atmosphere

MDPI

Article

Physical and Statistical Links between Errors at the Surface, in the Boundary Layer, and in the Free Atmosphere in Medium-Range Numerical Weather Predictions

Stéphane Bélair [1,*], Nasim Alavi [1], Sylvie Leroyer [1], Marco L. Carrera [1], Maria Abrahamowicz [1], Bernard Bilodeau [1], Dragan Simjanovski [2], Dorothée Charpentier [2] and Bakr Badawy [2]

[1] Environment and Climate Change Canada, Atmospheric Science and Technology, Dorval, QC H9P 1J3, Canada
[2] Environment and Climate Change Canada, Canadian Centre for Meteorological and Environmental Prediction, Dorval, QC H9P 1J3, Canada; bakr.badawy@ec.gc.ca (B.B.)
* Correspondence: stephane.belair@ec.gc.ca

Abstract: The adequate representation of interactions between the land surface and the atmosphere is of crucial importance in modern numerical weather prediction (NWP) systems. In this context, this study examines how errors in the planetary boundary layer (PBL) depend on the quality of near-surface prediction over land for medium-range NWP. Two series of 10-day forecasts from Environment and Climate Change Canada (ECCC)'s global deterministic prediction system were evaluated: one similar to what is currently used in ECCC's operational systems and the other with improved land surface modeling and land data assimilation. An objective evaluation was performed for the 2019 summer season in North America, with a special emphasis on three specific areas: northern Canada, the central US, and the southeastern US. The results indicate that the impact of the new land surface package is more difficult to interpret in the PBL than it is at the screen level. The error differences between the two experiments are quite distinct for the three regions examined. As expected, random errors (standard deviations) for air temperature and specific humidity in the PBL are directly linked with their own random errors at the screen level, with correlation coefficients decreasing from a value of one at the surface to values of about 0.2–0.3 a few kilometers above the surface. Less expected, however, is the fact that random errors in the lower atmosphere also strongly depend on changes in air temperature biases at the surface. Warmer near-surface conditions lead to increased random errors for air temperature in the lower atmosphere, in association with the development of the deeper PBL, with greater spatial variability. This finding is of particular interest when evaluating new configurations of NWP systems for implementation in national meteorological and environmental prediction centers.

Keywords: land–atmosphere interactions; numerical weather prediction; planetary boundary layer; near-surface prediction; medium range; bias

Citation: Bélair, S.; Alavi, N.; Leroyer, S.; Carrera, M.L.; Abrahamowicz, M.; Bilodeau, B.; Simjanovski, D.; Charpentier, D.; Badawy, B. Physical and Statistical Links between Errors at the Surface, in the Boundary Layer, and in the Free Atmosphere in Medium-Range Numerical Weather Predictions. *Atmosphere* **2024**, *15*, 1012. https://doi.org/10.3390/atmos15081012

Received: 2 July 2024
Revised: 13 August 2024
Accepted: 19 August 2024
Published: 21 August 2024

1. Introduction

A wide range of studies performed in the last few decades have conclusively demonstrated the substantial impact that the land surface has on the quality of numerical weather prediction (NWP) in the lower atmosphere. Not surprisingly, numerical predictions close to the surface, for instance, at the screen level, located two meters above the surface, are quite responsive to the lower boundary condition provided by land surface schemes, e.g., [1–3]. Gains in near-surface NWP have been reported in association with more realistic surface description fields [4,5], physically more appropriate land surface modeling [6,7], and more sophisticated approaches to land data assimilation, combined with the inclusion of observations from new sensors [8–10].

In contrast, fewer studies have presented the impact of land surface changes on the operational numerical prediction of planetary boundary layer (PBL) and tropospheric characteristics. A few of them have reported substantial improvements in their PBL forecasts that could be attributed to modifications in their land surface packages, i.e., for land surface modeling or land data assimilation. This was the case two decades ago in [11,12] with their inclusion of the Interactions between Soil, Biosphere, and Atmosphere (ISBA) land surface scheme and sequential land data assimilation in Canada's regional deterministic prediction system. More recent results are presented in [9], which show substantial improvements in the lower troposphere (1000–900 hPa layer) due to the implementation of a moderately coupled land–snow–atmosphere assimilation method in the US National Oceanic and Atmospheric Administration (NOAA)'s High-Resolution Rapid Refresh (HRRR) model. In other studies, such as [2,10], the benefits of improving the land surface representation are less conclusive.

Changes in the representation of land surface processes and initial conditions in NWP models can have a considerable influence on the evolution of the surface layer, boundary layer, and, possibly, the free troposphere. These processes and initial conditions are dominant factors in the determination of the amount of heat, water, and momentum transferred between the land surface and the atmosphere. Interactions between the land surface and the atmosphere have been under investigation for a few decades, notably in the context of the Global Land–Atmosphere Coupling Experiment [13,14] and of the Global Energy and Water Exchanges (GEWEX) initiative for Local Land–Atmosphere Coupling (LoCo) [15].

Changes in soil conditions related to either wetness, roughness, or vegetation influence surface turbulent exchanges of momentum, heat, and water with the atmosphere. These fluxes are determining factors controlling the diurnal evolution of the PBL, which, in turn, has some impact on entrainment processes and on the production of clouds and precipitation. This chain of processes has been well described and examined in the context of LoCo-related studies, e.g., [16,17]. How well an NWP model's atmosphere "reacts" and "interacts" with the land surface essentially determines whether a more realistic representation of the land surface will lead to better atmospheric forecasts in the PBL and above, e.g., [18].

Such issues regarding land–atmosphere interactions appear to have played a major role in the objective evaluation of the land surface package in [19]. This package features changes to both land surface modeling and land data assimilation. Its testing in Environment and Climate Change Canada (ECCC)'s Canadian Global Deterministic Prediction System (GDPS) led to substantial improvements in the medium-range prediction of near-surface air temperature and humidity when compared with screen-level analyses.

Although it was not shown or described in [19] or in other recent studies, an upper-air evaluation indicates that the impact of the new land surface package on the PBL is less conclusive, i.e., that it has considerable spatial variability in terms of its effects, both beneficial and detrimental. In the evaluation process toward an implementation proposal of the new land surface package at the Canadian Centre for Meteorological and Environmental Prediction Centre (CCMEP), the decision was made more difficult by these mixed results and by the lack of understanding as to why there was no clear correspondence between the positive results at the surface and those in the PBL.

The main objectives of this study are thus to present this PBL extension to [19]'s objective evaluation against analyses and to determine the physical and statistical links between near-surface and upper-air forecast errors at the medium range. This work is based on the evaluation of NWP forecasts against near-surface and atmospheric analyses and on simple diagnostics documenting the evolution of the boundary layer. The structure of this article is as follows: materials and methods are detailed in the next section, followed by a presentation of the main results (Section 3) and a discussion of important issues (Section 4), ending with conclusions (Section 5).

2. Materials and Methods

In [19], an objective evaluation versus the authors' own screen-level analyses is presented in order to determine the impact of a new land surface package on medium-range deterministic predictions of screen-level air temperature and humidity from ECCC's GDPS system. The GDPS is based on the Global Environmental Multiscale (GEM) atmospheric model, described in [20–23] for its numerical and dynamical aspects and in [24] for its representation of physical processes. The upper-air initial conditions for GEM's global medium-range forecasts are provided by the 4D-EnVar assimilation system presented by [25] and further described in [26,27].

2.1. The OP and NEW Configurations

The land surface aspects modified and tested in [19] include the mapping of land surface characteristics, land surface modeling, and land data assimilation. These modifications account for the differences between two different GEM configurations for global medium-range forecasting: the OP configuration, which is close to the global prediction system currently used at ECCC for its medium-range numerical guidance, and the NEW configuration, which is the same as OP except for changes to the land surface component. The OP and NEW configurations are described in detail in [19]. The following is just a brief summary.

In OP, land surface processes are represented with the ISBA land surface scheme, presented in [11,12]. In that scheme, a two-layer force–restore approach is applied for both soil moisture and surface temperature [28]. The land surface is modeled in NEW with the Soil, Vegetation, and Snow (SVS) scheme, which is based on ISBA but with several additions, such as the inclusion of multiple energy budgets at the surface [7], new hydrology computed on several soil layers with vertical diffusive processes and the production of surface, lateral, and base flows [29], and stomatal resistance determined with the photosynthesis module of the Canadian Terrestrial Ecosystem Model (CTEM) [30,31].

In addition to the more sophisticated land surface modeling provided by SVS, substantial changes were made to the production of the land surface initial conditions for surface temperature, soil moisture, and snow. The initial conditions for ISBA are provided in OP by a sequential optimal interpolation (OI) of screen-level observations, introduced by [32] and further described in [12]. In NEW, the SVS initial conditions are produced by the Canadian Land Data Assimilation System (CaLDAS) [33,34]. In CaLDAS, a one-dimensional ensemble Kalman filter (1D-EnKF) is applied for soil moisture and surface temperature. A simpler approach is used for snow depth based on an ensemble of OI analyses. The precipitation forcing for the model's first guess comes from the Canadian Precipitation Analysis System (CaPA) [35,36]. Space-based L-band brightness temperature observations from the Soil Moisture Active Passive (SMAP) and the Soil Moisture Ocean Salinity (SMOS) missions are directly assimilated in CaLDAS for soil moisture. Land surface retrievals from the Atmospheric Infrared Sounder (AIRS), CRoss-track Infrared Sounder (CRIS), and Infrared Atmospheric Sounding Interferometer (IASI) are assimilated for surface temperature [37].

Other differences between OP and NEW are related to the mapping of land surface characteristics. In NEW, more recent databases are used to specify the surface properties associated with vegetation and soil texture.

2.2. Surface Fluxes and PBL in GEM

For both ISBA and SVS, the fluxes of heat, moisture, and momentum between the land surface and the atmosphere are described in [28,38]. These formulations are based on classical aerodynamic equations with coefficients for thermal and momentum turbulent exchanges obtained from Monin–Obukhov stability theory [39–41]. The surface fluxes or the coefficients are spatially aggregated to the grid scale and are provided as a lower boundary condition to the one-dimensional vertical diffusion scheme representing atmospheric turbulence, which is based on eddy coefficients obtained from a 1.5-order turbulent kinetic energy (TKE) closure [38,42,43]. The representation of PBL turbulent mixing is the

same for the OP and NEW experiments. In NEW, the effect of subgrid-scale orography is considered with [44]'s form drag, a dynamic thermal roughness length is used, the value of the minimum Monin–Obukhov length has been modified, spatial filtering is applied to the aggregated surface exchange coefficients, and an effective land surface state is estimated for the flux calculations.

2.3. Vertical Profiles in GEM

An objective evaluation at the land surface is performed in the same way as in [19] based on a comparison of model forecasts with own analyses of air temperature and specific humidity at the screen level. In the atmosphere, the evaluation is directly performed on the GEM computational levels. As presented in [22], GEM uses a hybrid coordinate of a log hydrostatic pressure type. This ζ coordinate is written as

$$\ln \pi = \zeta + B_d s \tag{1}$$

in which π is the hydrostatic pressure, B_d is a vertical decay function, and s is to be determined from the boundary condition at the surface, i.e.,

$$\ln \pi_s = \zeta_s + s \qquad \text{or} \qquad s = \ln \pi_s - \zeta_s \tag{2}$$

The ζ vertical coordinate is discretized as $\zeta = \ln\left(\eta p_{ref}\right)$, where p_{ref} is a reference pressure, taken as 1000 hPa, and η is user-defined and varies from a low value (at the model top) to unity (at the model surface). Therefore, at the surface, $\zeta_s = \ln(p_{ref})$. In the present study, vertical profiles of errors and correlations are displayed on the model η levels, which can be approximated near the surface by the ratio of atmospheric pressure to surface pressure in a 15 km atmospheric model, such as the one used in this study.

As a rule of thumb, a vertical variation of 0.1 for the η hybrid coordinate approximately corresponds to differences of 100 hPa in pressure and 1 km in height.

The main advantage of performing the objective evaluation directly on model levels is that no vertical interpolation is needed, taking advantage of the fact that the η levels and orographic height are identical in the model forecasts and in the 4D-EnVar upper-air analyses for both OP and NEW. It should be recognized, however, that the height and pressure are not identical, since the surface pressure is not the same. These random variations in the model levels' height and pressure remain small and can be considered part of the model forecast errors. This additional error is similar to the uncertainty associated with vertical interpolation to constant pressure or height levels typically performed in objective atmospheric upper-air evaluations (e.g., see [45]).

2.4. Evaluation Metrics and Diagnostics

The error bias and standard deviation are estimated for each land grid point as a function of GEM's prediction lead time based on matched time series over the entire list of cases available for the evaluation period. This point-by-point approach allows for the detailed mapping of forecast errors at the screen level and in the PBL and is most appropriate for investigating the statistical links between these errors. Numerical predictions for OP and NEW are directly compared against their own analyses on the same computational grids:

$$STDE_{i,t}^2 = \frac{1}{N_d} \sum_{d=1}^{N_d} \left[(f-a)_d - \overline{(f-a)} \right]_{i,t}^2 \tag{3}$$

$$B_{i,t} = \frac{1}{N_d} \sum_{d=1}^{N_d} \left[(f-a)_d \right]_{i,t} \tag{4}$$

in which $STDE$ is the error standard deviation, B is the bias, d indicates the validation dates, N_d is the number of cases (validation dates), i is model grid points, t is the forecast

lead time, and f and a are the model forecasts and analyses being compared. The overline in Equation (3) represents time averaging.

The evaluation is performed at the near-surface versus screen-level analyses and in the atmosphere for each hybrid (η) GEM level versus 4D-EnVar analyses. Maps of errors are produced, as well as vertical profiles for spatially averaged errors over a few selected areas. Since the GEM results are evaluated here with a latitude–longitude mapping, spatial averages feature normalization based on latitude, as in [19]:

$$\delta_{area,t} = \frac{\sum_{area} \delta_{i,t} \cos \phi_i}{\sum_{area} \cos \phi_i} \tag{5}$$

where δ stands for the spatially averaged error metric, *area* is for the region over which spatial averaging is performed, and ϕ is the latitude of each model grid point i.

To understand the links between errors at the surface and in the lower troposphere for the OP and NEW experiments, Pearson correlation coefficients between STDE error differences in the lower troposphere and error differences at the screen level are displayed as vertical profiles and histograms based on spatial information for specific forecast ranges.

Further physical understanding is provided by diagnostics for the PBL diurnal evolution. The PBL height is calculated at every grid point for a few selected areas for all cases during the evaluation period, and then the values are averaged spatially over specific areas and temporally for several forecast days. The approach is based on the vertical profiles of potential temperature (θ), defined as

$$\theta = T\left(\frac{p_s}{p}\right)^{R_d/c_p} \tag{6}$$

in which T is the air temperature, the surface pressure p_s is used as a reference pressure, R_d is the gas constant for dry air, and c_p is the specific heat capacity at constant pressure. In this algorithm, the model η levels are examined one by one, starting from the model top, to detect an inversion for θ in the lowest few kilometers of the atmosphere. If no inversion is found, a reference is set a few levels above the model surface. The first inflection point identified above the detected inversion (or reference level) is then diagnosed as the PBL height. The parameters involved in this simple method were optimized through tests and trials with the visual inspection of several tens of vertical θ profiles. The objective of these PBL diagnostics is not to provide a rigorous estimate of the PBL height (which is always a difficult task) but rather to serve as a common basis for comparison between the OP and NEW experiments.

2.5. Experimental Setup

The experimental setup is identical to what is described in [19], except that only the results from the summer season are evaluated in the present study. The evaluation period is from 16 June 2019 to 31 August 2019, with a series of OP and NEW 10-day forecasts launched every 12 h each day, at 0000 UTC and 1200 UTC. Only the 0000 UTC runs are evaluated here, similar to the procedure followed in [19].

For NEW, the spinup of SVS prognostic variables (soil moisture and surface temperature) is achieved with an offline open-loop run from October 2017 to 15 May 2019. This is followed by a one-month offline assimilation cycle with CaLDAS, all the way to 16 June 2019, which is the first day of the summer evaluation period when the land surface and upper-air data assimilation systems are weakly coupled. For OP, no special spinup is necessary since the initial conditions from ECCC's operational GDPS cycles are used.

3. Results

In this section, medium-range prediction errors of air temperature and specific humidity at the screen level and in the PBL are presented and discussed for both the OP and NEW experiments. These errors and their differences between the two experiments are displayed as maps and vertical profiles over North America and over three specific areas. The diurnal

cycle of PBL errors is investigated, and correlations between error differences in the PBL and at the screen level are described.

Figure 1 shows maps of medium-range forecast errors at the screen level and at hybrid level $\eta = 0.85$ (approximately 1.5 km above the surface) for the air temperature and specific humidity predicted by the NEW experiment. Forecast errors are temporally averaged for summer 2019 forecasts valid between 78 h and 144 h based on model outputs and analyses every 6 h. The air temperature forecast error map at the screen level (Figure 1's upper-left panel) displays larger errors in the central US and the Canadian prairies (on the order of 2.5 K), in northern Canada (with values close to 3.0 K in northern Yukon), and in the southeast US. For specific humidity (upper-right panel), screen-level errors are larger over the US's central and southern areas, with values on the order of 1.5 g kg^{-1}.

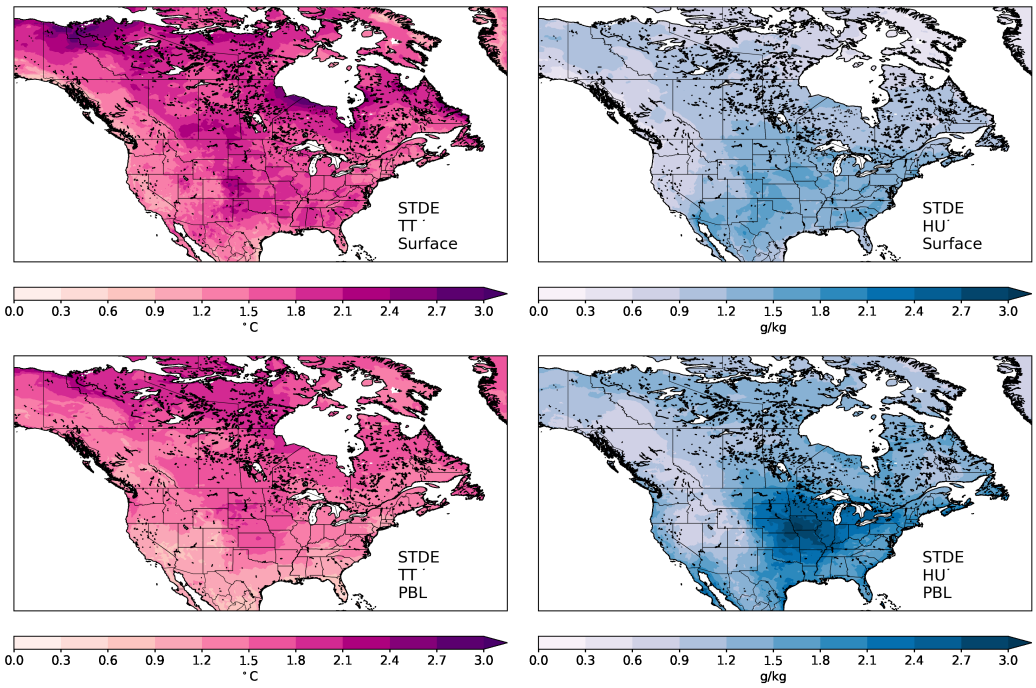

Figure 1. Maps over North America of NEW's STDEs at screen level (**upper panels**) and at hybrid level $\eta = 0.85$ (**lower panels**) for air temperature (K, with red shading, on the left) and for specific humidity (g kg^{-1}, with blue shading, on the right). Forecast errors are temporally averaged for summer 2019 forecasts valid between 78 h and 144 h based on model outputs and analyses every 6 h, i.e., at 0000 UTC, 0600 UTC, 1200 UTC, and 1800 UTC.

In the boundary layer, the spatial patterns of air temperature errors at level $\eta = 0.85$ are similar to what is found at the screen level (cf., Figure 1's left panels). These errors are generally smaller at $\eta = 0.85$ than at the screen level for air temperature, except for areas in northern Canada (i.e., Territories and Yukon), where the amplitude is similar. For specific humidity, the errors' spatial distribution at $\eta = 0.85$ is substantially different from that at the screen level. Large errors are detected over the US just southwest of the Great Lakes, with values greater than 2.5 g kg^{-1} over certain areas. Forecast errors for specific humidity are generally larger at $\eta = 0.85$ than at the screen level, especially over the central-eastern portion of the US. Interestingly, it seems that the maximum errors for specific humidity at $\eta = 0.85$ are slightly shifted east compared with those for air temperature at the same level (see lower panels of Figure 1).

Maps of the STDE differences between NEW and OP for the same period and forecast ranges are shown in Figure 2, both at the surface and at level $\eta = 0.85$. Forecast error differences are temporally averaged for summer 2019 forecasts valid between 78 h and 144 h based on model outputs and analyses every 6 h. The results at the screen level (Figure 2's upper panels) are consistent with [19]. For screen-level air temperature, the

STDE is generally smaller with NEW compared to OP. The largest impact is found over Canada's northern areas, but the air temperature STDE is also decreased at the screen level over the US's western and central regions. There is a smaller increase in STDE with NEW over the southeastern US and southern Canada. For specific humidity, the spatial patterns for STDE differences between the two experiments at the screen level are similar to those for air temperature, with a substantial error decrease over northern areas and the western-central US and an STDE increase over the southeastern US.

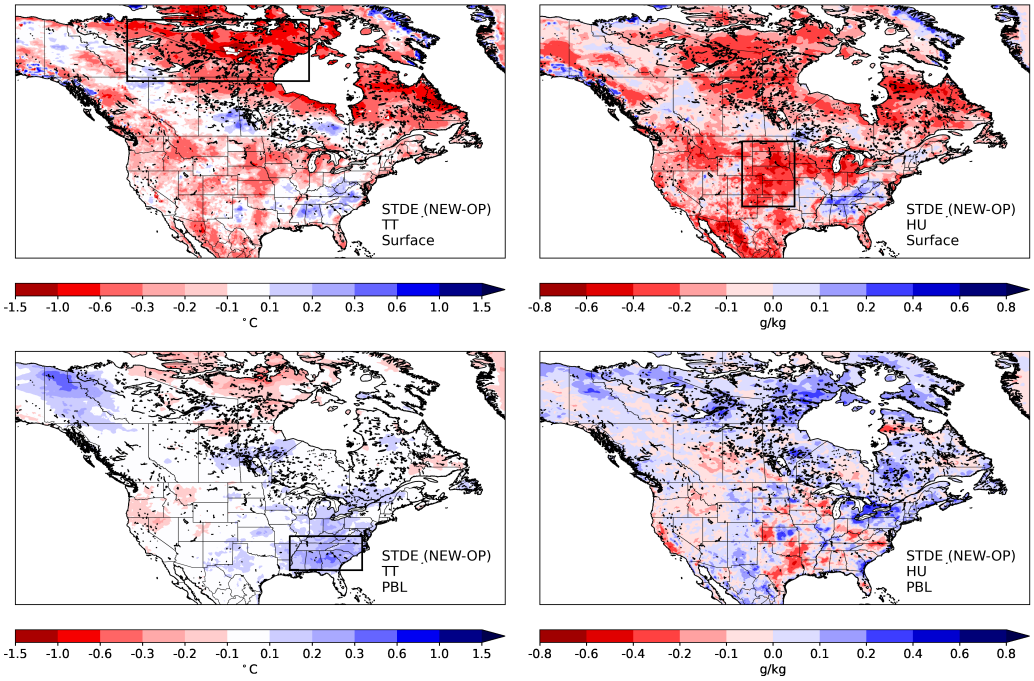

Figure 2. Maps over North America of STDE differences between NEW and OP at the screen level (**upper panels**) and at hybrid level $\eta = 0.85$ (**lower panels**) for air temperature (K, on the left) and specific humidity (g kg^{-1}, on the right). The forecast error differences are temporally averaged for summer 2019 forecasts valid between 78 h and 144 h based on model outputs and analyses every 6 h, i.e., at 0000 UTC, 0600 UTC, 1200 UTC, and 1800 UTC. Areas in red indicate smaller errors for NEW, and the reverse for areas in blue. The black rectangles indicate areas used in the figures below for spatial averaging, which are referred to as "North Canada", "Central US", and "Southeast US".

The impact of NEW vs. OP is quite different at level $\eta = 0.85$ (Figure 2's lower panels). Overall, the differences between the two experiments are smaller at this level. Large portions of the lower-left panel for air temperature at $\eta = 0.85$ are white, indicating small differences. For air temperature at this level, the largest impact is related to an STDE increase for NEW over the southeastern US and over the northern portion of Yukon. The STDE differences for specific humidity at $\eta = 0.85$ are different than for air temperature. These differences are greater than for temperature at this same level and are mostly increased with NEW, except for some areas in the central and southern US and also, to a lesser degree, in the Canadian prairies (southern Alberta and Saskatchewan).

Vertical profiles of STDEs for OP and NEW are shown in Figure 3 for spatial averages over North America and over the specific areas highlighted by the black rectangles in Figure 2 over northern Canada, the central US, and the southeastern US. The profiles are temporally averaged for forecasts valid every 6 h for ranges between 78 h and 144 h. For air temperature over North America, both experiments have the STDE increasing upwards from the surface (level $\eta = 1.0$) to reach its maximum values just below level $\eta = 0.95$ and then decreasing above (see upper-left panels in Figure 3). The STDE errors are smaller for NEW compared with OP near the surface, all the way up to approximately level $\eta = 0.92$ (i.e., nearly up to 1 km above the surface); these errors are slightly larger around level

$\eta = 0.85$, which is the level used for the PBL maps in Figures 1 and 2. The values of STDE at the surface are approximately 2.1 K for NEW and 2.3 K for OP. For specific humidity, the STDE is relatively constant near the surface up to about level $\eta = 0.90$ (lowest km) and then increases to reach its maximum values around level $\eta = 0.80$. NEW has a smaller STDE compared with OP for the lowest layer (below $\eta = 0.90$), but it is slightly larger around level $\eta = 0.80$.

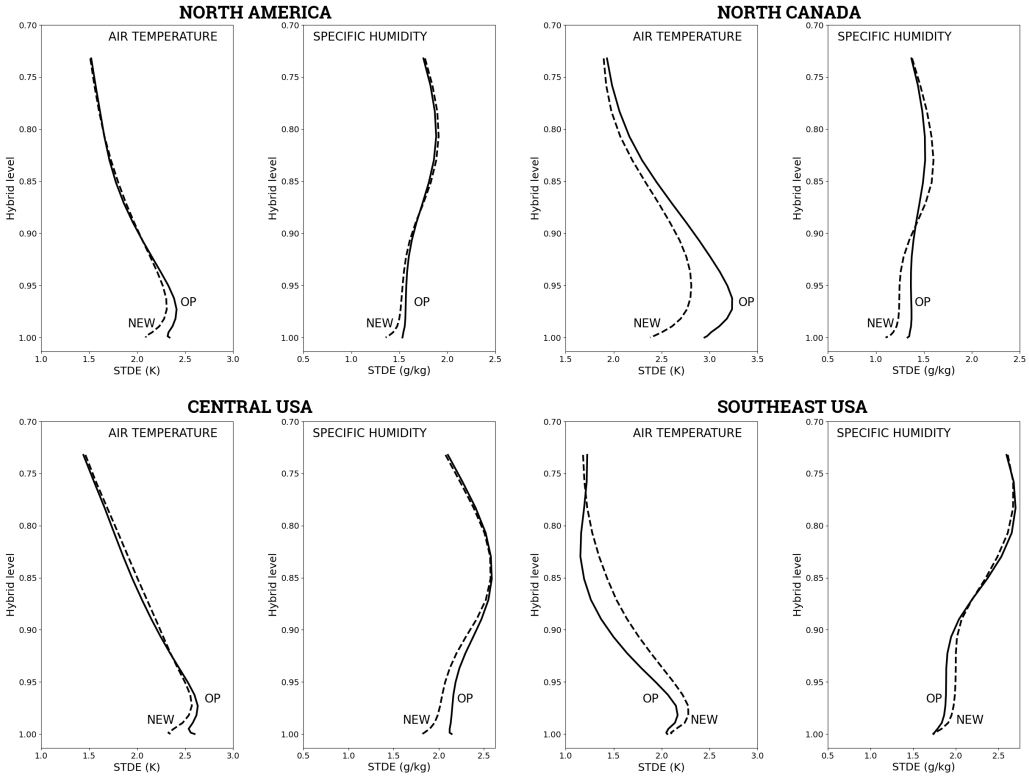

Figure 3. Vertical profiles of STDE errors for air temperature (K) and specific humidity (g kg^{-1}) spatially averaged over the land portion of North America and of the specific areas depicted by rectangles in Figure 2 for northern Canada, the central US, and the southeast US. Results are shown for the NEW (dashed lines) and OP (full lines) experiments. The profiles are temporally averaged for the forecasts valid at 0000 UTC, 0600 UTC, 1200 UTC, and 1800 UTC for ranges between 78 h and 144 h (i.e., forecasts for days 4 to 6). The model's hybrid η levels are used as vertical coordinates.

The other panels in Figure 3 reveal that the STDE vertical profiles and the relative performance of NEW versus OP are quite dependent on the location considered. Over northern Canada, the differences between OP and NEW are substantial for both air temperature and specific humidity. The air temperature STDE is smaller for NEW compared with that from OP for the entire atmospheric layer evaluated here (up to above level $\eta = 0.75$). The differences for the lowest atmospheric km are on the order of 0.5 K. For specific humidity, the STDE is also smaller for NEW near the surface but is significantly increased compared with OP in the layer between levels $\eta = 0.90$ and $\eta = 0.75$.

The opposite is found over the southeastern US (Figure 3's lower right panels), where there is a substantial deterioration in the air temperature STDE for NEW from the surface to level $\eta = 0.75$, with differences on the order of a few tenths of a degree. For specific humidity, the STDE is slightly worse for NEW in the lower km and slightly better for the layer between levels $\eta = 0.85$ and $\eta = 0.75$.

A different situation is observed for the domain over the central US, where STDE profiles are generally similar to what is previously described for North America. Over that region, the STDE is decreased with NEW for air temperature for the first few hundred meters and increased above level $\eta = 0.90$. The specific humidity STDE is decreased below

level $\eta = 0.85$ for NEW, with differences on the order of a few tenths of g kg^{-1} in the lowest few hundred meters. These STDE vertical profiles are consistent with the error difference maps presented in Figure 2.

The diurnal cycles of NEW's STDE vertical profiles and how they differ between OP and NEW are presented in Figure 4. The forecasts are valid at 0000 UTC, 0600 UTC, 1200 UTC, and 1800 UTC, and the errors are temporally averaged for day 4 to day 6 predictions. According to Figure 4's left panels, the air temperature STDE increases after 1200 UTC (morning) and becomes more vertical (i.e., more constant with height) from the surface to approximately hybrid level $\eta = 0.90$. It should be noted that the air temperature STDE for the southeast US remains larger near the surface; i.e., the error profile does not become vertical at 1800 UTC. For the three domains, the air temperature STDE further increases after 1800 UTC to reach its maximum value at 0000 UTC, i.e., in the evening. This increase late in the day is particularly pronounced for the southeastern US domain. The STDE difference profiles are consistent with the maps shown in Figure 2 and the vertical profiles in Figure 3, with very distinct characteristics for the three domains.

For northern Canada, the air temperature STDE differences (in favor of NEW for that domain) are amplified and deepen vertically throughout the day, as shown by the progression from 1200 UTC (morning) to 1800 UTC (afternoon) to 0000 UTC (evening). For the central US domain, the vertical profiles indicate a reduction in air temperature STDE in the lowest 500 m or so (below level $\eta = 0.95$) with a slight deterioration above. The error differences at low levels increase in vertical depth in the morning, from 1200 UTC to 1800 UTC, followed by a decrease late in the afternoon or early evening, from 1800 UTC to 0000 UTC. For the southeastern US, the air temperature STDE differences are in favor of OP and appear to decrease between 1200 UTC and 1800 UTC and then substantially increase between 1800 UTC and 0000 UTC, with a slight decrease after that (late evening, early night).

As shown in Figure 4's right panels, the diurnal evolution of the specific humidity STDE and of the impact of NEW versus OP is quite different from what is described just above for air temperature. For the three domains examined here, the specific humidity STDE becomes more vertically constant (or could be understood as vertically "well mixed") between 1200 UTC and 1800 UTC. This vertical mixing of the errors is associated with an increase in specific humidity STDE at the surface and in the lowest portion of the PBL. There is less diurnal variability in the atmospheric layer that could be interpreted as being above the PBL (above levels $\eta = 0.85$ or $\eta = 0.80$), except for the central US domain, where the specific humidity STDE substantially increases between 1800 UTC and 0000 UTC.

The diurnal evolution of specific humidity STDE differences is quite dissimilar for the three domains. For northern Canada, these errors are generally smaller for NEW near the surface (below $\eta = 0.90$) and greater for the layer above. The vertical profiles become more vertical (constant or well mixed) later in the day, as can be seen with the 0000 UTC profile. The opposite is found for the southeastern US domain, with a larger specific humidity STDE in the lower km and more neutral values above, except in the evening at 0000 UTC, when NEW's STDE is substantially smaller than OP's above level $\eta = 0.90$. Over the central US, the specific humidity STDE differences reveal NEW's better performance over that area for that variable, but a pattern for the diurnal evolution of STDE differences is difficult to discern, except for a substantial increase in this difference very close to the surface, in favor of NEW.

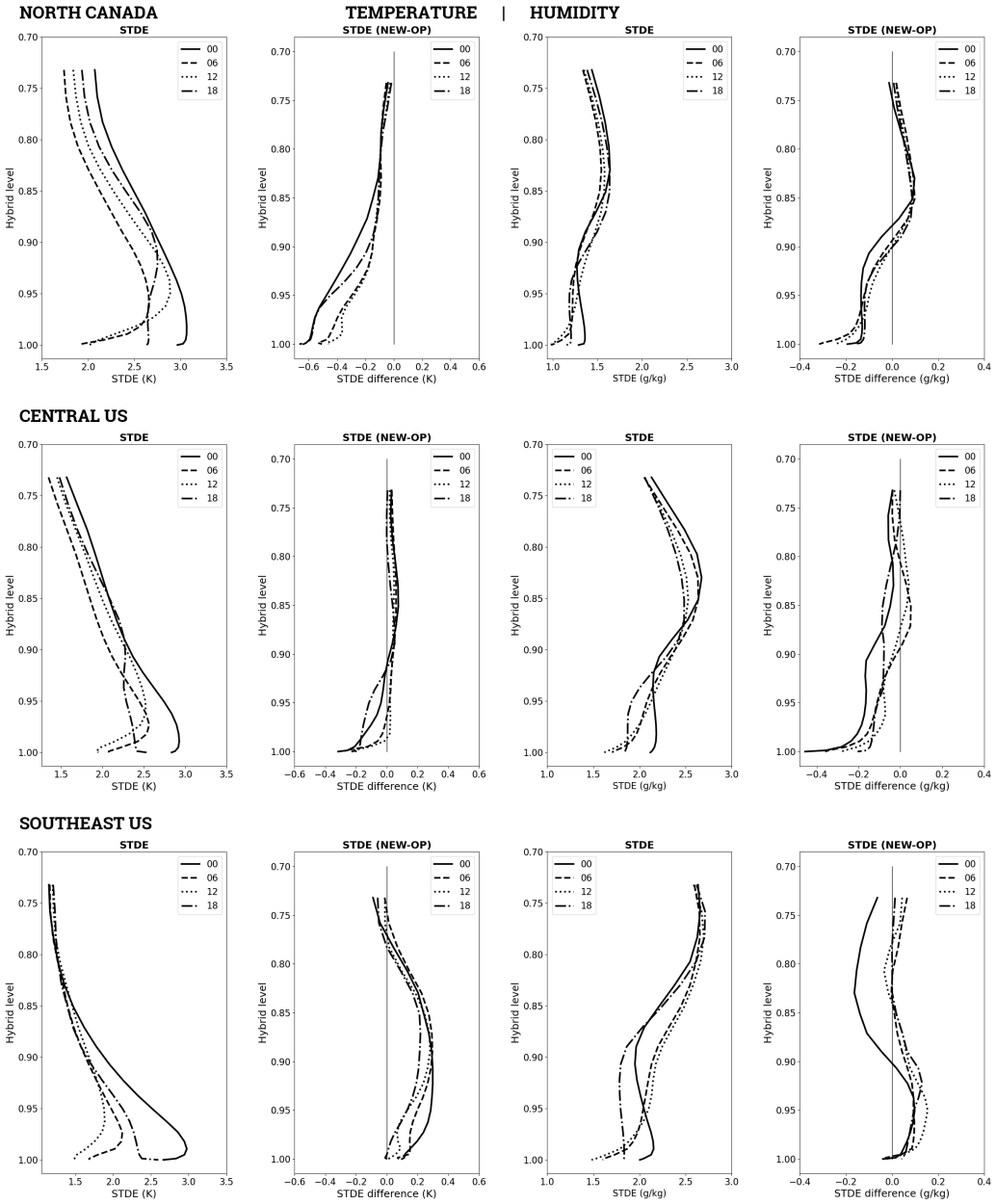

Figure 4. Vertical profiles of STDEs for air temperature (K) and specific humidity (g kg^{-1}) for the NEW experiment spatially averaged over northern Canada, the central US, and the southeast US for forecasts valid at 0000 UTC, 0600 UTC, 1200 UTC, and 1800 UTC, temporally averaged for day 4 to day 6 forecasts. Spatial averages of STDE differences (NEW minus OP) are also shown for the same areas and for the same forecast ranges.

One of the main objectives of this study is to determine how these very particular and complex vertical error structures for air temperature and specific humidity relate to errors at the land surface. The Pearson correlation coefficient is used here to determine the linear relationship between the atmospheric STDE differences between the two experiments and the error differences (STDE and bias) at the screen level. It provides information on how STDE is improved or deteriorated in the lower troposphere in relation to error changes at the surface. It can also help determine whether STDE deteriorations in the PBL (e.g., all the way up to hybrid levels $\eta = 0.85$ and $\eta = 0.80$) can be linked to similar deteriorations at the screen level, and whether biases play a role.

In Figure 5, these correlations are shown for STDE differences in air temperature (left panel) and specific humidity (right panel). The correlations are calculated over the North American domain and are shown for forecasts valid at 0000 UTC, averaged over the ranges 96 h, 120 h, and 144 h. For air temperature, the strongest link in the lower troposphere is with air temperature STDE differences at the screen level (left panel, full line), with a correlation coefficient of 1.0 at the screen level (by construction), decreasing almost linearly upward to a value of approximately 0.3 at level $\eta = 0.82$. There is also a good relationship with specific humidity STDE differences at the screen level, but to a much lesser degree. The correlation coefficient is about 0.4 at the screen level and decreases to a value close to 0.1 at level $\eta = 0.85$.

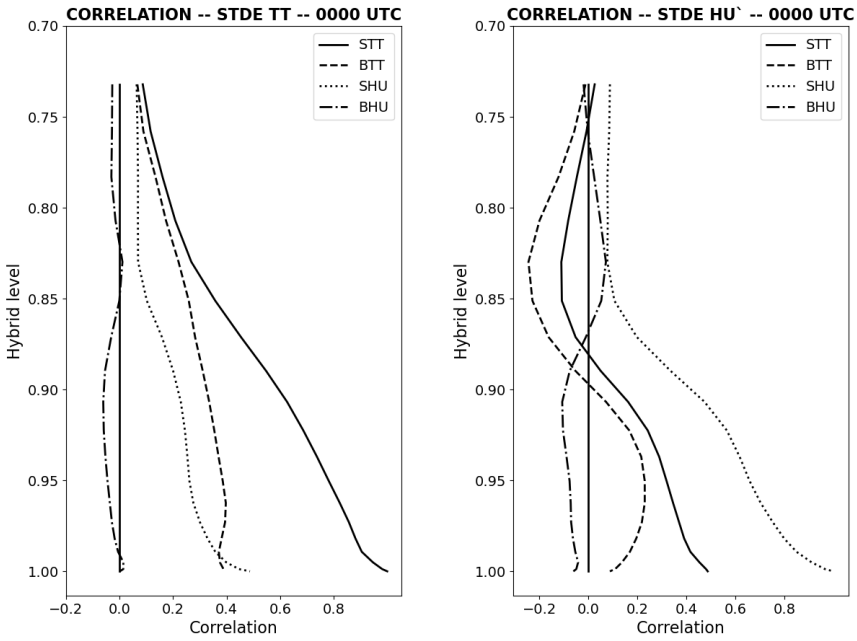

Figure 5. Vertical profiles of the Pearson correlation coefficient linking the STDE differences between NEW and OP for lower tropospheric air temperature and specific humidity with error differences in the same two variables at the screen level. The correlations are calculated over the North American domain and are shown for forecasts valid at 0000 UTC, averaged over the ranges 96 h, 120 h, and 144 h. In the legend, "STT" and "SHU" are for the correlations with STDE differences for air temperature and specific humidity at the screen level; "BTT" and "BHU" are for the correlations with bias differences for air temperature and specific humidity at the screen level.

The most notable feature in Figure 5's left panel is the fact that atmospheric STDE differences for air temperature correlate quite well with air temperature bias differences at the screen level. For levels $\eta = 0.85$ to $\eta = 0.80$, the correlation coefficient is almost as large as the one linking the same errors with air temperature STDE differences. It seems, however, that screen-level humidity biases do not play a substantial role, since the correlation coefficient for the "BHU" line in Figure 5's left panel remains small from the surface all the way up to levels above $\eta = 0.75$.

For specific humidity, the strongest link in the lowest portion of the atmosphere is with specific humidity STDE differences at the screen level (right panel, dotted line). This is the same as for air temperature, with a correlation coefficient of 1.0 at the screen level (by construction). This coefficient decreases rapidly to a value close to 0.1 at $\eta = 0.85$. The correlation coefficient with screen-level air temperature STDE differences is about 0.5 at the screen level but rapidly decreases with height to become negative (value of about -0.1 at level $\eta = 0.85$).

Again, the most noteworthy aspect is the link of specific humidity STDE differences with near-surface air temperature biases. In Figure 5's right panel, the dashed line for

"BTT" increases with height from the surface to reach a maximum value of about 0.2 close to level $\eta = 0.95$ and then decreases to reach strong negative values of less than -0.2 at $\eta = 0.85$. In other words, having cooler predictions at the surface has a positive impact on the humidity STDE for the lower part of the PBL, and the opposite is true for its upper portion. Considering that the link with specific humidity biases at the surface appears to be weak, it can be concluded that specific humidity STDE improvements or deteriorations are more strongly linked with air temperature bias changes than specific humidity bias changes at the screen level.

4. Discussion

The results presented in the previous section provide a statistical overview of the links between random errors for air temperature and humidity in the lower troposphere. In the discussion below, these statistics are further explored in order to determine the physical processes that could explain how atmospheric forecast errors are linked with model performance at the land surface. They help us better understand the differences in the errors diagnosed over the three areas specifically examined in this study, i.e., over northern Canada, the central US, and the southeast US. This analysis also informs on how errors at the surface and in the PBL relate to the diurnal evolution of the PBL, and on how they may influence the objective evaluation process traditionally used for NWP systems.

4.1. Statistically Explaining STDE in Upper Boundary Layer

The diagrams shown in Figure 6 provide a summary of the statistical links between screen-level errors and the STDE at level $\eta = 0.85$, which is approximately 1.5 km above the surface and could be considered to be in the upper portion of the boundary layer. The correlations displayed in the upper panel of that figure are obtained for the North American domain and are used here to determine whether errors at $\eta = 0.85$ for the three domains are consistent with the behavior previously described. The results in this section are only discussed for 0000 UTC, in the late afternoon and evening for North America, when the STDE is largest in the lower troposphere (see Figure 4).

Based on North American error statistics for this time of day, it seems that improving surface predictions (in terms of STDE) indeed has a positive impact on the STDE for both air temperature and humidity for the entire PBL, as revealed in the vertical profiles shown in Figure 5 and as summarized in the upper panel of Figure 6. But, the statistical links with air temperature biases at the screen level are also large in the upper part of the PBL and appear to be a crucial factor for both air temperature and specific humidity (see the BTT portion of the histograms in Figure 6's upper panel). In comparison, specific humidity errors at the screen level appear to have less impact on errors in the PBL.

For northern Canada (second panel in Figure 6), the air temperature STDE decrease at level $\eta = 0.85$ with NEW (decreased by 0.15 K) could be related to the substantial decrease in the air temperature STDE at the screen level (decreased by 0.64 K) and smaller values for near-surface air temperature with NEW (decreased by 1.16 K), both aspects positively correlated with the boundary-layer air temperature STDE reduction (see top panel). Similarly, the specific humidity STDE increase with NEW at level $\eta = 0.85$ (increased by 0.09 g kg^{-1}) can be mostly linked with the same cooling at the screen level (negatively correlated) and, to a lesser degree, with the air temperature STDE decrease at the screen level (also negatively correlated). The improvement for screen-level specific humidity STDE with NEW is, however, not consistent with the positive correlations displayed for both air temperature and humidity at $\eta = 0.85$ in the upper panel of Figure 6 since the humidity STDE is actually increased over that area at $\eta = 0.85$ (see left portion of Figure 6's second panel).

Figure 6's bottom panel describes the situation for the southeast US subdomain, another area where the STDE differences between the two experiments are large at level $\eta = 0.85$. Over that area, the large increase in air temperature STDE at $\eta = 0.85$ for NEW (about 0.24 K) could be related to increases in the air temperature STDE (0.10 K) and, mostly,

to a larger bias (1.13 K warmer with NEW) at the screen level. On the other hand, the same warming at the surface with NEW could be used to explain the substantial decrease in the specific humidity STDE at level $\eta = 0.85$ for NEW (0.14 g kg^{-1}).

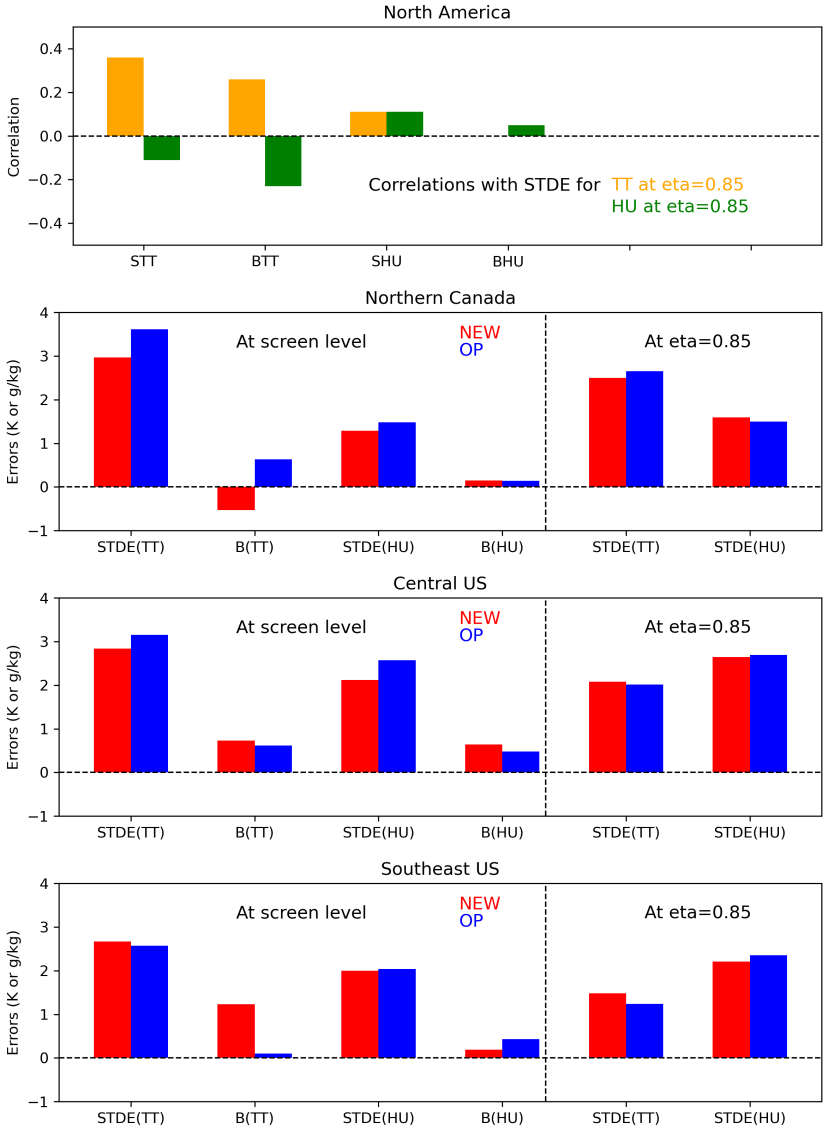

Figure 6. The upper panel provides histograms showing the statistical relationship between screen-level errors (STDE and bias) and the STDE at level $\eta = 0.85$ for the North American domain (in yellow for air temperature and in green for specific humidity). The lower three panels show the STDEs and biases for air temperature and specific humidity at the screen level and the STDEs at $\eta = 0.85$ for the three subdomains indicated by the black rectangles in Figure 2. The results for NEW are in red, whereas the results for OP are in blue. In the legends, "B" refers to bias, while "TT" and "HU" refer to air temperature and specific humidity.

For the central US subdomain (third panel in Figure 6), the STDE differences at $\eta = 0.85$ between the two experiments are much smaller compared with northern Canada and the southeast US. The slight deterioration with NEW for the air temperature STDE at $\eta = 0.85$ (0.07 K) appears to be linked with warmer air at the screen level (increased by 0.11 K), in spite of the contradictory signal from the air temperature and specific humidity STDE differences at the screen level, which are both reduced with NEW. The small improvement with NEW for the specific humidity STDE at $\eta = 0.85$ could also be

related to the warmer surface and to a substantial decrease in the specific humidity STDE (decreased by 0.45 g kg^{-1}).

4.2. Physical Links with PBL Diurnal Evolution

The strong correlation between STDE differences in the lower troposphere and STDE differences in the same variable at the screen level is to be expected. The results indeed show that the air temperature STDE in the lowest few hundred meters is well correlated with the air temperature STDE at the surface. The same can be said for specific humidity. More interesting is the relationship between air temperature biases at the surface and boundary-layer STDEs for air temperature and specific humidity in the first few kilometers. The role and impact of near-surface temperature biases are investigated in Figure 7, which provides some physical interpretation by showing the spatially and temporally averaged diurnal evolution of the boundary-layer height over the three subdomains examined in this study. In this figure, the PBL height is temporally averaged for days 4, 5, and 6 of all summer 2019 forecasts.

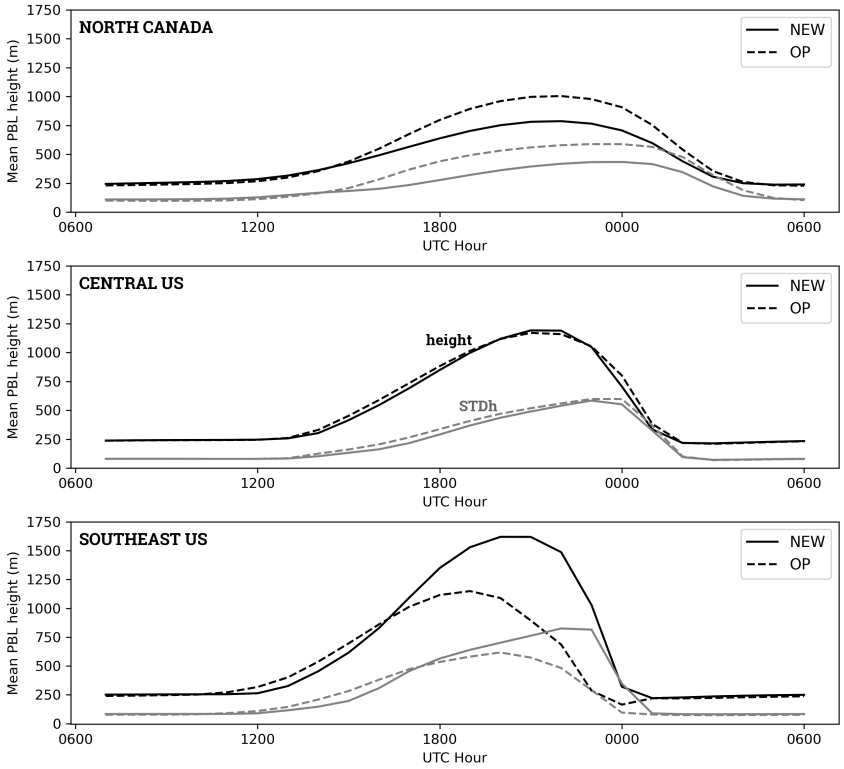

Figure 7. The diurnal evolution of the PBL height (m), spatially averaged over the subdomains shown in Figure 2 and temporally averaged for days 4, 5, and 6 of all summer 2019 forecasts. The spatial mean (in black) and standard deviation (in gray) are shown for the NEW and OP experiments (full and dashed lines, respectively).

Over northern Canada (Figure 7's upper panel), the maximum PBL height is on the order of 750 m for NEW and 1000 m for OP, with a spatial variability that is also substantially larger for OP, i.e., close to 500 m for OP compared with 350–400 m for NEW. For both experiments, the PBL is most developed for a few hours before 0000 UTC, while the spatial variability is largest just after 0000 UTC. Going directly to Figure 7's lower panel, the situation is reversed for the southeast US, where the PBL is substantially deeper for NEW, with peak values greater than 1500 m around 2100 UTC, compared with a maximum PBL height of about 1000 m for OP, peaking a few hours earlier than NEW (i.e., just after 1800 UTC). Differences are also found for the spatial variability, which is close to 750 m for NEW just before 0000 UTC, compared with a standard deviation of about 500 m, peaking around 2000 UTC for OP. In contrast, the PBL appears to evolve in a similar manner for the

central US domain (Figure 7's middle panel), with a maximum PBL depth slightly greater than 1000 m around 2100 UTC and a maximum standard deviation of about 500 m near 0000 UTC.

From this figure, the results indicate that the PBL height and spatial variability are directly related to the screen-level air temperature bias differences. Deeper, more developed boundary layers are related to greater turbulent activity and variance (both spatial and temporal) in the lowest few kilometers. This increased turbulent activity is represented in the atmospheric model by subgrid-scale vertical diffusion and by grid-scale upward motion. In either case, the increased entrainment of warmer and drier air at the top of the PBL associated with larger air temperature biases (i.e., warmer air) at the screen level could partly explain how the STDE in the lower troposphere is physically related to errors at the land surface.

4.3. PBL Errors and Upper-Air NWP Evaluation

The results of this study suggest that land areas where air temperature is decreased at the screen level (i.e., colder forecasts) should have smaller air temperature STDEs in the lower troposphere, at least for medium-range summertime forecasts valid at 0000 UTC. This could have an impact on the objective evaluation traditionally performed at prediction centers to determine the performance of their NWP systems. For instance, the STDE or RMSE for air temperature forecasts at 850 hPa is often used as one of the main target variables to improve. It is often displayed as part of scorecards and is included in the calculation of aggregated NWP performance indices [46,47]. Based on the results obtained in the present study, this practice might be problematic since random errors of air temperature at 850 hPa (similar to level $\eta = 0.85$ here) are subject to variability in atmospheric turbulence (parameterized or grid scale) and, in addition, could be sensitive to changes in biases for screen-level air temperature.

In fact, this dependence on the boundary-layer turbulence and near-surface temperature bias could have a more profound impact. The geopotential height, arguably the most popular variable used to evaluate the performance of global NWP forecast systems, is also sensitive to perturbations in air temperature STDE errors associated with near-surface biases.

In Figure 8, the performance differences for atmospheric temperature STDEs between NEW and OP (NEW minus OP) at 0000 UTC (i.e., 96 h, 120 h, and 144 h forecast ranges) over a series of land surface points selected on narrow percentile bands on the cold side (between 20th and 30th percentiles) and warm side (between 70th and 80th percentiles) of the spatial distribution of temperature bias differences at the screen level are compared. The distributions include land grid points over North America for forecasts valid at 0000 UTC for the ranges 96 h, 120 h, and 144 h. In this figure, the dashed line shows the vertical profile of STDE differences (NEW minus OP) for points where the bias difference at the screen level between NEW and OP is smaller than the mean (i.e., could be considered areas where NEW is cooler). And, the dash-dotted line displays the same but for land points where NEW could be considered warmer.

In both cases, NEW's temperature STDE is smaller than OP's for the North American domain. The STDE differences are larger at the surface and decrease to near zero near level $\eta = 0.87$. Above that level, the differences become negative again for the cold subset of points, while they remain near zero for the warm subset. For the first few kilometers above the surface, the STDE differences (i.e., gains for NEW versus OP) are more negative (more in favor of NEW) for the cold points. The performance difference between the cold and warm series of land points reaches about 0.05 K at η levels 0.93 and 0.78.

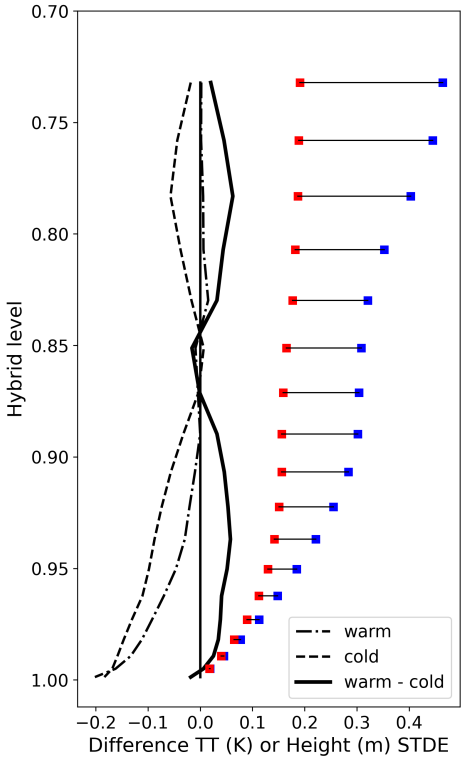

Figure 8. Vertical profiles of air temperature STDE differences between NEW and OP (NEW minus OP) for land surface points selected between the 20th and 30th percentiles ("cold" in the figure) and the 70th and 80th percentiles ("warm" in the figure) of the spatial distribution of air temperature bias differences between NEW and OP at the screen level. These distributions are over the land portion of North America for forecasts valid at the 96 h, 120 h, and 144 h ranges. The blue and red squares represent the differences in height differential errors (m) based on Equation (9) for the "cold" (blue) and "warm" (red) vertical profiles.

A rough estimate of the quantitative impact that this distinct performance over cold and warm areas could have on the geopotential height STDE is provided based on a simple development from the atmospheric hydrostatic equilibrium equation and the ideal gas law:

$$\frac{\partial p}{\partial z} = -\rho g \qquad ; \qquad p = \rho RT \tag{7}$$

in which ρ is the air density, g is the acceleration due to gravity, and R is the universal gas constant. Rewriting this in a discretized form and using the fact that η levels can be approximated by the ratio of pressure to surface pressure near the surface, the following can be obtained:

$$\partial z = -\frac{RT}{g}\frac{\partial p}{p} \approx \frac{RT}{g}\frac{\partial \eta}{\eta} \tag{8}$$

which can then be used to estimate the height difference Δz_k, or uncertainty, associated with the air temperature STDE difference ΔT:

$$\Delta z_k \approx \frac{R}{g}\sum_{surface}^{k}\left(\frac{\Delta \eta}{\eta}\right)\Delta T \tag{9}$$

The horizontal distance between the colored squares in Figure 8 indicates that this near-surface bias could have an impact on the order of 0.3 m (for height errors, i.e., 0.03 dam for geopotential height errors) when integrated from the surface to level $\eta = 0.73$. This difference is not negligible and is sufficiently large to have an influence on decisions based on the geopotential height STDE or RMSE.

5. Summary and Conclusions

In this study, the impact of a new land surface package with modifications to surface fields, land surface modeling, and land data assimilation is presented in terms of random errors (STDE) in the lower troposphere in medium-range summertime forecasts. Using correlation coefficients, the statistical relationships between STDEs at the screen level and in the lower atmosphere are investigated. Diagnostics for the mean PBL diurnal evolution provide insight into the physical processes that could explain some of the links between surface and atmospheric errors.

In agreement with previous studies, it is clear that improvements at the surface or screen level do not always or automatically translate into improvements in the lower troposphere or in the PBL. Although the NEW configuration evaluated in this study generally improves medium-range forecasts over North America for both screen-level and lower tropospheric results, the comparison with the control OP configuration depends on the geographical area. Three subdomains are used to illustrate situations in which the impact of NEW versus OP is quite different. Diagnostics over these three subdomains reveal a very distinct diurnal evolution of STDE errors, with maximum errors at 0000 UTC, i.e., in the evening, at the end, or just after the daytime boundary-layer growth.

The STDEs for air temperature and specific humidity in the lower troposphere correlate well with STDEs at the screen level, with coefficients smoothly decreasing with height. Interestingly, the air temperature bias at the screen level is also found to have a significant influence, often as important (or even more important) as the near-surface STDE for the same variable. These statistical relationships derived at the continental scale appear to properly explain the differences in the STDE profiles for the three subdomains examined in this study.

The evolution of the PBL is also closely linked with near-surface temperature biases and is likely to be one of the main physical causes explaining the impact of near-surface errors on the STDE in the lower troposphere. It appears that a lower air temperature at the surface is beneficial for the air temperature STDE in the PBL; i.e., areas that are colder seem to have lower temperature STDEs in the upper part of the boundary layer. This leads to a note of caution to NWP modelers concerning the inclusion of 850 hPa (or 925 hPa) air temperature in the scorecards now systematically used for NWP model evaluation. This extends to the inclusion of 850 hPa air temperature in the production of NWP indices. Caution is also warranted when looking at random errors for geopotential height because changes in near-surface temperature bias and its impact on the diurnal PBL evolution can affect or perturb these errors.

There are, of course, several limitations to this study. Some of them are listed below, along with possible extensions for this work:

- The same analysis could be carried out during the winter season and over other areas, when and where the nature of the links between errors at the surface and in the atmosphere could be different.
- It is not clear whether the same kind of statistical and physical links would be found in other atmospheric models.
- Do we see the same kind of behavior for weather prediction based on AI forecasting systems?
- Diagnostics for land–atmosphere coupling could be examined in order to have a more complete understanding of the links between errors at the surface and in the boundary layer.

Work is ongoing at ECCC to evaluate the impact of a similar land surface package in its km-scale short-range deterministic prediction system.

Author Contributions: Conceptualization, S.B.; methodology, S.B. and M.L.C.; software, S.B., S.L., N.A., M.L.C., M.A., B.B. (Bernard Bilodeau), D.S., D.C. and B.B. (Bakr Badawy); validation, N.A., D.S. and D.C.; formal analysis, S.B. and N.A.; investigation, S.B. and N.A.; resources, S.B. and M.L.C.; data curation, D.S. and D.C.; writing—original draft preparation, S.B.; writing—review and editing, N.A.,

S.L., M.L.C., M.A., B.B. (Bernard Bilodeau), D.S., D.C. and B.B. (Bakr Badawy); visualization, S.B. and N.A.; supervision, S.B., M.L.C. and M.A.; project administration, S.B.; funding acquisition, S.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The NWP model forecasts and verification analyses produced for this study will be kept internally at ECCC for a period of 5 years after publication of this article. A subset could be made available for reasonable requests, pending the availability of human resources at ECCC to convert these analysis files into a format readable by external users.

Acknowledgments: The authors would like to thank the scientific management at ECCC's ASTD and CCMEP branches for their support of this project.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AIRS	Atmospheric Infrared Sounder
CRIS	CRoss-track Infrared Sounder
EnKF	Ensemble Kalman Filter data assimilation system
EnVar	Ensemble Variational data assimilation system
CaLDAS	Canadian Land Data Assimilation System
CaPA	Canadian Precipitation Analysis
CCMEP	Canadian Centre for Meteorological and Environmental Prediction
CTEM	Canadian Terrestrial Ecosystem Model
ECCC	Environment and Climate Change Canada
GDPS	Global Deterministic Prediction System
GEM	Global Environmental Multiscale model
GEWEX	Global Energy and Water EXchanges initiative
HRRR	High-Resolution Rapid Refresh
IASI	Infrared Atmospheric Sounding Interferometer
ISBA	Interactions between Soil, Biosphere, and Atmosphere scheme
LoCo	Local Land–Atmosphere Coupling
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical weather prediction
PBL	Planetary boundary layer
SMAP	Soil Moisture Active Passive mission
SMOS	Soil Moisture and Ocean Salinity mission
SVS	Soil, Vegetation, and Snow scheme
TKE	Turbulent kinetic energy
UTC	Universal Time Coordinated

References

1. Kala, J.; Evans, J.P.; Pitman, A.J. Influence of antecedent soil moisture conditions on the synoptic meteorology of the Black Saturday bushfire event in southeast Australia. *Q. J. R. Meteorol. Soc.* **2015**, *141*, 3118–3129.
2. Zheng, W.; Zhan, X.; Liu, J.; Ek, M. A preliminary assessment of the impact of assimilating satellite soil moisture data products on NCEP Global Forecast System. *Adv. Meteorol.* **2018**, *2018*, 7363194 .
3. Crow, W.T.; Gomez, C.A.; Sabater, J.M.; Holmes, T.; Hain, C.R.; Lei, F.; Dong, J.; Alfieri, J.G.; Anderson, M.C. Soil moisture–evapotranspiration overcoupling and L-Band brightness temperature assimilation: Sources and forecast implications. *J. Hydrometeorol.* **2020**, *21*, 2359–2374.
4. Champeaux, J.L.; Arcos, D.; Bazile, E.; Giard, D.; Goutorbe, J.P.; Habets, F.; Noilhan, J.; Roujean, J.L. AVHRR-derived vegetation mapping over Western Europe for use in numerical weather prediction models. *Int. J. Remote Sens.* **2000**, *21*, 1183–1199.
5. Mailhot, J.; Bélair, S.; Charron, M.; Doyle, C.; Joe, P.; Abrahamowicz, M.; Bernier, N.; Denis, B.; Erfani, A.; Frenette, R.; et al. Environment Canada's experimental numerical weather prediction systems for the Vancouver 2010 Winter Olympic and Paralympic Games. *Bull. Am. Meteorol. Soc.* **2010**, *91*, 1073–1086.
6. Campbell, P.C.; Bash, J.O.; Spero, T.L. Updates to the Noah land surface model in WRF-CMAQ to improve simulated meteorology, air quality, and deposition. *J. Adv. Model. Earth Syst.* **2019**, *11*, 231–256.

7. Husain, S.Z.; Alavi, N.; Bélair, S.; Carrera, M.; Zhang, S.; Fortin, V.; Abrahamowicz, M.; Gauthier, N. The multibudget Soil, Vegetation, and Snow (SVS) scheme for land surface parameterization: Offline warm season evaluation. *J. Hydrometeorol.* **2016**, *17*, 2293–2313.
8. Bakketun, Å.; Blyverket, J.; Müller, M. Using a reanalysis-driven land surface model for initialization of a numerical weather prediction System. *Weather Forecast.* **2023**, *38*, 2155–2168.
9. Benjamin, S.G.; Smirnova, T.G.; James, E.P.; Lin, L.F.; Hu, M.; Turner, D.D.; He, S. Land–snow data assimilation including a moderately coupled initialization method applied to NWP. *J. Hydrometeorol.* **2022**, *23*, 825–845.
10. Sassi, M.Z.; Fourrié, N.; Guidard, V.; Birman, C. Preliminary assimilation of satellite derived land surface temperature from SEVIRI in the surface scheme of the AROME-France model. *Tellus A* **2023**, *75*, 88–107.
11. Bélair, S.; Brown, R.; Mailhot, J.; Bilodeau, B.; Crevier, L.P. Operational implementation of the ISBA land surface scheme in the Canadian regional weather forecast model. Part II: Cold season results. *J. Hydrometeorol.* **2003**, *4*, 371–386.
12. Bélair, S.; Crevier, L.P.; Mailhot, J.; Bilodeau, B.; Delage, Y. Operational implementation of the ISBA land surface scheme in the Canadian regional weather forecast model. Part I: Warm season results. *J. Hydrometeorol.* **2003**, *4*, 352–370.
13. Koster, R.D.; Sud, Y.; Guo, Z.; Dirmeyer, P.A.; Bonan, G.; Oleson, K.W.; Chan, E.; Verseghy, D.; Cox, P.; Davies, H.; et al. GLACE: The Global Land–Atmosphere Coupling Experiment. Part I: Overview. *J. Hydrometeorol.* **2006**, *7*, 590–610.
14. Guo, Z.; Dirmeyer, P.A.; Koster, R.D.; Sud, Y.; Bonan, G.; Oleson, K.W.; Chan, E.; Verseghy, D.; Cox, P.; Gordon, C.; et al. GLACE: The Global Land–Atmosphere Coupling Experiment. Part II: Analysis. *J. Hydrometeorol.* **2006**, *7*, 611–625.
15. Santanello, J.A.; Dirmeyer, P.A.; Ferguson, C.R.; Findell, K.L.; Tawfik, A.B.; Berg, A.; Ek, M.; Gentine, P.; Guillod, B.P.; Van Heerwaarden, C.; et al. Land–atmosphere interactions: The LoCo perspective. *Bull. Am. Meteorol. Soc.* **2018**, *99*, 1253–1272.
16. Santanello, J.A., Jr.; Lawston, P.; Kumar, S.; Dennis, E. Understanding the impacts of soil moisture initial conditions on NWP in the context of land–atmosphere coupling. *J. Hydrometeorol.* **2019**, *20*, 793–819.
17. Lawston-Parker, P.; Santanello, J.A., Jr.; Kumar, S.V. Understanding the impacts of land surface and PBL observations on the terrestrial and atmospheric legs of land–atmosphere coupling. *J. Hydrometeorol.* **2021**, *22*, 2241–2258.
18. Dirmeyer, P.A.; Halder, S. Sensitivity of numerical weather forecasts to initial soil moisture variations in CFSv2. *Weather Forecast.* **2016**, *31*, 1973–1983.
19. Bélair, S.; Alavi, N.; Carrera, M.L.; Abrahamowicz, M.; Bilodeau, B.; Simjanovski, D.; Charpentier, D.; Badawy, B.; Leroyer, S. Evaluating the impact of land surface on medium-range weather forecasts using screen-level analyses. *Q. J. R. Meteorol. Soc.* **2023**, *149*, 2146–2175.
20. Côté, J.; Gravel, S.; Méthot, A.; Patoine, A.; Roch, M.; Staniforth, A. The operational CMC–MRB Global Environmental Multiscale (GEM) model. Part I: Design considerations and formulation. *Mon. Weather Rev.* **1998**, *126*, 1373–1395.
21. Côté, J.; Desmarais, J.G.; Gravel, S.; Méthot, A.; Patoine, A.; Roch, M.; Staniforth, A. The operational CMC–MRB Global Environmental Multiscale (GEM) model. Part II: Results. *Mon. Weather Rev.* **1998**, *126*, 1397–1418.
22. Girard, C.; Plante, A.; Desgagné, M.; McTaggart-Cowan, R.; Côté, J.; Charron, M.; Gravel, S.; Lee, V.; Patoine, A.; Qaddouri, A.; et al. Staggered vertical discretization of the Canadian Environmental Multiscale (GEM) model using a coordinate of the log-hydrostatic-pressure type. *Mon. Weather Rev.* **2014**, *142*, 1183–1196.
23. Husain, S.Z.; Girard, C. Impact of consistent semi-Lagrangian trajectory calculations on numerical weather prediction performance. *Mon. Weather Rev.* **2017**, *145*, 4127–4150.
24. McTaggart-Cowan, R.; Vaillancourt, P.; Zadra, A.; Chamberland, S.; Charron, M.; Corvec, S.; Milbrandt, J.; Paquin-Ricard, D.; Patoine, A.; Roch, M.; et al. Modernization of atmospheric physics parameterization in Canadian NWP. *J. Adv. Model. Earth Syst.* **2019**, *11*, 3593–3635.
25. Buehner, M.; McTaggart-Cowan, R.; Beaulne, A.; Charette, C.; Garand, L.; Heilliette, S.; Lapalme, E.; Laroche, S.; Macpherson, S.R.; Morneau, J.; et al. Implementation of deterministic weather forecasting systems based on ensemble–variational data assimilation at Environment Canada. Part I: The global system. *Mon. Weather Rev.* **2015**, *143*, 2532–2559.
26. Buehner, M. Local ensemble transform Kalman filter with cross validation. *Mon. Weather Rev.* **2020**, *148*, 2265–2282.
27. Caron, J.F.; Buehner, M. Implementation of scale-dependent background-error covariance localization in the Canadian Global Deterministic Prediction System. *Weather Forecast.* **2022**, *37*, 1567–1580.
28. Noilhan, J.; Planton, S. A simple parameterization of land surface processes for meteorological models. *Mon. Weather Rev.* **1989**, *117*, 536–549.
29. Alavi, N.; Bélair, S.; Fortin, V.; Zhang, S.; Husain, S.Z.; Carrera, M.L.; Abrahamowicz, M. Warm season evaluation of soil moisture prediction in the Soil, Vegetation, and Snow (SVS) scheme. *J. Hydrometeorol.* **2016**, *17*, 2315–2332.
30. Arora, V.K. Simulating energy and carbon fluxes over winter wheat using coupled land surface and terrestrial ecosystem models. *Agric. For. Meteorol.* **2003**, *118*, 21–47.
31. Arora, V.K.; Boer, G.J. A parameterization of leaf phenology for the terrestrial ecosystem component of climate models. *Glob. Chang. Biol.* **2005**, *11*, 39–59.
32. Mahfouf, J.F. Analysis of soil moisture from near-surface parameters: A feasibility study. *J. Appl. Meteorol. Climatol.* **1991**, *30*, 1534–1547.
33. Carrera, M.L.; Bélair, S.; Bilodeau, B. The Canadian Land Data Assimilation System (CaLDAS): Description and synthetic evaluation study. *J. Hydrometeorol.* **2015**, *16*, 1293–1314.

34. Carrera, M.L.; Bilodeau, B.; Bélair, S.; Abrahamowicz, M.; Russell, A.; Wang, X. Assimilation of passive L-band microwave brightness temperatures in the Canadian Land Data Assimilation System: Impacts on short-range warm season numerical weather prediction. *J. Hydrometeorol.* **2019**, *20*, 1053–1079.
35. Fortin, V.; Roy, G.; Donaldson, N.; Mahidjiba, A. Assimilation of radar quantitative precipitation estimations in the Canadian Precipitation Analysis (CaPA). *J. Hydrol.* **2015**, *531*, 296–307.
36. Fortin, V.; Roy, G.; Stadnyk, T.; Koenig, K.; Gasset, N.; Mahidjiba, A. Ten years of science based on the Canadian Precipitation Analysis: A CaPA system overview and literature review. *Atmosphere-Ocean* **2018**, *56*, 178–196.
37. Heilliette, S.; Garand, L.; Bilodeau, B.; Carrera, M.; Belair, S. Assimilation of land surface skin temperature observations derived from GOES imagery. In Proceedings of the 21st TOVS Study Conference, Darmstadt, Germany, 29 November–5 December 2017.
38. Benoit, R.; Côté, J.; Mailhot, J. Inclusion of a TKE boundary layer parameterization in the Canadian regional finite-element model. *Mon. Weather Rev.* **1989**, *117*, 1726–1750.
39. Beljaars, A.; Holtslag, A. Flux parameterization over land surfaces for atmospheric models. *J. Appl. Meteorol. Climatol.* **1991**, *30*, 327–341.
40. Delage, Y.; Girard, C. Stability functions correct at the free convection limit and consistent for for both the surface and Ekman layers. *Bound.-Layer Meteorol.* **1992**, *58*, 19–31.
41. Delage, Y. Parameterising sub-grid scale vertical transport in atmospheric models under statically stable conditions. *Bound.-Layer Meteorol.* **1997**, *82*, 23–48.
42. Bélair, S.; Mailhot, J.; Strapp, J.W.; MacPherson, J.I. An examination of local versus nonlocal aspects of a TKE-based boundary layer scheme in clear convective conditions. *J. Appl. Meteorol. Climatol.* **1999**, *38*, 1499–1518.
43. McTaggart-Cowan, R.; Zadra, A. Representing Richardson number hysteresis in the NWP boundary layer. *Mon. Weather Rev.* **2015**, *143*, 1232–1258.
44. Beljaars, A.C.; Brown, A.R.; Wood, N. A new parametrization of turbulent orographic form drag. *Q. J. R. Meteorol. Soc.* **2004**, *130*, 1327–1347.
45. Carminati, F.; Migliorini, S.; Ingleby, B.; Bell, W.; Lawrence, H.; Newman, S.; Hocking, J.; Smith, A. Using reference radiosondes to characterise NWP model uncertainty for improved satellite calibration and validation. *Atmos. Meas. Tech.* **2019**, *12*, 83–106.
46. Ling, F.; Ouyang, L.; Larbi, B.R.; Luo, J.J.; Han, T.; Zhong, X.; Bai, L. Is artificial intelligence providing the second revolution for weather forecasting? *arXiv* **2024**, arXiv:2401.16669.
47. Yokota, S.; Banno, T.; Oigawa, M.; Akimoto, G.; Kawano, K.; Ikuta, Y. JMA operational hourly hybrid 3DVar with singular vector-based mesoscale ensemble prediction system. *J. Meteorol. Soc. Jpn. Ser. II* **2024**, *102*, 129–150.