Answer to Referee 2

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

1. Page 3016, lines 1-3: What if A = 80% within a 4x4 km2 MOD29 pixel? I guess, in this case we are not talking about actual "ice surface temperature"? Could the separately involved passive- microwave derived ice concentration be employed to correct Ti in such situations? In any case, this seems like another source of error to be addressed.

We now use a linear weighting of the MOD29 derived surface temperatures with the different ice concentration data sets and the water temperature. The interaction of brightness temperatures of different surfaces in the same pixel is non-linear but since the ice concentrations are mostly above 90 % (Fig. 6) a linear weighting is a good approximation. A description is included in the Sect. 2 (par. 4):

"The measured MOD29 temperature, however, represents the surface temperature of a whole pixel and is also influenced by open water areas in that pixel. Therefore, the MOD29 surface temperatures are linearly weighted using the ice concentration products to obtain a better estimate of the actual ice surface temperatures."

2. Page 3019: While the "Box model" includes standard ABL calculations, subgrid-scale effects seem to have been ignored. What about the role of pressure ridges and ice concentration in determining the (effective) roughness length (see e.g. Stössel and Claussen, 1993)? This should be added to the list of limitations or sources of errors with the chosen methods.

We have partly addressed this topic in Sect. 5.2 already where we pointed out that we used constant roughness lengths since we did not have enough information about the sea ice topography. We also described already that greatly different values for the roughness lengths had been tested which had only a small impact on the correlations. The cited papers already address the role of pressure ridges (e.g. Garbrecht et al. 2002) and ice floe edges (e.g. Lüpkes and Birnbaum 2005). Note that parameterizations of sea ice roughness as a function of sea ice concentration have a large influence only in the marginal ice zones or in summer Arctic sea ice conditions where the topography is dominated by ice floe edges and/or edges at melt ponds (Andreas et al. (2010) and Lüpkes et al. (2012)).

"Additional uncertainties which might be responsible for the large RMSE arise, for example, from the used roughness lengths for momentum z_0 and heat z_T which are set constant in the model (…) For this reason, different constant values of z_0 and z_T over sea ice have been tested but the impact was moderate on both RMSE and biases and the explained variance changed only little. However, it cannot be excluded that variable values accounting for the sea ice topography (Andreas et al., 1984; Garbrecht et al., 1999, 2002; Vihma et al., 2003; Guest and Davidson, 1987; Lüpkes and Birnbaum, 2005, Stössel and Claussen,1993) would have a larger impact. This cannot be tested here because the sea ice topography is dominated by pressure ridges in regions with large sea ice concentration and topography data are not available. Estimating the variability of drag coefficients by parameterizations accounting for sea ice concentration (Andreas et al., 2010; Lüpkes et al., 2012) shows that its impact on drag coefficients is only small in our

case. This would be different during summer or in the marginal sea ice zones where the surface topography is determined by ice floe edges and edges at melt ponds so that it can be parameterized as a function of sea ice concentration as described in the above mentioned literature."

3. Page 3021, lines 21-23: Since errors accumulate along a trajectory, the impact of the "large uncertainties" of the far-field remain after passing the near-field, where "the uncertainties are" (presumably) "much smaller". Also, according to Fig.2, on 20. April, there is a near-field difference in wind direction (or angle of trajectory) between ERA and JRA of about 60 degrees, which most likely amplifies the uncertainties considerably.

We do not think that errors always accumulate along the whole trajectory because there is a damping of the errors. If we imagine, for example, a trajectory starting at a distance of 1000 km with two different air temperatures of +20 and -20 °C the fluxes in the two cases will lead to and adjustment of the air temperature of the drifting parcel so that an equilibrium temperature is reached after a certain time. The corresponding spatial scale is the radius of impact which we investigate in Sect. 4.5.

The figure shows indeed examples with large angle differences, on average, however, these differences are much smaller. We added "on average" to the corresponding sentence:

"The large uncertainties in the trajectory positions cause large uncertainties in the estimation of the impact of remote areas but in the near environment of about 100 km the uncertainties are ***on average*** *much smaller."*

4a) Page 3022, lines 7-16: For this example, it would be interesting to investigate the impact of ice concentration by setting it to a fixed value, e.g. 100%, and redo this for ice surface temperature to separate their contribution to the air temperature variability, respectively.

We appreciate your idea. Instead of applying it only to the example 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. 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)."

4b) It would also add to the value of this paper if the corresponding near-surface (2- or 10-m) air temperature of the ERA and JRA analyses were looked at for comparison. This could lead to an insightful discussion on the impact of the specified surface boundary conditions on the near-surface air temperature in the respective analyses.

The corresponding ERA 2-m temperatures have been plotted in Fig. 5 and their evolution is described and discussed in Sect. 4.2:

"The corresponding ERA 2-m temperatures are too high along the trajectory path with a value of -12.6 °C arriving at Tara. This example shows the important role of the specified surface boundary conditions of a model on the calculated air temperature evolution. While the box model which uses ice concentrations and ice surface temperatures derived from remote sensing data reproduces the measured 2-m air temperature quite well the temperature of the reanalysis is about 4 °C too high. This is probably due to the sea ice boundary conditions in ERA-Interim with fixed values for the ice thickness of 1.5 m (White, 2006) and for the ice concentration of 100 % north of 82.5° N (Inoue et al., 2011) which reduce the surface temperature variability."

5a) Page 3028, line 22: Whether "thin ice" is counted toward ice concentration or not should actually make a large difference for the sensible heat flux calculations, so I would suggest to elaborate on this topic.

We write now:

"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. Since ice concentration data along the trajectories are above 90 % in most cases (Fig. 6) the effect of an underestimated ice concentration by a few percent in the presence of thin ice can be expected to be small."

5b) The next question is then what the ice surface temperature looks like when thin ice is around?

As you pointed out the MOD29 surface temperature already includes subpixel effects from open water areas. Let us consider an example with 90 % ice concentration, 5 % thin ice and 5 % open water. If thin ice is counted as ice the corrected ice surface temperature will be larger which means increased fluxes over ice. The water area decreases, however, which decreases the net heat flux. If, in the other case, thin ice is counted as water the ice surface temperature and therefore the fluxes over ice decrease while the fluxes over water increase due to the larger area. Therefore the overall effects are small.

Due to the answer to question 5a) we did not include this discussion in the text.

6a) Page 3031, lines 20-26: You tested 4 different "ice concentration products" using the same "erroneous trajectories positions" for each. Why should the latter then be "masking the inaccuracy of the ice concentrations"?

The main reason for large errors are the erroneous trajectory paths so that the changes induced by different sea ice concentration data sets are much smaller. We removed this sentence nonetheless since our sensitivity studies showed that the results hardly change for uncertainties of the ice concentration of the order of 5 % which is why the results hardly differ for different ice concentrations. The sentence reads now:

"The results depend only weakly on the sea ice concentration products although they show significant differences in the sea ice distributions. For example, the correlation coefficient between measured and calculated 2-m temperatures at the different sites changes only by 10^{-4} for a 5 % change of sea ice concentration. "

6b) Next sentence: Assuming that a 4x4 km2 pixel contains leads, the "prescribed ice surface temperature" somehow depends on ice concentration, though not on the one used in equation (1) (see above).

See above (1.)

6c) Next sentence: I am not sure what this sudden switch to a "fully coupled model" is all about; again, the satellite-derived ice surface temperature is affected by subpixel-scale ice concentration.

This sentence is relevant in terms of comparing temperatures obtained using our simple model to those from a reanalysis (see 4b). We rephrased this section for clarification:

"The results depend only weakly on the sea ice concentration products although they show significant differences in the sea ice distributions. (…) This small sensitivity can be explained by the independence of measured ice surface temperatures and ice concentrations. However, in an atmospheric model coupled with a thermodynamic sea ice model, such as ERA-Interim, the ice surface temperature adjusts to the ice concentration and ice thickness and thus changes of the ice concentration would have a larger effect (Lüpkes et al., 2008). "

6d) When it comes to listing reasons for mismatches one should also include possible effects arising from subgrid-scale heterogeneities that could enhance the heat flux (effective roughness length, etc., see above).

See above (2.)

7. Fig.10: The statistics for Tara is based on data from only one month of one year, whereas that of the other 2 stations is based on 2-3 months of several years. One should thus not put these side-by-side on the same figure, unless perhaps separated by a thick line or explained in the caption. The same holds for Figs.7-9.

We agree that the results for Tara are less significant due to the shorter time period. We added the respective time periods to the figure captions and included the following sentence in the Data section:

"As the thermal differences between sea ice and open water surfaces are small in summer, only one month (April 2007) of Tara data was used in the analysis. Despite the short timeseries Tara provides invaluable data since measurements from the Central Arctic are sparse."

8. Page 3013, line 17: "...can cause a temperature change of up to 3.5 K." What horizontal scales are being considered here?

The cited study used a 1 D model and therefore the horizontal scale is not considered. 3.5 K was found in the quasi-stationary solution after 2 days of simulation. We write now:

"Lüpkes et al. (2008) used a one dimensional atmospheric model coupled with a sea ice model to investigate the influence of a change in ice cover on the atmospheric boundary layer temperatures. They found that, under clear skies in winter and for ice concentrations close to 100 %, a change in ice concentration of 1 % can cause a change of the near-surface equilibrium temperature by up to 3.5 K after 2 days of development."

9.Page 3013, lines 17-21: "Reducing the ice cover from 100 to 50 %..." in an uncoupled atmosphere GCM simulation leads to rather unrealistic results. I don't think this citation fits to the other ones presented in this paragraph.

We agree that these results represent a strong idealization and thus removed the citation.

10. Page 3015, lines 10-12: How do the "surface wind fields" of JRA differ from the "10m wind fields" of ERA? Or is the former a typo?

It is indeed a typo. Both wind fields are at 10 m height. The sentence is corrected to:

"Backward-trajectories arriving at the stations are calculated from the 10 m-wind fields of the Japanese 25-year reanalysis (JRA) and of the European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis (ERA-Interim)."

11. Page 3021, line 19: "750 km after 30 h". According to Fig.2, the "separation" is at most 400 km.

Only exemplary trajectories were shown in Fig. 2. The numbers in Sect. 4.1 are averages over all trajectories. We hope that we could clarify this by rephrasing the first sentences of Sect. 4.1 to:

"The trajectories calculated using the different reanalyses show large inconsistencies. Examples are shown in Fig. 3. The paths of all trajectories are compared by calculating the mean spatial distances between JRA and ERA (1.5) trajectories which differ for the three stations."

11. Page 3031, line 5-8: It sounds like this summary sentence on the AT and the TV method is not consistent with what is described in section 3.

The sentence has been rephrased for clarification:

"For the AT method the modeled temperatures at the stations were compared to the measured ones and for the TV method temperature changes between the model temperature at the trajectory starting point and in situ measurements at the stations were compared to mean sensible heat fluxes."

12. Fig.6: is there any reason for the banded structure of the model temperatures?

A "band" consists of successive trajectories with almost overlapping paths. Therefore, the modeled temperatures show almost no variation. The in situ temperatures, however, show a significant cooling or warming during this time period which is not captured by the model. We presume that the observed temperature change might be due to changing cloud cover along the trajectory path during that time period. Since only daily ice surface temperatures are used such a change is not captured by the model. These undetected cloud cover changes are presumably the largest uncertainty in our calculation methods.

12. Other technical comments

Done