

Screen-level variable	$\sigma^o$	$\sigma^b$
T2m	<i>1.5 K</i>	<i>2.0 K</i>
RH2m	<i>7%</i>	<i>10%</i>

Table 4: Prescribed observation errors (standard deviation) in the soil moisture and soil temperature analyses.

Observation	$\sigma_o$ (standard deviation)
$Tb_{H,V}(L)$	3.0 K
$Tb_{H,V}(C)$	3.0 K
$Ts(IR)$	3.0 K
$T_{2m}$	2.0 K
$RH_{2m}$	10 %

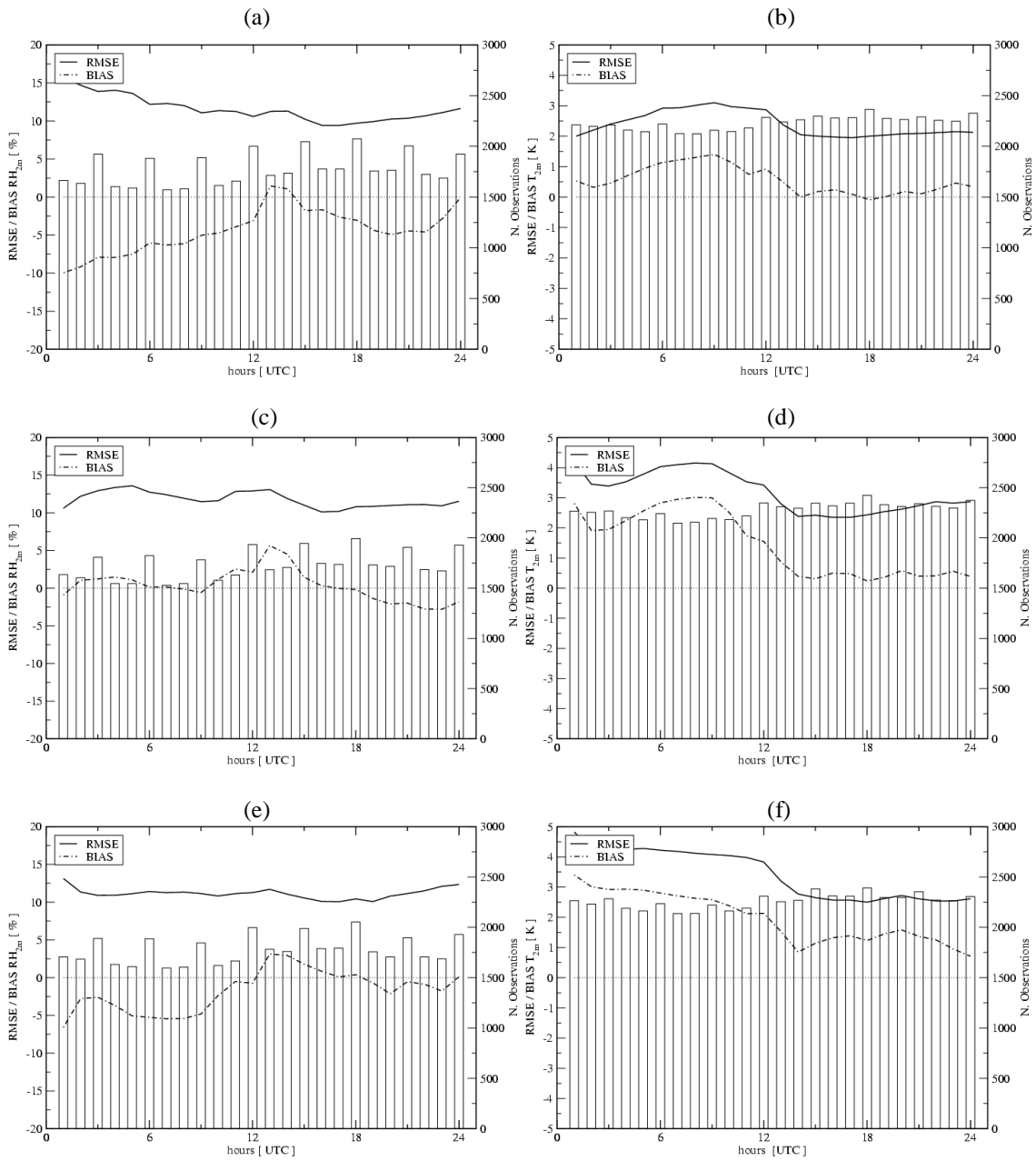


Figure 1: Hourly BIAS (*dashed line*) and RMSE (*full line*) for GEM screen-level temperature (*a, c, e*) and relative humidity (*b, d, f*) for three days : 1 May (*a, b*), 15 July (*c, d*) and 1 October (*e, f*) 2005. Bars show the number of observations (right axis).

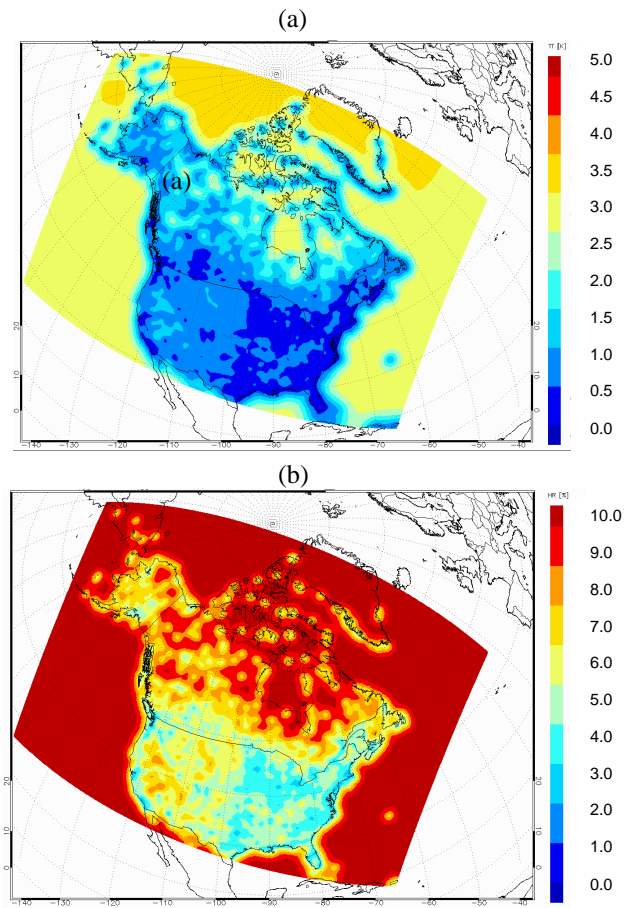


Figure 2: Observation error patterns (standard deviation) for screen-level temperature (*a*,  $T_{2m}$  [K]) and relative humidity (*b*,  $RH_{2m}$  [%]) from OI analyses on 1 May 2005 at 18 UTC.

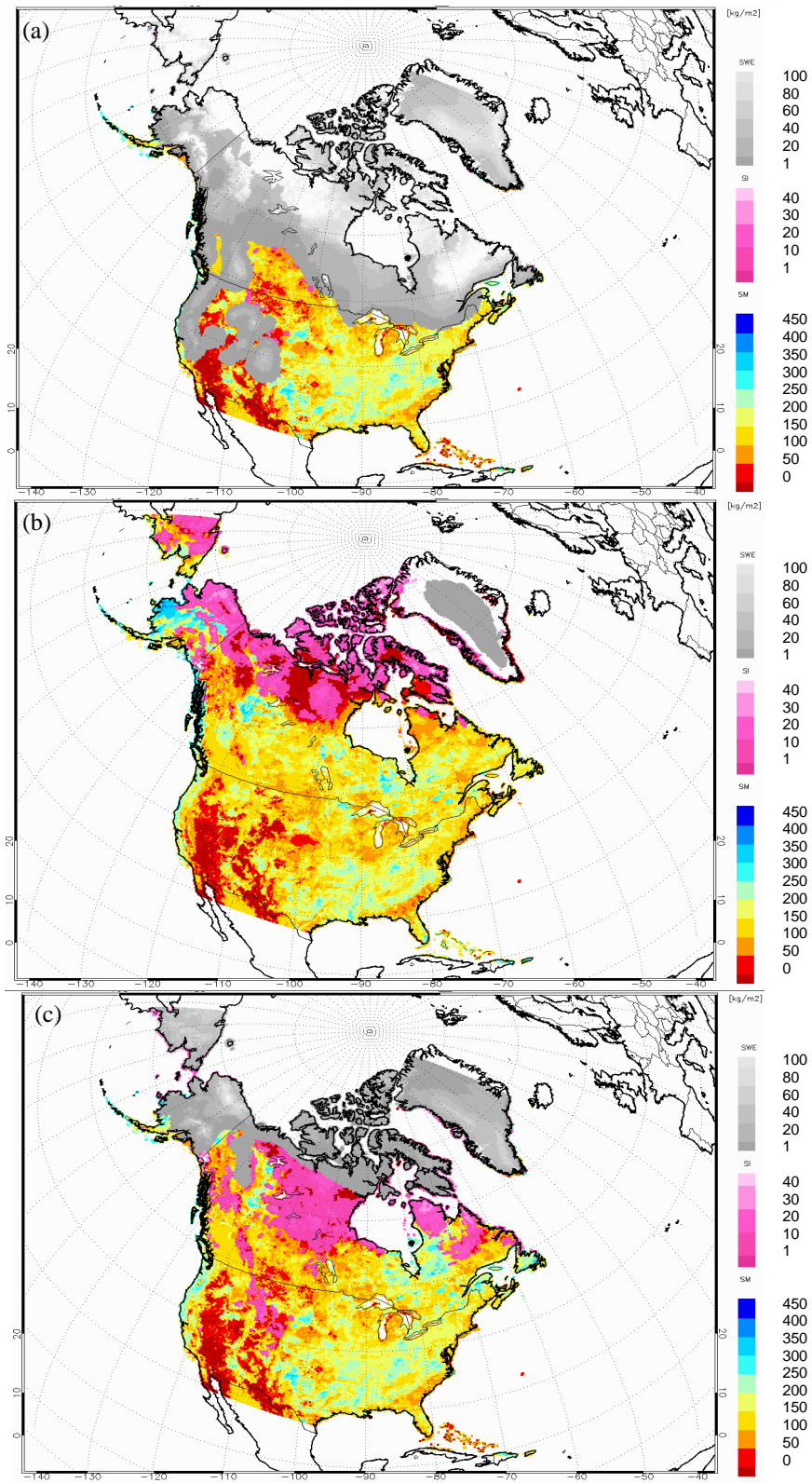


Figure 3: Land surface water state (total soil moisture, ice and snow water equivalent  $[\text{kg}\cdot\text{m}^{-2}]$ ) on 1 May (a), 15 July (b) and 1 October (c) 2005.

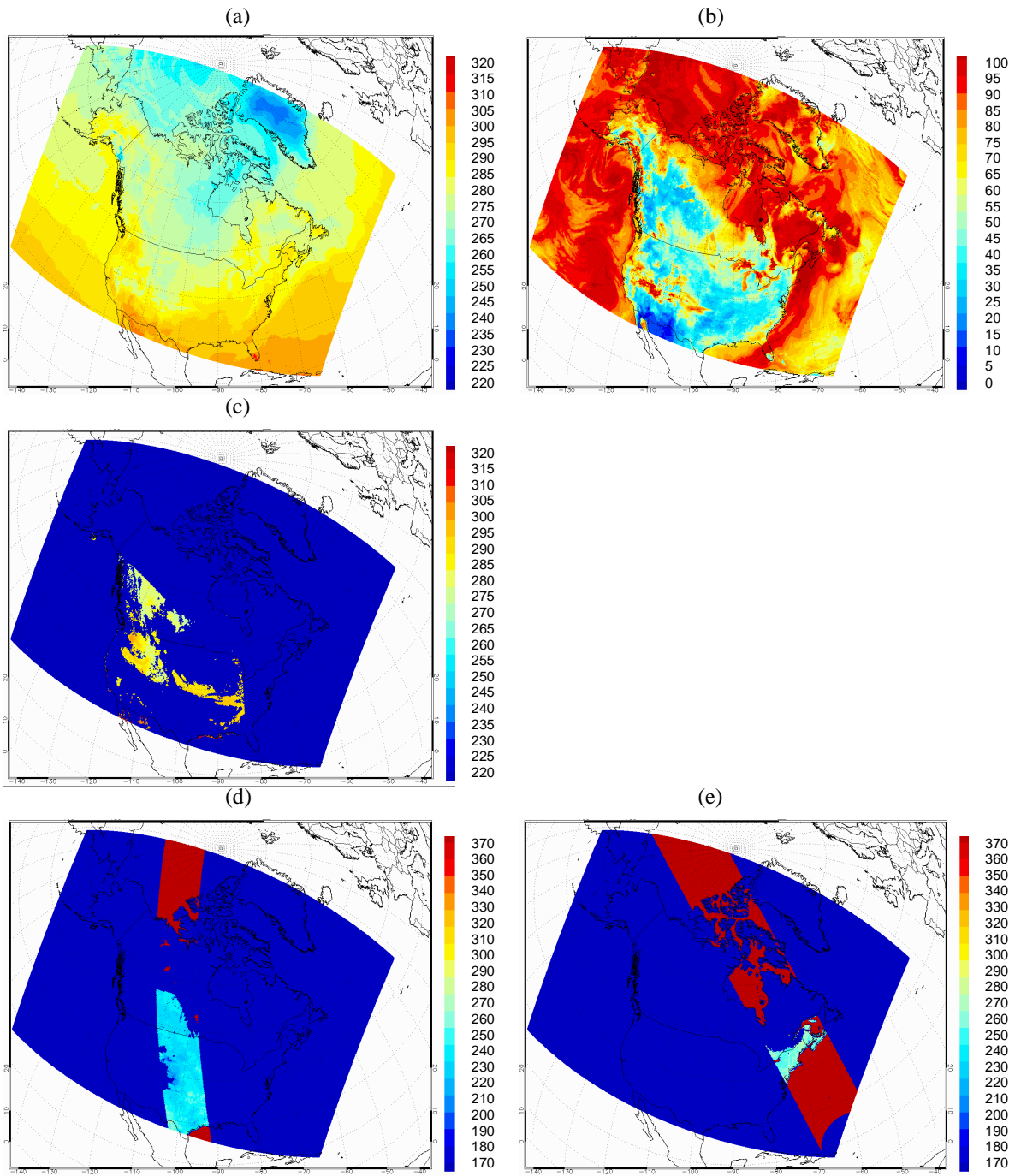


Figure 4: Examples of simulated observations for 1 May 2005: screen-level temperature (a), relative humidity (b), infrared skin temperature (c), microwave brightness temperature (H component) for L-band (d), and C-band (e).

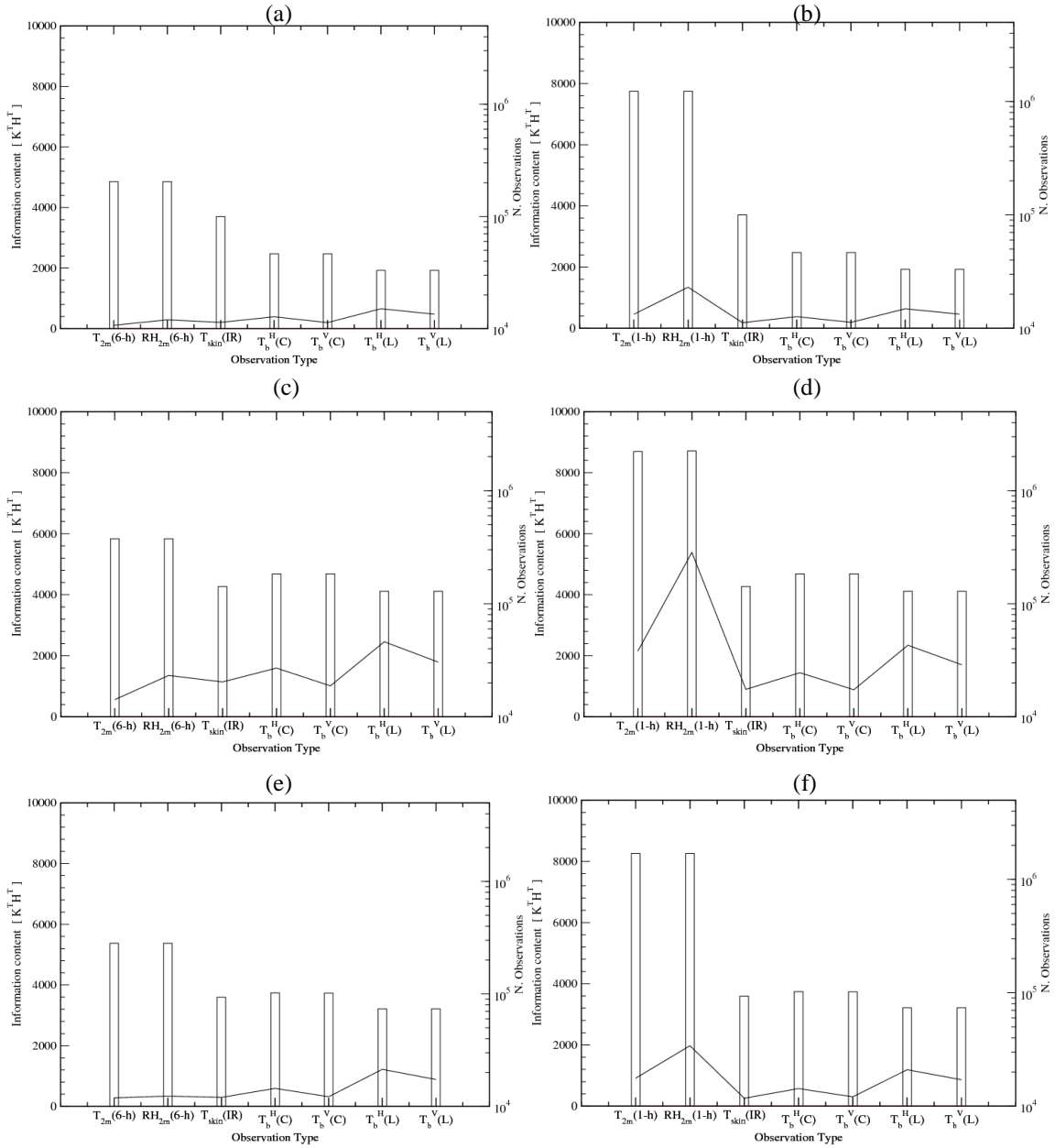


Figure 5: Information content ( $\mathbf{K}^T \mathbf{H}^T$ ) of the observations in a daily soil moisture analysis assimilating 6-hourly (a, c, e) and hourly (b, d, f) screen-level observations and satellite observations for three days: 1 May (a, b), 15 July (c, d) and 1 October (e, f) 2005. Bars show the number of assimilated observations (log axis).

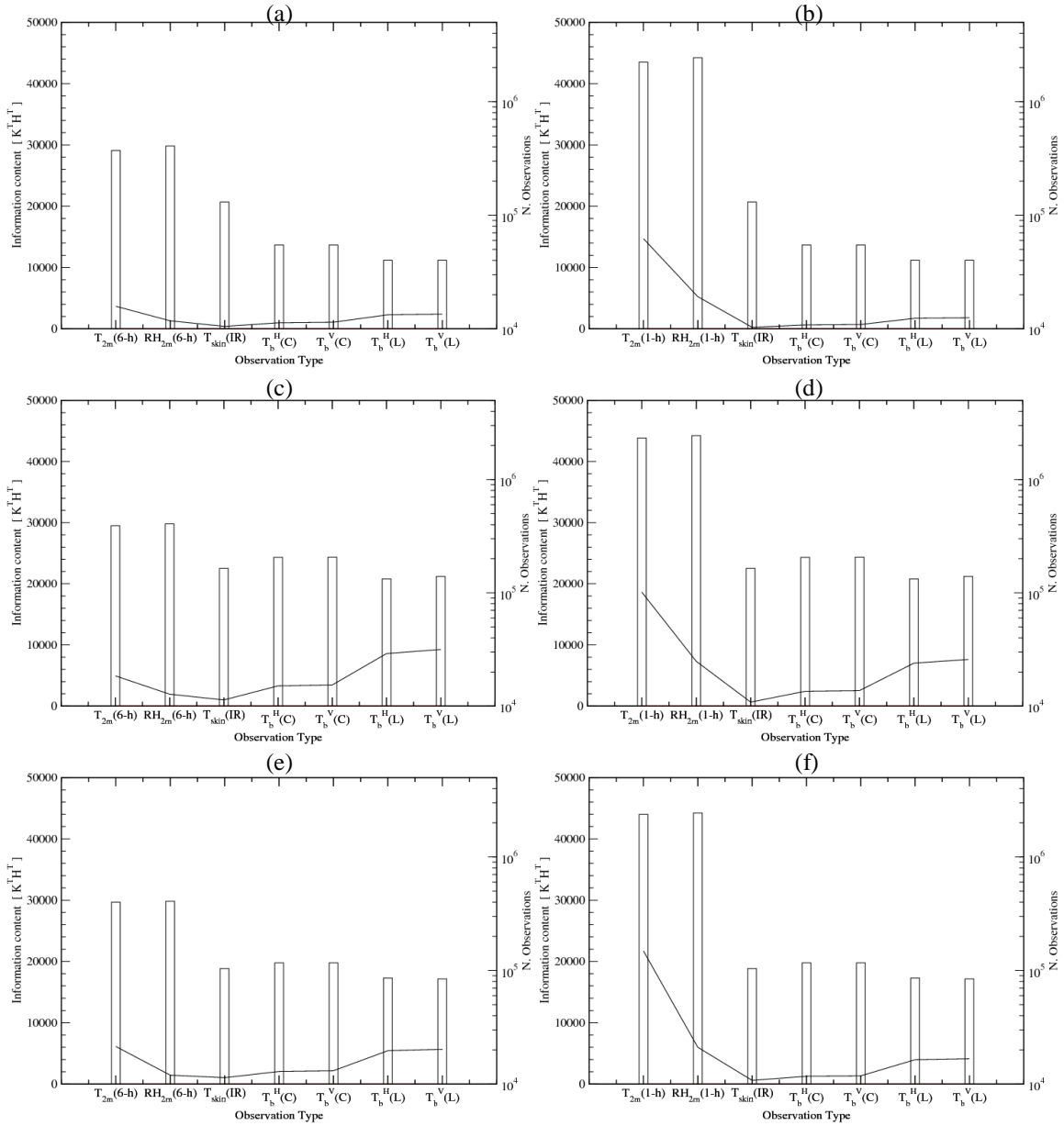


Figure 6: Information content ( $\mathbf{K}^T \mathbf{H}^T$ ) of the observations in a daily soil temperature analysis assimilating 6-hourly (a, c, e) and hourly (b, d, f) screen-level observations and satellite observations for three days: 1 May (a, b), 15 July (c, d) and 1 October (e, f) 2005. Bars show the number of assimilated observations (log axis).

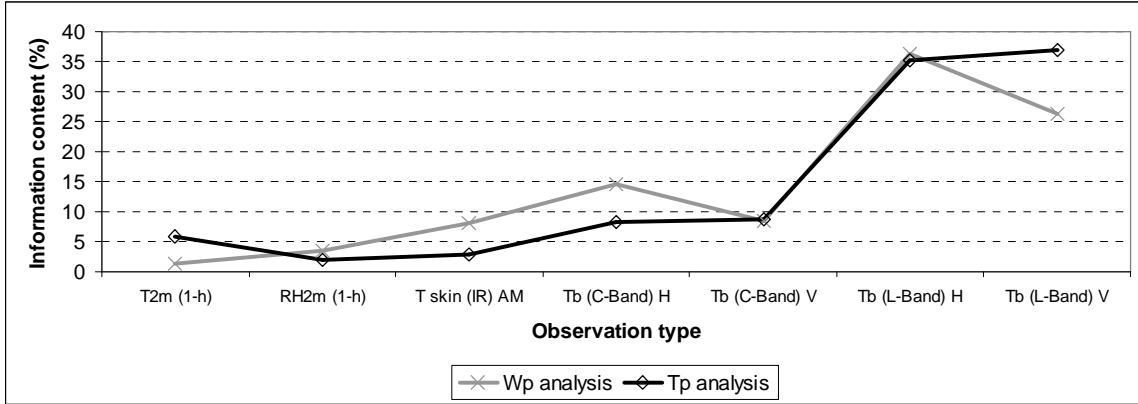


Figure 7: Normalized Information content (% weight in the analysis) of a single observation in a daily soil moisture (gray) and temperature (black) analyses assimilating hourly screen-level observations and satellite observations (average over three days).

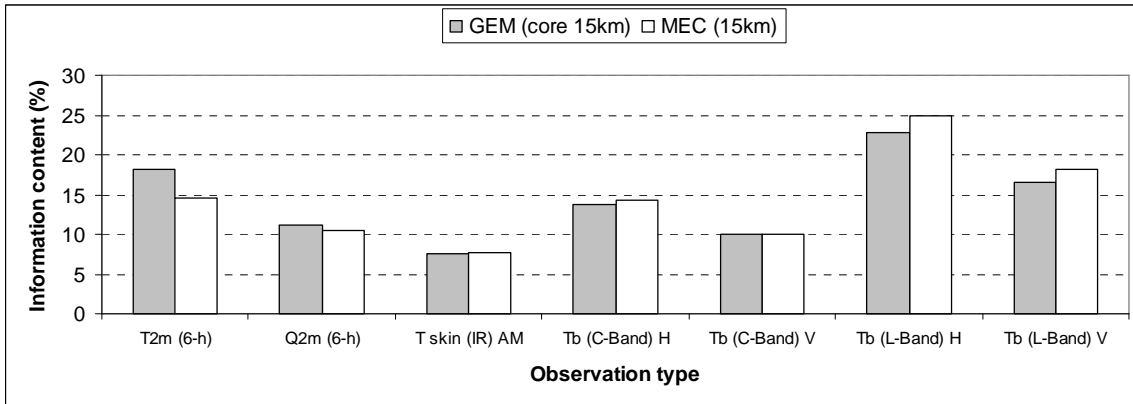


Figure 8: Relative Information content (% weight in the analysis) of the observations in a daily soil moisture analysis assimilating 6-hourly screen-level observations and satellite observations using the atmospheric coupled system (GEM 15 km) and the off-line system (MEC 15 km) for 5 July 2004 (Number of considered points =89223).

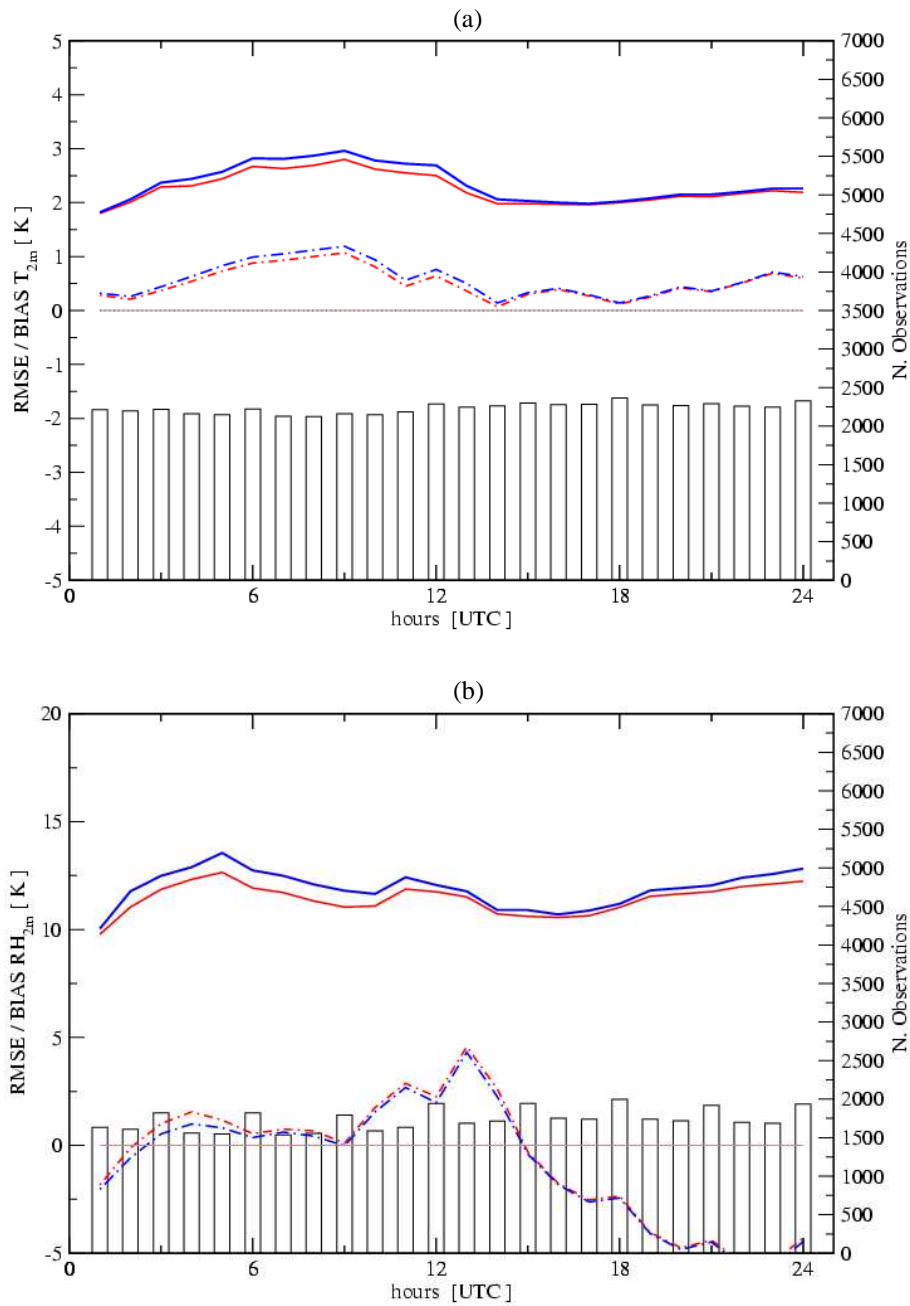


Figure 9: RMSE and BIAS for screen-level temperature (*a*,  $T_{2m}$  [K]) and relative humidity (*b*,  $RH_{2m}$  [%]) along a 24-hour off-line integration (before and after a land surface analysis) of  $T_p$  and  $W_p$ . Bars show the number of observations (right axis).

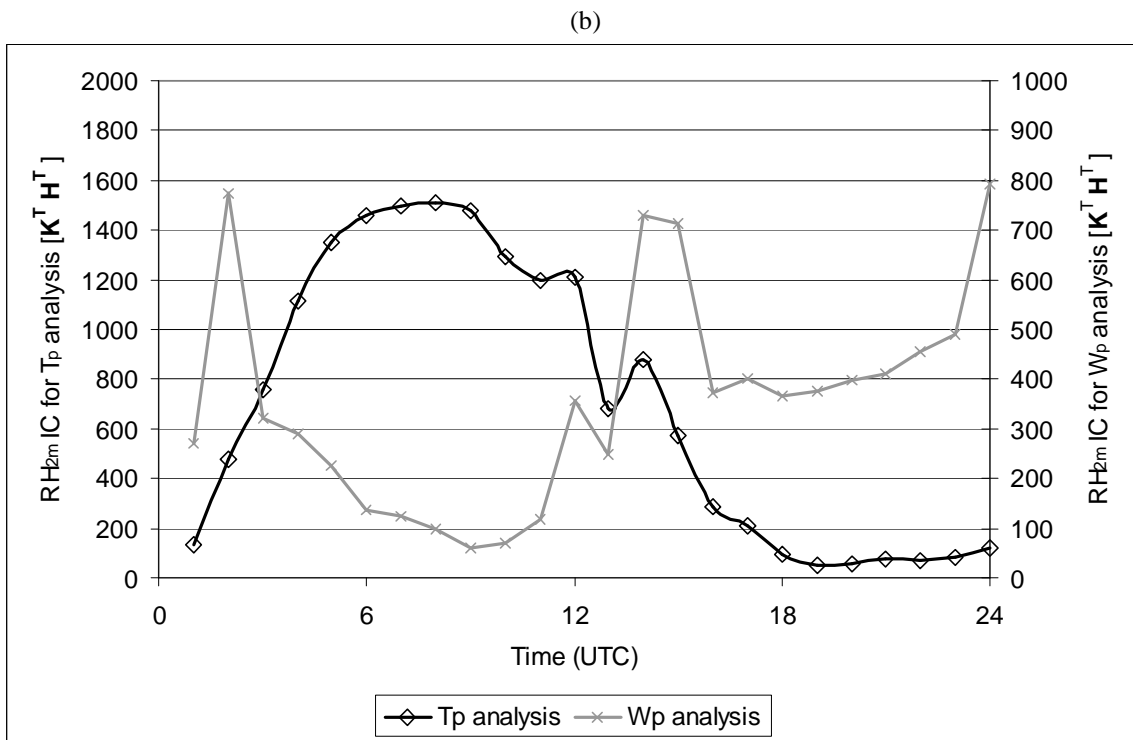
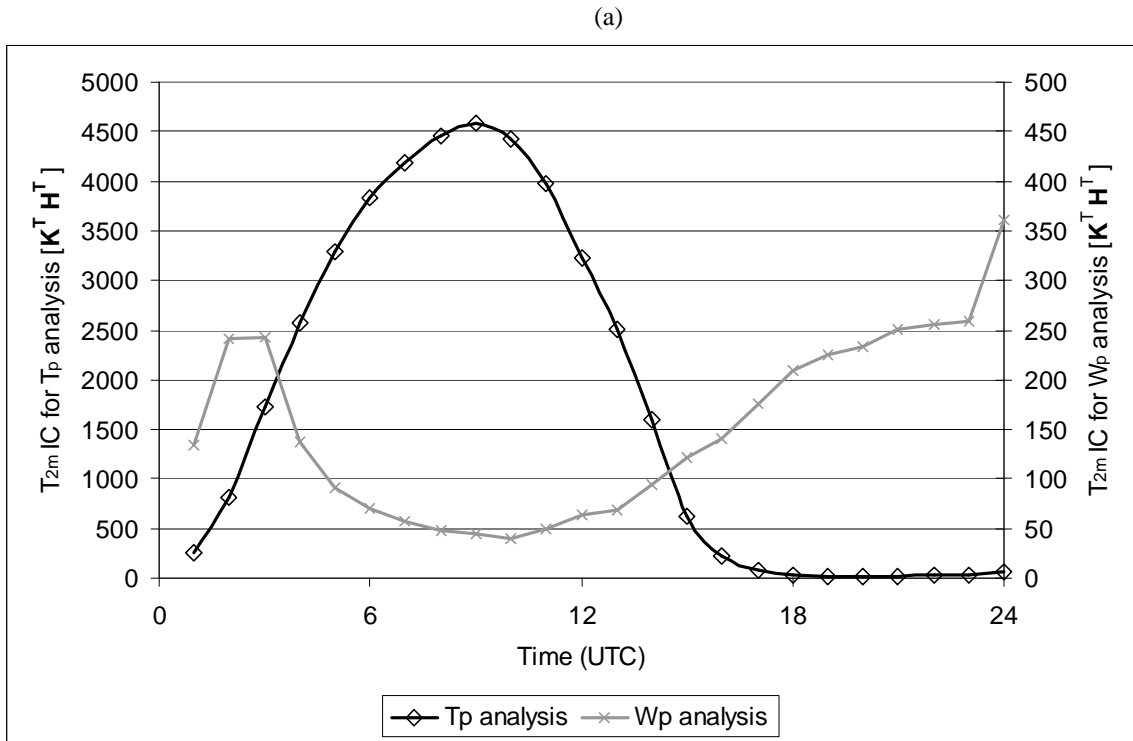


Figure 10: Information content of hourly  $T_{2m}$  (a) and  $RH_{2m}$  (b) when assimilating one time-slot in a 24-h time window for the daily soil moisture ( $W_p$ , in gray) and soil temperature ( $T_p$ , in black) analyses. Number of assimilated observations each hour ( $\sim 10^5$ ).