

Answer to Referee 3

We would like to thank the referee for his constructive criticism which helped to improve the manuscript.

1. My main scientific concern deals with assumptions on the nature of the BL. In late winter and early spring, especially under clear skies, the High Arctic BL is often a strong surface-based temperature inversion with very high static stability. As a result the wind and temperature profiles show large gradients very close to the ground which are not captured by the reanalysis data sets. This calls into question not only the wind speed and wind directions used in the back trajectories but also the validity of the assumed boundary layer depths which become the mixing heights as I understand their approach. I think it is important to anchor the reanalysis data with the station radiosonde measurements of the boundary layer profile. Was this done? If not then I suggest that this be used to reduce the errors introduced from the reanalysis data. Under strong stability conditions there is a decoupling between the surface air and the air just above it. It might be better to compare your model predicted temperatures with the temperature above 2-m at the station if the boundary layer is very stable. This issue needs to be discussed. The results should be stratified according to the surface-based inversion strength to test for a dependence on stability.

We agree that our model assumptions might not be suitable for cases with strong surface-based inversions. Therefore we used the station radiosonde data to determine the frequency of strong inversions. We could only use soundings at Barrow since too few soundings were available at Tara. We also neglected the soundings at Alert which were conducted further away from the coast and at higher elevation than the measurements of the 2-m air temperature and are therefore not fully comparable. We calculated the stratification in terms of the Richardson number in the lowest 30 m for different wind speed classes (see new Fig. 1). We found that for wind speeds between 2 and 4 m/s more than 90 % of the Ri numbers were below the critical value of 0.25 and 75 % below 0.1. Therefore, we limited our study to those cases with wind speeds mainly above 3 m/s to exclude cases with strong inversions. Furthermore, we found that our model is not necessarily limited to well mixed BL (For details see 2.). A description is included in Sect. 3.3 (par. 3):

“To verify the assumption of a well mixed BL radiosonde data are analyzed. Since soundings at Tara during the considered period are sparse and the soundings at Alert are conducted at higher elevations than the temperature measurements only data from Barrow are considered. Soundings from the University of Wyoming dataset are used which are available every 12 h. Only those soundings with wind direction from the ice are considered. The stratification is expressed in terms of the Richardson number (Ri) in the lowest 30m as a function of wind speed (Fig. 1). For wind speeds between 2 and 4 ms⁻¹ about 90 % of the Ri numbers are below the critical value of 0.25 and 75 % below 0.1. Therefore, the assumption of a well mixed BL seems to be valid for wind speeds above 3 ms⁻¹. In addition, the few sounding from Tara all show a well mixed BL. Therefore, only trajectories with 90 % of the wind speeds above 3 ms⁻¹ are considered. This limit is lowered to 80 % of the wind speed above 2 ms⁻¹ for Alert since too few cases remain if the stricter criterion is applied.”

By limiting our study to cases with stronger winds we exclude most surface inversion cases and therefore do not expect a decoupling in the lower boundary layer.

We are still facing the uncertainties in the 10-m wind fields of the reanalyses. In principle, we could anchor the reanalysis data at the stations using radiosonde data. These are, however, only point measurements which cannot be assumed to be valid along the whole trajectory. In addition, only soundings from Barrow with a time resolution of 12 h could be used (see above) which is not sufficient for hourly trajectories for all three stations.

2. This also relates to the BL assumptions. On page 3018, line 18, it states that “the BL is assumed to be well mixed with a constant potential temperature above a reference height of 10 m”. While this may be true when the surface winds are strong, I doubt it is true for cases of light winds and so this assumption needs to be checked using the radiosonde measurements at the 3 stations used.

The integration of our model equation is not only valid for a well mixed BL. It can be shown, that the same equation is valid when we use a linear temperature profile or a power law with a time constant vertical temperature gradient. We include this in Sect. 3.3 (par. 2):

“In general, the solution of Eq. 1 depends on the specified temperature profile. However, it can be shown (Appendix A) that the solutions are identical for a well mixed BL with height constant θ and for a more general power law temperature profile. This holds for the assumption that both the difference between the temperatures at 10 m height and at the BL top and the mixed layer height H are not depending on time. In the following, however, a well mixed BL is assumed since in this case the latter condition ($H = \text{const.}$) is not necessary. Furthermore, a constant flux layer is assumed below the reference height of $z_{\text{ref}} = 10$ m, with logarithmic profiles of wind and potential temperature.”

However, by limiting our study to cases with wind speeds mainly above 3 ms^{-1} (see 1.) the well-mixed assumption is valid most of the time.

3. The scaling functions and parameters used in the Monin-Obukhov similarity theory should be updated using the results from the SHEBA experiments. I am aware of at least one paper from that field project that improves the characterization of sensible heat fluxes in the Arctic (Grachev et al., 2005, Boundary-Layer Meteorology, 116, 201-235). The authors should use that improved parameterization.

We now use the scaling functions by Grachev et al.

Sect. 3.3 (below Eq. 3): “(...) the *Psi*-functions for momentum and heat are chosen according to Grachev et al. (2007).”

4. The real test of the results is the RMSE from the model predictions. The RMSE values are 3 to 4 °C which are disappointingly large. The authors do a good job explaining sources of error but the reader is left wondering if a better job could not have been done in isolating the main error source. I think it is necessary to perform a detailed case study where some of the error sources from the reanalysis and microwave satellites can be removed. Why not pick a case of completely clear skies and use MODIS to deduce openings in the sea ice. In March the visible images can also be used for half of the day. Also there are frequency overpasses every day at high latitudes which

allow higher temporal resolution. This approach may fail if the sea ice is highly broken at a scale less than 250 metres but I think good cases can be found. Also pick a case with a simple synoptic-scale weather pattern so that the regional wind field will be similar to the radiosonde measurement at the station. By performing a few case studies where many of the error sources are reduced you might be able to make a stronger conclusion on the quality of your model.

We agree that it is necessary to identify the main error sources. However, we think that a detailed case study, as you suggested, would exceed the framework of this study. Instead, we performed a sensitivity study using all trajectories. We successively used a constant ice concentration, ice surface temperature and wind speed for the calculations. By varying the constant values we could separate the impacts of the different variables on the air temperature variability and thus identify the main error sources. The results are presented in Sect. 5.1:

Wind speed:

“Increasing the wind speed by 1 ms^{-1} in a sensitivity study (not shown) caused changes of the model temperature by up to 1°C for individual trajectories. However, the mean impact on the correlation and RMSE for the ensemble of trajectories was found to be small.”

Ice concentration:

“The impact of a constant error in the ice concentration of 5 % was investigated in sensitivity studies (not shown) and found to be small causing model temperature changes of less than 0.5°C .”

Ice surface temperature

“The impact of these large uncertainties is investigated by assuming a constant offset between MOD29 and real ice surface temperatures of 1°C . The average changes in the modeled temperature were of the order of 1°C resulting in changes of the bias and RMSE of up to 1°C (not shown). This means that the largest source of uncertainties in the used methods is due to inaccurate ice surface temperatures which are mainly caused by inaccurate trajectory positions and radiative effect from undetected clouds.”

Thus, we conclude that surface temperature uncertainties have the largest impact on the model results. We attribute these uncertainties to uncertainties in trajectory positions and radiative effects of undetected clouds:

“In addition, the considered cases may still contain clouds which notably influences the ice surface temperature (Vihma and Pirazzini, 2005). There are uncertainties concerning the cloud mask and fog is sometimes not classified as clouds (Hall et al., 2006). Furthermore, even if there were no clouds present during the overpass of the satellite there might still be cloudy conditions at the time of the trajectory path. An attempt to use cloud data from the reanalyses turned out to be impracticable due to the larger grid sizes. Additional uncertainties arise because of the inaccurate trajectory positions. A displacement of 20 km can cause uncertainties in the MOD29 ice surface temperatures of up to 2°C (not shown).”

5. The radiative cooling of the boundary layer is being ignored in the model. Why not include a fixed radiative cooling rate (it probably does not change much if the skies are clear) to the model?

We added a constant cooling rate of 2 °C/day to the model (see Eq. 1) as in Vihma et al. (2003). The effect was, however, small.

“Radiative cooling of the air column is also accounted for assuming a constant cooling rate c of 2 °C per day as in Vihma et al. (2003).”

Recalculations for the three methods are done but only cases with higher wind speeds are included (see 2.). Furthermore, scaling functions by Grachev et al. are used (see 3.) as well as a constant cooling rate. The results for the radius of impact notably improve, giving a clearer result of about 200 km for Barrow and Alert using all three methods. The new results are included in Sect. 4.5:

“For Barrow, biases and RMSE decrease by about 1 °C for trajectory lengths between 2 and 10 h and remain nearly constant for larger lengths using the AT method. The minimum RMSE using the IST method is found for trajectory lengths of about 10 to 20 h. The bias from IST increases for shorter trajectory lengths from -1 to 4 °C. Both methods suggest a value $R_t=10$ h for the characteristic time scale which corresponds to $R=180$ km for an average wind speed of 5 ms^{-1} . For Alert, the results using the AT method improve only slightly for longer trajectories. Distinct changes can be found in the curves for bias and RMSE of the IST”

“These results are supplemented by explained variances calculated using the TV method (Fig. 11). (...) the largest impact in this method is also seen in the first 10 h where the slope of the curves is the largest. Therefore, a radius of main impact R_t can be defined by relating it to the region with the largest slope of the curves. By this definition, is reached at the transition from steeper to shallower slopes. This transition is pronounced for all stations at a trajectory length of 10 h which is consistent with the results from the AT method.”

The implemented changes also affect the RMSE and the bias. For Barrow, the RSME decreases from about 3.5 °C to less than 3 °C ($H=350$ m) while changes of the biases are small. For Alert and Tara, the biases decrease however from about -1 to -2 °C. Thus, the induced changes do not affect the main error sources which we attribute to undetected clouds (see above). We agree that 3 °C for RMSE is large but not too large to draw some interesting conclusions. We think that the explained variances are an important measure for the radius of impact.

6. How did you handle cases when the measured surface winds were calm at the station even though the reanalysis had a non-calm wind which was used on the back trajectory calculations? This could be another source of larger RMSE.

Our analysis of the soundings at Barrow (new Fig. 1) shows that calm winds below 1 ms^{-1} amount to less than 5 % of the cases. Because of the small occurrence, those cases with calm winds at the station but higher wind speeds in the reanalysis are not treated explicitly but are already included in the discussion of erroneous trajectory positions in Sect. 5.1.

7. The results are restricted to mostly clear sky cases. Quite different conclusions are possible, particularly under low altitude cloudy skies when large downward IR radiances occur and if stronger winds are mixing the BL to greater depths. I suggest adding “clear skies” somewhere in the title of the paper.

The title reads now:

“The impact of heterogeneous surface temperatures on the 2-m air temperature over the Arctic Ocean under clear skies in spring”

8. In Figure 4 please plot the actual measured air temperature at Tara, the plots only show model output.

The figure now also shows the measured temperature at Tara.