Soil Field Model Interoperability: Challenges and Impact on Screen Temperature Forecast Skill during the Nordic Winter

JØRN KRISTIANSEN AND DAG BJØRGE
Norwegian Meteorological Institute, Oslo, Norway

JOHN M. EDWARDS AND GABRIEL G. ROONEY
Met Office, Exeter, United Kingdom

(Manuscript received 31 July 2011, in final form 31 January 2012)

ABSTRACT

The high-resolution (4-km grid length) Met Office (UKMO) Unified Model forecasts driven by the coarser-resolution (8-km grid length) High-Resolution Limited-Area Model (HIRLAM), UM4, often produce significantly colder screen-level (2 m) temperatures in winter over Norway than forecast with HIRLAM itself. To diagnose the main error source of this cold bias this study focuses on the forecast initial and lateral boundary conditions, particularly the initialization of soil moisture and temperature. The soil variables may be used differently by land surface schemes of varying complexity, representing a challenge to model interoperability. In a set of five experiments, daily UM4 forecasts are driven by alternating initial and lateral boundary conditions from two different parent models: HIRLAM and Met Office North Atlantic and Europe (NAE). The experiment period is November 2007. Points for scientific examination into the topics of model interoperability and sensitivity to soil initial conditions are identified. The soil moisture is the main error source and is therefore important also in winter, rather than being a challenge only in summer. The day-to-day variability in the forecast error is large with the larger errors on days with strong longwave heat loss at the surface (i.e., the forecast sensitivity to soil moisture content is significant but variable). The much drier soil in HIRLAM compared to NAE reduces the heat capacity of the soil layers and affects the heat flux from the surface soil layer to the surface. Normalizing the respective soil moisture fields reduces these differences. The impact of ground snow is quite limited.

1. Introduction

In numerical weather prediction (NWP) the issue of model interoperability (i.e., starting one model with the fields from another model) is an important topic (Di Giuseppe et al. 2011). The initial atmospheric and land surface conditions are most often estimated from the available observations employing complex and costly data assimilation (DA) systems (e.g., Gustafsson et al. 2001; Simmons 2006; Houser et al. 2010). If the model interoperability approach is successful, however, different NWP models can be initialized from the same analysis, thereby increasing the flexibility and applicability of NWP models in weather forecasting and research, as shown, for instance, in the recent studies by García-Moya et al. (2011), Kristiansen et al. (2011), and McInnes et al. (2011).

Two NWP models are employed in this study. The high-resolution simulations, and the focus of this paper, are run using the nonhydrostatic Met Office (UKMO) Unified Model (MetUM; Davies et al. 2005). MetUM is currently used by several operational forecasting centers (at a range of different resolutions) around the world as well as in the academic research community. At the Norwegian Meteorological Institute (met.no), MetUM is configured with a 4-km grid mesh (UM4) and focuses on the Norwegian mainland. UM4 is configured as a direct dynamical downscaling of the High-Resolution Limited-Area Model (HIRLAM; Undén et al. 2002); that is, both its initial and lateral boundary conditions are taken from the driving model. The quasi-hydrostatic HIRLAM is developed and used operationally by several European countries. At met.no, HIRLAM is configured with a horizontal grid spacing of 8 km (HL8) and relative to UM4 it covers a large geographical domain.
Since 2007 met.no has used the website yr.no (www.yr.no) to freely distribute site-specific weather forecasts. There has been a shift from viewing yr.no as a novelty to expecting high-quality forecasts every day. To this end, both HL8 and UM4 have proven useful. Because of the higher resolution and nonhydrostatic dynamics, UM4 has been a valuable contribution to the prediction of the finer-scale precipitation structures and convection (as shown for polar lows by McInnes et al. 2011 and Kristiansen et al. 2011). UM4 forecasts driven by HL8 have, however, in winter often been significantly colder than forecast with HL8 itself. Figure 1 shows the mean error (measured against observations) as a function of forecast lead time for the winters [November–February (NDJF)] 2008/09, 2009/10, and 2010/11. In UM4 the mean error (solid line) grows almost linearly with lead time. The HL8 forecast bias (dashed line) is, however, small for all lead times. The unsystematic forecast errors [standard deviation of the error (SDE)] are even larger than the systematic errors (not shown). An example of the day-to-day variability in the forecast performance of UM4 is shown in Fig. 2. The forecast error may increase by about 5°C from one day to the next and over a period of a few days the error may gradually increase or decrease by up to 10°C. There also appears to be a dependency on the weather conditions with larger errors on the colder days. The main purpose of this study is to investigate more closely the effect of driving UM4 with HL8 data and identify the error source. To this end, model-consistent data from the Met Office North Atlantic and Europe (NAE) configuration of the MetUM is employed instead of HL8 data to drive UM4. That is, NAE is MetUM with coarser resolution (12-km grid mesh) than UM4.

The NWP models calculate atmospheric variables such as temperature and velocity on discrete model levels. The lowest temperature level is at about 20 (30) m above the surface in UM4 (HL8). The surface temperature is updated by the land surface scheme. The model variables need to be interpolated in the vertical to standard screen-level observation heights. For instance, assuming that Monin–Obukhov similarity theory (Monin and Obukhov 1954) is valid for the surface layer, the screen-level temperatures at 2 m can be diagnosed. The soil properties can significantly impact the screen-level temperature—as well as humidity, low clouds, and precipitation—by influencing the surface fluxes (e.g., Betts 2004). In NWP models, the exchange of heat and water between the land surface and the atmospheric boundary layer is simulated by an interactive surface scheme. Soil observations are still scarce and can vary significantly over short distances (e.g., because of the spatial variability of the soil physical properties, vegetation, and orography) such that measurements made at one location are not representative of the neighboring conditions. Strong local nonlinearities further complicate the interpretation of the relationships between the areal averaged surface fluxes and the corresponding soil field averages, even on km-scale resolutions (Koster et al. 2009). The surface characteristics such as soil conductivity may together with other uncertainties also introduce inaccuracies in the forecasts. To produce the most accurate atmospheric fluxes the soil variables may therefore be carried through the model rather than represent real-world values (e.g., Drusch and Viterbo 2007). For instance, Rooney and Claxton (2006) found that the modeled surface moisture flux became less accurate when observed rather than modeled soil moisture contents were used as input. Differences in absolute soil moisture magnitudes are also found on the global scale when land models are driven (offline) with precisely the same meteorological forcing data but the temporal variability tends to be very similar between the models (Koster...
et al. 2009). Because the initial state of the soil often is known only approximately, it is important to assess the model sensitivity to the associated prognostic variables. Yang et al. (2011) found that the summertime seasonal predictions over the continental United States were quite different depending on the initialization of the land state but the sensitivity was largest (and significant) when the El Niño–Southern Oscillation (ENSO) signal was weak.

Two surface schemes of varying complexity are employed here. The two-layer force–restore method of the Interactions between Soil, Biosphere, and Atmosphere (ISBA) scheme (Noilhan and Planton 1989) is employed by HIRLAM, whereas MetUM uses the more complex diffusive method of the Met Office Surface Exchange Scheme-II (MOSES-II; Essery et al. 2003a). Data assimilation for the land surface has received much less attention than the atmosphere, probably because of a lack of observations of the soil variables. As such, there is an argument for employing simplistic and conservative land surface models. On the other hand, for the NWP models to benefit from their land surface parameterizations, the land surface schemes should realistically represent the land surface processes and parameters (e.g., Wetzel and Chang 1988; Kurkowski et al. 2003). Soil field mapping between different models is difficult to do and one may argue that it should be avoided if possible. Nevertheless, if there exists a soil mapping that works, the benefits could be great. In this study we present the results from a first attempt at mapping the HIRLAM–ISBA soil fields to MetUM–MOSES using a simple approach. The analysis of the corresponding impact on the MetUM screen temperature forecasts should be interesting and informative.

Another goal of this work is to estimate the wintertime forecast sensitivity to soil initial conditions helping other developers of short-range, mesoscale prediction systems to make informed decisions about model interoperability.

After this introduction, section 2 gives an overview of MetUM and HIRLAM with focus on their respective surface schemes. Section 3 presents the experimental design; the different UM4 configurations employed in this study, including the surface field conversion between HIRLAM (source model); and MetUM (target model). The screen temperature forecasts are investigated for the main error source in section 4. The soil moisture initial conditions in HIRLAM and MetUM are compared both as raw fields and after normalization in section 5. Section 6 concludes.

2. Model description

a. UM4

UM4 has since its implementation at met.no been a direct dynamical downscaling of the HL8 forecasts. The met.no configuration UM4 employed in this study is based on the “Ported Unified Model” version 6.1 with 4-km horizontal grid spacing and 38 vertical levels below 40 km. The nonhydrostatic MetUM employs a semi-implicit semi-Lagrangian advection scheme on a rotated latitude–longitude horizontal grid with hybrid-height terrain-following atmospheric levels (Davies et al. 2005). The horizontal grid staggering is the Arakawa C grid while a Charney–Phillips staggering is employed in the vertical. Longwave and shortwave radiative transfer are represented using the Edwards–Slingo radiation scheme (Edwards and Slingo 1996). The scheme of Lock et al. (2000) operates over the 13 boundary levels.

In the stable boundary layer turbulence is represented using a Richardson number scheme. UM4 employs the long-tails scheme [McCabe and Brown 2007, their Eq. (5)] to enhance the vertical mixing and thereby increase the screen temperatures in stable conditions. Other changes to UM4 compared to the reference version is the use of a canopy model to represent canopy snow processes for needleleaf trees (Essery et al. 2003b), a prognostic grid box snow cover, and a decoupled screen temperature diagnostic for stable conditions (Edwards et al. 2011).

Each land grid box consists of any mixture of eight surface types, and the fluxes between the land surface and atmosphere are modeled by MOSES-II (Cox et al. 1998, 1999; Essery et al. 2001, 2003b). The surface temperature is calculated for each tile from the surface energy balance expressed as

\[ C_c \frac{dT_s}{dt} = R_N - H - L_s E - L_f S_M - G_0, \]  

(1)

where \( C_c \) is the canopy heat capacity (zero for non-vegetated tiles); \( H \) is the surface sensible heat flux; \( E \) is the evapotranspiration or moisture flux; \( L_s \) and \( L_f \) are the latent heat of vaporization/sublimation and fusion, respectively; \( S_M \) is the rate of snowmelt; \( G_0 \) is the ground flux; and \( R_N \) is the surface net radiation. Surface temperature is interpreted as a surface skin temperature for nonvegetated and canopy layer temperature for vegetated tiles but for simplicity both are referred to as surface temperature. The surface radiation is

\[ R_N = SW_N + \varepsilon LW + \varepsilon \sigma T^4, \]  

(2)

where \( SW_N \) is the net shortwave radiation and the two last terms are, respectively, the absorbed downwelling and emitted surface longwave radiation components. The emissivity and Stefan–Boltzmann constant are denoted \( \varepsilon \) and \( \sigma \), respectively. The sensible and latent heat fluxes are derived from the mean surface and atmospheric (first model level) values of temperature and humidity using bulk aerodynamic formulae.
The ground flux combines radiative and turbulent fluxes below vegetation canopies and conductive fluxes for the nonvegetated tiles, and is parameterized as a function of the surface temperature and surface soil layer temperature (Best and Hopwood 2001; Essery et al. 2001, 2003b):

\[
G_0 = f_c [e_c e_s (T^4 \text{surf} - T^4) + c_p \Delta T_{\text{can}}(T^* - T_1) + (1 - f_c) \frac{2\lambda}{\Delta z}(T^* - T_1)],
\]

where \( f_c \) is the canopy fraction of the modeled grid box; \( e_c \) and \( e_s \) are the emissivities of the canopy and soil, respectively; \( T_1 \) is the constant soil layer temperature; \( c_p \) is the specific heat capacity at constant pressure; \( \Delta T_{\text{can}} \) is the turbulent exchange coefficient between the canopy and the underlying soil layer; \( \Delta z \) is the thermal conductivity; and \( \Delta z_1 \) is the thickness of the surface soil layer.

MOSES incorporates a multilayer subsurface moisture and temperature model. There are four soil levels of thickness—\( \Delta z_1 = 0.1 \text{ m} \), \( \Delta z_2 = 0.25 \text{ m} \), \( \Delta z_3 = 0.65 \text{ m} \), and \( \Delta z_4 = 2.0 \text{ m} \)—from the surface downward. Soil temperature and moisture content are homogeneous across a grid box. The temperature of the \( n \)th soil layer (\( T_n \)) is updated by the vertical diffusive heat fluxes into (\( G_{n-1} \)) and out of (\( G_n \)) of the layer and the net heat flux (\( J_n \)) advected from the layer by the moisture flux:

\[
C_A \Delta z_n \frac{dT_n}{dt} = G_{n-1} - G_n - \Delta z_n J_n,
\]

where \( \Delta z_n \) is the layer thickness and \( C_A \) is the volumetric heat capacity of the \( n \)th layer (Cox et al. 1999; Essery et al. 2001). At the lower boundary, \( J_4 = G_4 = 0 \). MOSES takes account of the soil moisture (both liquid and frozen) when calculating the thermal conductivity and the volumetric heat capacity and also accounts for the soil moisture freezing and thawing (Essery et al. 2001).

The soil moisture content is updated by a simple mass balance accounting for flux across layer boundaries and evapotranspiration extraction. The latter also takes into account the root density, and the parameterization employed here assumes that the root density decreases exponentially with depth. The snowpack is combined with the surface layer of the soil model rather than being modeled as a distinct layer. Both the thermal conductivity and density of the snow are constants and the heat capacity of the snow is neglected (Essery et al. 2001). However, the increased thickness of the surface soil layer and the different conductivities of snow and soil reduce the thermal conductivity of the combined snow–soil layer.

\( b. \text{NAE} \)

We also employ model-consistent forecasts from the 12-km NAE model as initial and/or lateral boundary conditions for UM4. At the Met Office the principal soil moisture analysis is performed within the global model. Every 6 h the soil moisture is corrected according to errors in forecasts of the screen-level temperature and relative humidity. This nudging scheme is deactivated (i.e., the soil moisture runs freely from one model cycle to the next) in stable conditions and underneath ground snow. The global soil moisture analysis is interpolated to the NAE regional model once each day. The soil temperatures and ground snow in NAE are left to run free except for the surface and first soil level temperatures where an increment is applied if there is no ground snow. Orographic height differences between the NAE and UM4 grids are not accounted for in the interpolation of surface and soil temperatures.

\( c. \text{HIRLAM} \)

HIRLAM is a quasi-hydrostatic gridpoint model whose dynamical core is based on a semi-implicit semi-Lagrangian discretization of the multilevel primitive equations, using a hybrid coordinate in the vertical and a rotated latitude–longitude in the horizontal. In this study we have employed HIRLAM version 7.1.4 (see documentation at www.hirlam.org) with a horizontal grid mesh width of 8 km and 60 levels in the vertical below 10 hPa. The forecasts start from a HIRLAM three-dimensional variational data assimilation (3DVAR) analysis and every 3 h data from the 6-h-lagged European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecast System (IFS; Jung et al. 2010) at the lateral boundaries. The forecast length is 66 h and the model is run four times a day.

To model the land surface processes the force–restore ISBA scheme is employed (Noilhan and Planton 1989). The land surface is a mixture of three types. The soil is divided in two vertical layers: one surface layer, with a depth typically of 1 cm, that responds to the diurnal cycle and a total layer extending down to a depth of about 1 m with a time scale of some days. The lower boundary fluxes are specified using climatological monthly-mean fields of soil temperature and soil moisture content. Optimum interpolation based on screen-level analysis increments is employed to estimate the initial soil water and temperature fields (Mahfouf 1991; Navascués 1997). The analysis of soil water content is switched off on days when the observed screen values would not have been forced by the soil water—for example, on days with ground snow (Undén et al. 2002).
3. Methodology

a. Experiments

The aforementioned met.no-specific model changes described in section 2a improve the model performance of the operational UM4 configuration compared to the default settings. The systematic forecast cold bias is reduced by about 1.5°C at the end of the forecast range \( T + 66 \) (not shown). Still, as displayed in Figs. 1 and 2, respectively, an almost linear error growth remains and the day-to-day variability in the forecast skill is large. Further, the corresponding HL8 forecasts are relatively close to the observations (Fig. 1). This suggests that the main source of error has not been targeted with these changes.

We therefore designed several experiments focusing on the forecast initial and lateral boundary conditions but freezing the model settings of MetUM to that of the operational UM4 (in the following we therefore also refer to UM4 in experiments). To this end, the initial and lateral boundary conditions are alternated between the archived operational forecasts of HL8 and the NAE. The experiments are summarized in Table 1. In experiment ExpHL, the model configuration (model settings and initial and lateral boundary conditions) is identical to the operational UM4. ExpHLNAElbc (ExpHLNAEsurf) is the same as ExpHL but with NAE on the lateral boundaries (soil initial conditions). In the last two experiments, EH and EN, UM4 is driven by NAE but EH is directly using soil initial conditions interpolated from HL8.

The model domain size is \( 220 \times 220 \) grid points, focusing on southern Norway (see Fig. 10 below), and the forecast length is set to 48 h. The study domain is characterized by trees (mostly needleleaf) and grass below the tree line and shrubs and bare soil above. The soil is loamy in a broad region along the coast, whereas loamy sand is found further inland; averaged over the domain, each type is about equally abundant. The model performance is evaluated for daily 0000 UTC reforecasts during November 2007. Because of technical difficulties on 1, 2, 15, 16, and 19 November, the NAE fields were not retrieved. The screen-level temperature forecasts are compared against 42 surface synoptic observations (SYNOP). The 3-hourly observations are compared against 42 surface synoptic observations (SYNOP) point observations ensuring reasonable geographical coverage. The 3-hourly observations are compared against the forecasts at the nearest grid point (e.g., Edwards et al. 2011) but other interpolation methods had only marginal impact on the results. On average the model orography in UM4 is 108 m higher than at the 42 observation sites, and the screen temperature forecasts are adjusted to the station height using a standard atmosphere lapse rate of 0.65 K (100 m)\(^{-1}\). We use standard verification methods for real continuous scalar quantities (e.g., Déqué 2003), mean error (ME), and SDE. Each observation–forecast pair is given the same weight in the error scores. The forecast error is defined as modeled minus observed values.

b. HIRLAM to MetUM soil field conversion

Snow water equivalent (SWE), land surface temperature, soil moisture content, and soil temperature are first interpolated horizontally and then vertically from the HL8 analysis onto the UM4 grid. Based on the soil temperature, the fractions of frozen and unfrozen water contents are determined employing the MetUM code. Soil temperature employs a simple to implement direct linear interpolation from the source (HL8) to the target model (UM4):

\[
T^* = T_s, \\
T_1 = (T_s + T_d)/2, \\
T_2 = T_d, \\
T_3 = (T_d + T_c)/2, \quad \text{and} \\
T_4 = T_c,
\]

where \( T_s \) is the surface layer temperature, \( T_d \) its average over one day, and \( T_c \) is the climatological monthly-mean temperature in the two-layer HIRLAM–ISBA scheme. The MetUM soil temperatures are defined in section 2a above.

The soil moisture mapping is expressed as

\[
M_1 = \rho_w \Delta z_1 \omega_d, \\
M_2 = \rho_w \Delta z_2 \omega_d, \\
M_3 = \rho_w \Delta z_3 \left( \frac{\omega_d + \omega_c}{2} \right), \quad \text{and} \\
M_4 = \rho_w \Delta z_4 \omega_c,
\]

where \( M_n \) is the total soil water content of the \( n \)-th layer in UM4 and \( \omega_d \) and \( \omega_c \) are, respectively, the volumetric soil moisture content (VSMC) of the total soil layer (~1 m deep) and the climatological monthly mean in

<table>
<thead>
<tr>
<th>Experiment</th>
<th>AICs</th>
<th>SICs</th>
<th>LBCs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ExpHL</td>
<td>HL8</td>
<td>HL8</td>
<td>HL8</td>
</tr>
<tr>
<td>ExpHLNAElbc</td>
<td>HL8</td>
<td>HL8</td>
<td>NAE</td>
</tr>
<tr>
<td>ExpHLNAEsurf</td>
<td>HL8</td>
<td>NAE</td>
<td>HL8</td>
</tr>
<tr>
<td>EH</td>
<td>NAE</td>
<td>HL8</td>
<td>NAE</td>
</tr>
<tr>
<td>EN</td>
<td>NAE</td>
<td>NAE</td>
<td>NAE</td>
</tr>
</tbody>
</table>
of soil layer $n$ in MetUM, and $\rho_w$ is the density of liquid water. As will be shown in section 5 below, $\omega_z$, the VSMC of the surface layer (~1 cm deep), varies significantly from day to day and it was therefore not included in the mapping. Numerically, in MetUM, the soil water content and temperature are represented on the model full levels whereas the fluxes are given on the half levels. The force–restore method of ISBA does not require a vertical discretization as such, and the depth of each layer is set according to the characteristic time scales (for NWP) for interacting with the atmosphere above. The interpolation of the atmospheric fields between HIRLAM (source model) and MetUM (target model) is described in Kristiansen et al. (2011) and are not repeated here.

The soil variables could alternatively be normalized (before mapping) to account for soil textural differences between the source and target models (e.g., Di Giuseppe et al. 2011). For instance, the soil wetness index (SWI), a nondimensional index using the wilting point and the field capacity according to soil texture, is linked to evapotranspiration processes and should lead to more similar evaporation fluxes between the source and target models. But during winter periods such a scaling is less relevant, and Di Giuseppe et al. (2011) identified large challenges with the normalizing approach during autumn in southern Europe. Nevertheless, the soil parameter fields are highly similar between HL8 and UM4, resulting in only small corrections when normalizing. Koster et al. (2009) found that differences in model-simulated soil moisture extend beyond those associated with model-specific soil layer thicknesses or texture. It can be argued that $T_s$, as currently determined by Eq. (5), responds too much to $T_s$. An alternative way to set the weights for interpolation might be to assume a temperature wave traveling into the soil, which for a given day would result in a vertical temperature profile. This approach has not been pursued in this study.

4. Screen temperature forecast skill

a. Mean forecast error

Figure 3 shows for all experiments the mean forecast error as a function of lead time. The forecast error growth rate is clearly strongly dependent on the soil initial conditions. When HIRLAM is employed (i.e., ExpHL, ExpHLNAEibc, and EH), the growth rate is about $-0.75^\circ$C day$^{-1}$ and qualitatively similar to that of the operational forecasts shown in Fig. 1. On the contrary, when model-consistent soil fields are employed in experiments ExpHLNAEsurf and EN, the linear error growth is marginal and the forecast error is dominated by its diurnal cycle. Around sunrise the temperatures are on average well predicted whereas the largest errors (excluding the spinup period) occur around sunset (about $-0.5^\circ$C). The coarser grid size in NAE compared to HL8 (12 versus 8 km) results at initial time in larger negative biases in ExpHLNAEsurf and EN. The spatial variability in temperature is greater at the surface than in the atmosphere 30 m above, explaining the even larger error at initial time in EH. The forecast skill is only slightly improved by applying model-consistent atmospheric initial and lateral boundary conditions. Although the model integration domain is relatively large and the screen parameters typically experience a nonnegligible forcing from the land surface below, the latter results could on a given day and location vary with the size and placement of the model domain. In the following, only experiments EN and EH are therefore considered.

b. Daily and spatial distributions of forecast errors

Figure 4 displays the evolution of the error for each of the daily forecasts (thin lines). The mean error is repeated from Fig. 3 (thick black). In EN the standard deviation of the error (green line) is less than 1$^\circ$C for most lead times whereas in EH the spread in forecast skill increases more rapidly and crosses 1$^\circ$C at $T + 24$. (If we include the variability in skill between the individual stations, the SDE is doubled.) In EH a significant contribution to both the systematic and unsystematic errors comes from a few days with very large forecast errors (>5$^\circ$C) whereas in EN the forecasts are never more than 2$^\circ$C too cold.

From Fig. 4, two 4-day composites have been selected (Table 2). The first composite (C1) is performed on 4 consecutive days, 11–14 November, where the forecast skill in EH is particularly low. These days are marked as
blue lines in Fig. 4. The second composite (C2) is performed on days with biases that are smaller than the monthly-mean error and where the forecast skill differs little between EN and EH (red lines). Figure 5 shows the daily-mean forecast error at each of the 42 observation stations for these two composites of forecasts. The results are separated by experiment [EH (Figs. 5a–d) and EN (Figs. 5e–h)], by composite [C1 (Figs. 5a,b,e,f) and C2 (Figs. 5c,d,g,h)], and forecast day [1 (Figs. 5a,c,e,g) and 2 (Figs. 5b,d,f,h)]. The screen temperature forecast errors are small at most stations in C2 [the mean error is less than $2^\circ$ at all but two (five) stations on day 1 (day 2)] and generally similar between EH and EN. The forecasts are typically too warm at the coastal stations and also in the south and east of the domain. The warm bias in the south becomes more prominent on day 2, especially in EN.

In C1, on the other hand, the use of HL8 fields as soil initial conditions has an immediate effect on the UM4 screen temperature forecasts. The forecasts are too cold on most stations, especially in the mountainous region, and the daily-mean error is larger than $3^\circ$ on 19% of the stations (Fig. 5a). When model-consistent NAE soil conditions are used instead, the forecasts come significantly closer to the observed values and for only one station is the daily-mean forecast error larger than $3^\circ$ (Fig. 5e). On day 2, the forecast cold biases in EH generally increase by several degrees to more than $6^\circ$ on 10% of the stations (Fig. 5b). This is in contrast to EN, where the error growth is much smaller (Fig. 5f). Compared to C2, the forecast skill is, however, lower in C1 also for EN. We note that the observed temperatures at forecast initial time are lower in composite C1 than in C2. On the cold days the major, if not the main, contribution to the forecast error is the soil initial conditions interpolated from HL8.

Southern Norway is reliably experiencing snowfall in November and 2007 was no exception. The snow distribution is largely similar in EN and EH (Fig. 6) (i.e., between NAE and HL8). Experiments with different snow depths resulted only in small responses in the screen temperature forecasts (not shown), and a close inspection of the snow distributions and screen temperature forecast skill revealed no causality for correlation between the two (i.e., as indicated in Figs. 2 and 5). Further, in an additional experiment we artificially removed the ground snow in EH (not shown). This reduced the screen temperature mean error but the predictions were frequently too warm and the MAE was only slightly reduced. The fraction of frozen soil moisture is larger in EN than in EH except away from the snowpack. A larger frozen water fraction reduces the heat capacity of the soil layer for a given amount of soil moisture content (Essery et al. 2001).

The accurate prediction of screen temperatures depends heavily on the cloud amount, especially low clouds. The modeled cloud amount in the experiments was generally in good agreement with the National Oceanic and Atmospheric Administration (NOAA) IR images.

#### Table 2. The start dates of the C1 and C2 forecast composites.

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nov</td>
<td>11 Nov</td>
<td>3 Nov</td>
</tr>
<tr>
<td>Nov</td>
<td>12 Nov</td>
<td>21 Nov</td>
</tr>
<tr>
<td>Nov</td>
<td>13 Nov</td>
<td>29 Nov</td>
</tr>
<tr>
<td>Nov</td>
<td>14 Nov</td>
<td>30 Nov</td>
</tr>
</tbody>
</table>
but with a tendency for an underprediction (not shown). The winds were calm during the experiment period and the forecasts underestimated the observations by less than 0.6 m s\(^{-1}\) on average (not shown).

c. Point evaluation

To diagnose the main error source we localize the investigation of the screen temperature forecast error. For legibility, only results from Nesbyen (60°34' N, 9°08' E) are presented but the results are qualitatively similar for other locations. We employ the same 4-day composites as above (C1 and C2). The UM4 grid box holding Nesbyen contains mainly needleleaf trees (70%) with contributions from temperate \(\text{C}_3\) grass (20%) and bare soil (10%). The model represents the real vegetation cover well. Nesbyen is located in a valley, giving shelter from the wind. The wind speeds (model and observations) were mostly below 0.5 m s\(^{-1}\) and only rarely above 1 m s\(^{-1}\) for
the periods studied here. The winds were only slightly stronger in C2 than in C1. Nesbyen is as such a fairly representative station for inland conditions in Norway.

Table 3 presents the soil moisture contents of EN and EH. During winter with ground snow, the soil moisture varies only slowly and the values in the table therefore cover all lead times and both composites. There are large differences between EN and EH. Depending on the soil layer, the soil moisture is 3–10 times larger in EN than in EH. For instance, the 10-cm-thick surface soil layer holds only 8.3 kg of water per unit area in EH whereas in EN the amount is 33 kg m\(^{-2}\). The frozen soil moisture fractions are similar between EN and EH (not shown). Thus, the heat capacity of the soil is much larger in EN than in EH, and for a given net forcing, the soil temperature response would be larger in the drier EH soil than in the moister EN [Eq. (4)]. The local thermal heat conductivity and advective heat fluxes are also influenced by the soil moisture content but both play a lesser role in the evolution of the soil temperatures than the thermal heat capacity.

Compared to the observations (blue dashed) the C1 screen temperatures predicted by EH (blue solid) becomes increasingly colder with lead time and the error is 9°C at \(T + 48\) (Fig. 7a and Table 4). In EN, the screen temperature forecast is, on the other hand, too warm (at most 4°C; Fig. 8a), with daily-mean errors of 1.5°C (Table 4). Even though EH is significantly colder then EN, in both experiments the surface (red line) is slightly colder than the screen level (blue line) at night and warmer during the day. That is, the state of the land surface has a stronger influence

![Fig. 6. Snow depth in (top) EH and (bottom) EN at the (left) beginning and (right) end of November 2007. The shadings depict 0, 1, 5, 10, 50, and 100 cm.](image-url)
on the screen-level temperature than the atmospheric state does. Further, the surface temperature responds relatively quickly to the surface forcing and for lead times larger than about 12 h the screen temperature forecast error is not correlated with the error at initial time.

Higher surface temperatures are associated with larger longwave energy loss at the surface, and on day 1 the average net longwave flux is 49 W m$^{-2}$ in EN compared to 31 W m$^{-2}$ in EH (red line; Figs. 9 and 10). The net shortwave surface fluxes (blue line) and albedos are similar in the two experiments (20 W m$^{-2}$ and 0.14, respectively). Because of the light winds the sensible heat flux is low (e.g., Smith et al. 2010). The nocturnal atmosphere becomes decoupled from the surface and the sensible (orange) and latent (green) heat fluxes fall to very small values. The temperature difference between the surface and the soil is larger (by about 2°C) in EN than in EH, except for the last few hours of the forecast when the difference is small. As a consequence, the daily averaged upward ground heat flux (gray) is much larger in EN than in EH, respectively, 36 and 12 W m$^{-2}$ on day 1. Thus, the ground flux is the major term acting to offset the surface radiative cooling in EN but in EH its daily average is smaller than the solar heating term, giving a daily-mean negative heat flux balance at the surface (black). The results for EN are consistent with the findings in Smith et al. (2010). Studying cold air pools in shallow small-scale valleys, they found that when the sensible heat flux was shut off because of sheltering, the ground heat flux increased to large values.

In EN the two uppermost soil layers have effectively the same temperature (Fig. 8b), and the large upward ground heat flux is in effect the only forcing of $T_1$. Because of the large soil layer moisture content the forcing sensitivity is, however, low, and $T_1$ only starts to decrease on day 2. The soil is drier in EH, and the surface soil layer temperature therefore responds quickly to changes in the ground heat flux (Fig. 7b)—for instance, when the ground heat flux changes direction at midday.
because of the larger heating of the land surface than the underlying soil. Even with this reverse in the ground heat flux and an upward heat transport from the underlying layer, $T_1$ decreases by more than $6\,^\circ C$ over the 2-day forecast period. The soil is colder in EH than in EN but for layers 2–4 the respective intrasoil vertical gradients are about $3\,^\circ C$ in both experiments. The exception is found after the first few hours in EH when the reduced heat capacity is also felt in soil layer 2. Toward the end of day 1, the surface cooling in EH is beginning to influence the atmosphere above and on day 2 the total downward longwave flux reaching the surface is reduced compared to EN. Surface–atmosphere feedbacks complicate the interpretation of the temperature response on day 2. Nevertheless, from the energy balance of the surface soil layer [Eq. (4)] and in Fig. 7a we infer that a wetter soil in EH would have resulted in a warmer soil layer and therefore an increased upward ground heat flux, which, again, would have reduced the screen temperature cold bias and also the atmospheric cooling.

The model performance on composite C2 is presented in Figs. 11 and 12. The screen temperature forecast errors are small initially both in EN and EH. The following observed midday temperature increase and subsequent decrease are respectively under- and overpredicted. In EH the forecast error reaches $-3.1\,^\circ C$ at $T + 21$, which is similar to that of EH–C1. However, the subsequent almost linear error growth found in EH–C1 is not seen (see also Table 4). For EN the screen temperature was generally well forecast with a small cold (warm) bias on day 1 (2) (Table 4). The $-3\,^\circ C$ phase shift between the forecast and observed maximum on day 2 was caused by an underprediction of the midday cloud amount on 4 November.

The surface fluxes in C2 are displayed in Figs. 13 and 14. The latent and sensible heat fluxes are, as in C1, small. The abundant low clouds present in C2 (not shown) reduce the net longwave surface cooling, almost balancing the daily radiative fluxes (within $5\,W\,m^{-2}$ on day 1), and the ground flux is in effect of secondary importance for the evolution of the surface temperature. The soil temperatures are nearly constant between the two composites, except for the surface soil layer where EN (EH) is colder (warmer) in C2 than in C1 (Figs. 11 and 12). The surface is even warmer and the resulting ground heat flux is therefore smaller than in C1 and directed downward at midday both in EN and EH. Still, because of the drier soil, the ground heat flux is, as in C1, smaller in EH than in EN.

Forecast sensitivity to soil moisture may only be a symptom, with the higher soil moisture content and higher heat capacity in EN leading to a correction in the apparent results, not in the physics. The surface heat budget and its components were therefore examined more closely. The sensible heat flux was low at Nesbyen because the winds were very light. Observations at Sodankylä, northern Finland, for the period October–November 2008 show that light winds and strong longwave cooling can lead to ground heat fluxes of $40\,W\,m^{-2}$ or more, even though monthly averages may be only a third of this value. Hence, soil heat fluxes of $20\,W\,m^{-2}$ at Nesbyen (Fig. 10) are not sustained throughout the winter. The ground heat fluxes in Fig. 14, for C2, are most likely much closer to the climatological mean values. (The longwave cooling was stronger...
Many, though not all, periods of high net longwave surface cooling at Sodankylä correspond to high soil heat fluxes, and periods of quite light winds often matched the cold surface temperatures. The Met Office's climate model, which has the same soil model as the UM4, gives both at Nesbyen and Sodankylä climatological soil heat fluxes that are consistent with the sustained magnitudes expected from a simple conceptual soil model and with the observations from Sodankylä. In climate mode the various components of the model's surface flux budget are very much in line with the expected values. The winter mean sensible heat flux (warming) is about one-third of the net longwave radiation (cooling) at Nesbyen and about half at the more northern location of Sodankylä.

In stable boundary layers the height of the model levels may be an issue. Delage (1997) demonstrated that in the quasi-equilibrium stable boundary layer, surface similarity is applicable over reasonably thick atmospheric layers because of compensation between the decrease of the turbulent temperature scale and the local Obukhov length with height. For the choice of stability functions employed in the UM4, Delage finds that errors in surface similarity become important only when the lowest model layer is thicker than 20 m, which is the height of the lowest temperature level in UM4. This compensation is less effective with rapid surface cooling in light winds, and Monin–Obukhov similarity may become invalid in very stable conditions. However, in these cases the sensible heat flux will be small, while...
the screen-level temperature is diagnosed using the decoupled diagnostic.

5. Soil moisture content in HIRLAM and MetUM

a. Spatial distribution and temporal evolution

Units of soil water content are of importance. When considering soil water in units kg m$^{-2}$ as in MetUM, the soil depth is accounted for. By using the layer thickness, the NAE soil water is converted to VSMC (measured relative to the total soil layer volume) and compared to the HIRLAM fields. From the top down the VSMC of the NAE soil layers are denoted $\omega_1$, $\omega_3$, $\omega_2$, and $\omega_4$.

At a given location the monthly-mean NAE $\omega_1$ is up to 40% points greater than HL8’s $\omega_1$; only in a small region in the southwest is $\omega_1$ larger than $\omega_1$ (Fig. 15). The monthly-mean $\omega_1$ is largest along the Norwegian west coast, whereas $\omega_1$ has its peak values inland where we find the deepest snowpack (see Fig. 6). According to Fig. 16, $\omega_4$ is between 1.6 and 4.8 times larger than $\omega_1$ on a given day. The dominant day-to-day variability in $\omega_1$ is largely a result of the thinner soil depth (1 cm). However, it is surprising that the linear trends are of opposite sign over the 10-cm soil layer holding $\omega_1$ and $\omega_4$ (Figs. 16 and 17). The temporal variability in $\omega_1$ is generally dominated by the linear trend and is, as expected, smallest under deep snow. This is contrary to $\omega_4$. In addition, $\omega_4$ varies significantly on the smaller spatial scales.

The moisture content of the HIRLAM deep soil layer ($\sim$1-m depth), $\omega_1$, is next compared to a weighted average of $\omega_4$, $\omega_2$, and $\omega_3$, denoted $\omega_{1m}$. The soil is considerably wetter in NAE (Fig. 15), and from Fig. 16 we find that $\omega_{1m}$ is about twice as moist as $\omega_4$. Figure 15 shows that spatially the largest values are found centrally in the south Norwegian mountains (but with the maximum in $\omega_4$ located more to the east than in $\omega_{1m}$) and also in a region on the southern coast. Elsewhere, there is little spatial consistency between the two fields. The temporal variability as seen in Fig. 16 is, as expected, small. Lastly, we recognize the distribution of the climatological precipitation intensity in $\omega_4$ and $\omega_1$ but mostly in the former. The ratio $\omega_4 / \omega_1$ is large (about 10).

Figure 15 shows that the spatial distributions of VSMC are in a monthly-mean sense largely consistent between all four soil layers in NAE, whereas for HL8, $\omega_4$ differs considerably from $\omega_4$. Since $\omega_4$ represents the fast diurnal variations as opposed to the slower varying $\omega_1$ (time scale of a few days), we would have expected a larger consistency between their monthly means. It is important to resolve these different behaviors in HIRLAM and MetUM. This is potentially a challenging task. The soil moisture content and its variability in a given land model is determined by the model-specific land surface parameterizations, including evaporation, runoff, and infiltration, in addition to soil parameters such as porosity, hydraulic conductivity and wilting point, and horizontal and vertical resolution (Koster et al. 2009). The strength of the land–atmosphere coupling may also be an influencing factor (Yang et al. 2011). A fuller investigation is beyond the scope of this paper but we note from the discussion of the surface energy budget at Nesbyen above that there appears to be significant differences between HIRLAM and MetUM.

b. Soil moisture conversion

Koster et al. (2009, p. 4333) emphasized that the true information content “of a model soil moisture product lies not in its absolute magnitude but in its time variations.” Therefore, a soil moisture scaling could be developed as a basis for mapping soil moisture between different models. A scaling using the climatological mean and variability has been suggested by Koster et al. (2004, 2009). Similarly, the comparison of the VSMC in HIRLAM and MetUM suggests as an alternative to the soil moisture mapping in Eq. (6), normalizing the respective fields before mapping:

$$\tilde{\omega}_x = \frac{\omega_x - \omega_{\text{ref}}}{\text{std}(\omega_x)},$$

where $x$ refers to the indexation of the MetUM and HIRLAM soil layers, and std is standard deviation. The variable transformation is performed at each grid point. With respect to $\omega_4$ it is more appropriate to use its climatological mean rather than to develop a mapping based on $\omega_{1m}$. The climatology and variability in Eq. (7) are here estimated from data for the experiment period only. The following results should therefore for this period be regarded as an upper estimate of the accuracy of the suggested normalization.

The scaled variables are presented in Fig. 18. The scaling generally results in a closer correspondence between HL8 and NAE than without the scaling, especially for $\tilde{\omega}_{1m}$ and $\tilde{\omega}_4$. The scaled fields do, however, retain the trends of the raw data. For example, the largest differences between $\omega_4$ and $\omega_{1m}$ are found at the beginning and end of the 1-month experiment period. Using the average standard deviation, this difference of about 1 scaled unit would give an error of 0.9% points when $\omega_{1m}$ is estimated from letting $\tilde{\omega}_{1m} = \tilde{\omega}_4$. The errors may be larger locally on a given day. For $\omega_4$ and $\omega_1$ the corresponding average error is 3.8% points and the difference between the scaled fields reaches almost 2 scaled units. To illustrate the impact of the different linear trends, we have detrended the VSMC data prior to the scaling (Fig. 19). There is now a quite close relationship between $\tilde{\omega}_{1m}$ and $\tilde{\omega}_4$, and also evidence of a relationship between $\tilde{\omega}_4$ and $\tilde{\omega}_4$, which was not evident.
FIG. 15. The monthly mean of November 2007 in VSMC from (right column and top left) NAE–MOSES and (middle and bottom left) HIRLAM–ISBA. There are four soil levels in MOSES of thickness 0.1, 0.25, 0.65, and 2.0 m from the surface downward. The VSMC of these layers are denoted, respectively, $v_1$, $v_2$, $v_3$, and $v_4$. In addition, the VSMC of the first 1 m is estimated $v_{1m}$ for comparison to the total soil layer (about 1 m) in ISBA. In ISBA, the surface, total, and climatological layers are labeled $v_s$, $v_d$, and $v_c$. Starting at 0, the contours are drawn every 5% except at the climatological depth $v_c$, where the contour interval is 0.25%.
prior to the detrending. The RMS difference between the daily $v_{1m}$ and $v_d$ fields is also small but largest (0.3 scaled units) on the lee side of the dominating mountains in southern Norway, which coincides with the region of overestimation of precipitation by HL8 (not shown).

Although only based on a very short time period, these results suggests that by detrending and normalizing the soil moisture fields it may be possible to develop a proper soil field mapping for the upper 1-m soil layer. If a relationship between the MetUM fields at different levels can be established, $v_{1m}$ can be used to estimate $v_1$, $v_2$, and $v_3$. The mapping comes with a cost as it requires sufficiently long time series, preferably with the most recent (coupled) model versions (e.g., Koster et al. 2009). Still, perfect mapping will most likely never be achieved.

6. Conclusions

We have investigated the performance of UM4 during early winter in southern Norway with respect to the screen-level temperature forecasts. Considering all variables, UM4 performs quite well but in winter the screen temperature forecast errors may grow rapidly with lead time. UM4 is driven by the coarser-resolution HL8. To identify the source of this forecast cold bias, this study focuses on the forecast initial and lateral boundary conditions, particularly the initialization of soil moisture and temperature. The soil variables may be used differently by land surface schemes of varying complexity, representing a challenge to the model interoperability approach and the model performance. In a set of five experiments, daily UM4 forecasts are driven by alternating initial and lateral boundary conditions from two different parent models: HL8 and MetUM NAE.

Because of the shortness of the experiment period, this paper cannot draw too-general conclusions but serves as a reference for other similar studies. We have identified points for scientific examination into the topics of...

FIG. 16. Spatially averaged VSMC as a function of day starting on 3 Nov. The daily 0000 UTC soil initial conditions of HIRLAM–ISBA (dashed) and MetUM–MOSES (solid) are displayed for the respective soil layers; also shown is the VSMC of the top 1 m of soil in MetUM–MOSES, $\omega_{1m}$ (dotted). See text for details.

FIG. 17. Linear trend in (left) $\omega_1$ and (right) $\omega_2$ for November 2007. The contour interval is 0.25% day$^{-1}$ and only the model initial conditions are considered.
model interoperability and sensitivity to the soil initial conditions.

- Soil moisture is important in winter, rather than being a challenge only in summer. Good and consistent short-range forecasts of wintertime screen temperatures therefore require proper initialization of the soil moisture. When driven by HIRLAM, the UM4 screen temperature error growth rate is on average about $20.75 \text{C day}^{-1}$. Driven by model-consistent NAE fields, the error is small.
- The soil in UM4 is several times moister when model-consistent NAE fields are employed instead of converted HL8 fields. This results in large differences in the heat capacity of the soil and subsequently in the ground heat flux. To confirm the significance of the soil moisture content, we repeated the EH experiment but with the initial soil temperatures of EN, and the screen temperature forecast error was only slightly reduced (not shown).
- When driven by HL8, the day-to-day variability in the UM4 forecast error is large, suggesting significant sensitivity to the weather conditions. On days with large longwave heat loss at the surface and low temperatures, the ground flux is the major term acting to offset the surface radiative cooling. On other days the ground heat flux plays a smaller role in the surface energy budget.
- The large differences in volumetric soil moisture content between HIRLAM and MetUM are reduced by normalizing the fields before mapping. A larger dataset needs to be investigated to find whether the suggested soil moisture mapping is robust and whether there are consistent long-term linear trends in the two datasets.
- The impact of ground snow was quite limited for the time period studied. This might be related to the

![Fig. 18. As in Fig. 16 but for the normalized fields using Eq. (10). The climatological fields $\omega_4$ and $\omega_c$ have been omitted.](image1)

![Fig. 19. As in Fig. 18 but the VSMC data are detrended before normalization.](image2)
snowpack being combined with the surface layer rather than being modeled as a distinct layer.

The presence of snow covering the ground and vegetation influences the energy transfer between soil, land surface, and atmosphere. Deeper snow has an insulating effect on the soil and as such may reduce the sensitivity of the surface temperature to the soil temperatures. On the other hand, reduced snow conductivity may have the side effect that the model becomes more sensitive to other model errors. A multilayer snow scheme has been developed within the MetUM (Best et al. 2011). This study has identified topics worth including in its evaluation.

Acknowledgments. We thank George Pankiewicz for encouraging and including this work under the Met Office—met.no collaboration and Glenn Greed for setting up the data feed between the Met Office and met.no. Haldis Berge, Magne Simonsen, and Inger-Lise Aasen Askaotrud are acknowledged for their assistance on several figures. The observation data from Sodankylä are provided by the Finnish Meteorological Institute. We are also grateful for the valuable comments and suggestions made by two anonymous reviewers.

REFERENCES


