Anonymous Referee #2 Received and published: 29 June 2012

General Comments
This paper addresses an important area of development, namely ice surface temperature (IST). I agree with the authors on the potential significance of a IST product viewed as usable as a boundary condition in NWP, much as SST presently is. I would have like to have had more discussion of this nature, about the context and possible uses in NWP in this paper. The authors state there is "a potential for improving model predictions", but how? It is probably harder to achieve than for SST, because I suspect the radiometric temperature of ice is more closely coupled to air temperature, and so prescribing IST from observations (as is done for SST) could have negative impact. So, I think this aspect of the introduction could be strengthened – perhaps give more context from the Stammer reference.

Author statement:"...a potential for improving model predictions."
This comment is not so deep, it merely states that model predictions will improve if the IST observations can contribute to achive better initial boundary conditions for a model.

A few words on the context and perspectives of using IST in NWPs and sea ice models. We agree that the IST is closely coupled to air temperatures in case of constant and well known ice cover. However, ice drift and deformation constantly cause new exposure of water to the atmosphere and thereby changing the heat flux extensively. We have added following to the introduction:

“The drifting of Arctic sea ice constantly cause opening and closing of the sea ice cover and changes in ice cover of only a few percent can influence the heat flux between ocean and atmosphere drastically (Maykut, 1978; Marcq and Weiss, 2012). For a model to produce a realistic initial surface temperature boundary fields, detailed information of ice concentration and ice drift is needed. The ice concentration fields that are assimilate by e.g. the global deterministic NWP model at ECMWF, have uncertainties of up to 10% (Andersen et al., 2006) and contribute to surface and air temperature uncertainties of several degrees (Lüpkes et al., 2008). The sparsely distributed Arctic buoy observation network can not resolve these variations on the spatial scales on which the changes are occurring, thus emphasizing the potential of using satellite observations to estimate Arctic ice surface temperatures.”

New references are added (see below)

IST has been determined from AVHRRs before (by Comiso, for example) and for MODIS (part of the LST product), so what is the particular contribution of this paper? There is some new validation data collected and matched to Metop AVHRR, which is nice since there are >20000 points (although possibly only 2000 independent points if I understand correctly that multiple satellite points go with each in situ). But the algorithm is not a new formalism (it is a conventional split window) and (although I find this hard to believe) doesn’t use coefficients designed for Metop but for NOAA12. To me, this makes the paper premature. To make this work, there needs to be a way successfully define the retrieval coefficients for different sensors. The authors do define new coefficients by regression to the in situ, but then "validate" these against the same in situ. Of course, this does not prove that successful coefficients can be defined from match ups since by design the results improve when applied back to the data from which they arrive. The good results found are not convincing evidence that such results can be representative of the true errors.

I encourage more work on this topic, for sure. But a significant advance would be to report more than the collection of new data points and application of an old algorithm. Here are some proposals to make the paper of real significance: 1. define Metopspecific coefficients by regression against routinely available in situ data that could be routinely applied without special field campaigns, and then test this against the new data matched in this study to give an independent test of the retrieval performance, OR (and probably better) 2. define Metop coefficients by radiative transfer modelling and show the degree to which this works, OR (best) 3. move on to cutting edge algorithms like optimal estimation. In any case, something different than using coefficients from the wrong sensor.

We have specified the two most essential IST data sets available and accounted for the unique contribution of this METOP IST product. Following text is merged to the introduction in which also a few other edits are done:

"The Pathfinder data set is well suited for climatologically studies, but can not be used for recent or real time ice surface temperature analysis, due to irregular data set updates. Further, the Polar Pathfinder spatial resolution is 5km, which makes it less suitable for fine scale mapping and analysis. The MODIS IST product has very similar characteristics to the METOP IST product (see section 6), with product timeliness and sensor continuity as the main differences. Timeliness and data continuity are essential issues for the model communities to gain from data assimilation schemes (Stammer et al., 2007). The MODIS sea ice products have time lags of days, from observation to product availability, and the timeliness of present IST product is a couple of hours. The METOP AVHRR data stream, that is used for this IST production, is guarantied continuity and is scheduled until at least 2020, in contrast to the MODIS data stream that will end with the current Aqua and Terra missions."

Yes, each cloud free MIST data inside the 4x4km area is accounted for individually – giving approximately 20000 cases.

The error statistics for the MUisar data set was analysed using both mean of all induviduals, mean and median MIST values, without finding clear indications of which measure to use. The standard deviation of errors were practically equal and we decided to threat all MIST data individually. In figure 6 the MIST data are plotted with minimum, maximum and median values and the MISTnewcalibration is plotted as the average value. This is specified in section 5 with this comment:

"This validation strategy was based on experience from MUisar data. The MUisar error statistics was analysed using both mean of all induvidual data pairs, mean and median MIST values, without clear indications of a best measure. Thus, it was decided to threat all MIST-OBS data pairs individually."

We agree that it may not be optimal to adopt the NOAA12 calibration coefficients for the METOP IST algorithm. None the less, this solution seems to be equally good as other available solutions for the time being. We believe that the optimal calibration of the IST algorithm is achieved from an empirical relationship with in situ skin temperature data from positively cloud free conditions. Presently, we do not have sufficient in situ skin temperature data to perform such a calibration and we are left with the options of using either simulated top-of-atmosphere brightness temperatures for Arctic conditions or buoy and ship data for an empirical

calibration.

The NOAA12 calibration performed by Key et al. is retrieved from simulated brightness temperatures at the AVHRR infrared bands in Arctic conditions. This calibration is therefore applicable for the METOP algorithm, that is based on the same instrument as the NOAA12 AVHRR instrument, with identical band width and nearly identical spectral response functions. So, one can say that the calibration is spectrally specific, rather than specific for a certain satellite instrument. In present paper the Key calibration is proven to be competitive with a calibration to the full buoy data set. This is explained in section 4.5, where a clarefying comment is added (see next):

The 're-calibration tests' as they are called in the paper, are not performed to come up with new calibration coefficients, merely to test the performance of the applied algorithm. By 'recalibrating' to the entire buoy data set one find the best possible empirical calibration. This is now specified in the text, by adding this:

"... Hence the re-calicration is not performed to establish new calibration coefficients, but to compare the best possible empirical calibration from the Arctic buoys and ISAR measurements to the operating algorithm. If the re-calibration tests do not improve the performance significantly, the dominant errors are associated with other issues than algorithm calibration."

As mentioned, the re-calibration is used to estimate the error contribution from a possible poor NOAA12 calibration. However, the result show that we practically gain nothing from re-calibrating the algorithm to the full OBSarctic data set, thus indicating that the NOAA12 calibration is working well – or at least, contribute much less to the overall error than other error sources. A clearifying comment is added in the text in section 5 (Results):

"... i.e. indicating that the adopted NOAA 12 calibration coefficients works well for the METOP AVHRR instrument and that erroneous cloud screening is a dominating source of error."

However, we do agree that a new algorithm/calibration must be part of the future IST algorithm development. In this paper we believe that the important issue is to present the new product, to make sure the calibration working well and to present the product performance thoroughly. We have added following in section 3:

"The NOAA12 calibration coefficients are retrieved from RTM modelled brightness temperatures for the AVHRR infrared channels and related to model skin temperatures (Key et al., 1997). The channel centre, -width and spectral response function of the NOAA12 and METOP AVHRR instruments are nearly identical. We therefore considered the applied calibration equally valid for METOP AVHRR data than for NOAA AVHRR data.

Specific comments
Is ECMWF 2 m the right comparison to the IST? There is an ECMWF skin temperature field that is probably more like-for-like? Also, show the difference field rather than side by side. I had to look carefully to see there are actually very big differences.

It is a good point and we agree that the surface temperature and 2m temperature fields can be rather different for sea ice. The NWP panel in the figure is changed to surface temperatures (See illustration 1 below). The

discriptive text is adjusted accordingly. We prefer not to produce a difference image, as that will indicate the plots are quantitatively comparable, which we argue in the text that they are not. We prefer to leave the two panels as they are, for qualitative comparison.

The scatter plot figures would be much easier to view if plotted, first of all, square, and second, with the same axis range on the horizontal and vertical. Your bias was -3 K. Was the Hall bias mentioned +2 K, as written, or is it also negative, which it seems to read like? Although an agreement in bias would be a coincidence if using coefficients from a different sensor.

We perfer to leave the scatterplot symbol as they are (small diamonds) to avoid complete smering of the plot. The axis are changed to equal ranges. The Hall bias was -2.1K, as you correctly interpreted it. It is now corrected. With reference to the calibration points above, we do not consider the values coincidencial.

REFERENCES ADDED:

Andersen, S., R. Tonboe, S. Kern, and H. Schyberg. Improved retrieval of sea ice total concentration from spaceborne passive microwave observations using numerical weather prediction model fields: An intercomparison of nine algorithms. REMOTE SENS ENVIRON, 104, 4, 374–392, doi:10.1016/j.rse.2006.05.013, 2006.

Lüpkes, C., T. Vihma, G. Birnbaum, and U. Wacker. Influence of leads in the sea ice on the temperature of the atmosphere boundary layer during polar night. GEOPHYS RES LETT, 35, doi:10.1029/2007GL032461, 2008.

Maykut, G. A. Energy Exchange Over Young Sea Ice in the Central Arctic, J GEOPHYS RES-OCEAN, 83, C7, 3646–3658, doi:10.1029/JC083iC07p03646, 1978.

Marcq, S. and J. Weiss. Influence of sea ice lead-width distribution on turbulent heat transfer between the ocean and the atmosphere. THE CRYOSPHERE, 6, 143–156, doi:10.5194/tc-6-143-2012, 2012.

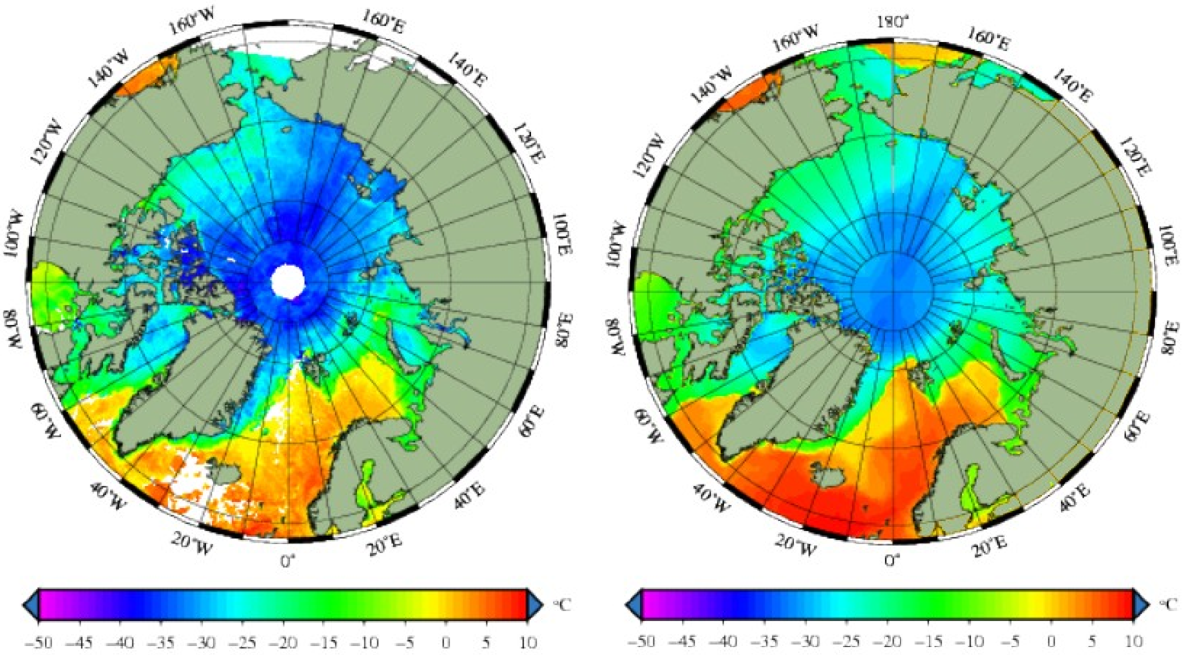

Illustration 1: New figure 4 with NWP surface temperatures in the right panel