Reply to RC2 for the manuscript "An Extended Global Earth System Data Record on Daily Landscape Freeze-Thaw Status Determined from Satellite Passive Microwave Remote Sensing" by Youngwook Kim, J. S. Kimball, J. Glassy, and J. Du

Dear Anonymous Referee #2, thank you for your constructive comments on our manuscript. Please find below our responses to all of the review comments. The resulting responses and manuscript changes (in *italic*) are provided immediately after each review comment (in ***bold***).

Responses to Referee #2's comments:

The paper presents a new algorithm for freeze-thaw detection from passive microwave data. Thresholds are defined using an empirical linear regression relationship and combined with a cosine function for different weighting of values close to 0C. It's a modified version of Kim et al. 2011. The results are compared to WMO station data, reanalyses data, river ice break-up records and another passive microwave (same dataset) based melt dataset from the Greenland ice sheet. It thus constitutes a classical research paper. It is not clear to me how the boundary between data and research paper is defined for ESSD, but I would see it more on the research side.

What seems to be not within scope (as written on the ESSD web page) is comparison to other methods. By comparing it to the Greenland melt dataset, it is actually compared to another method of freeze/thaw detection on the same dataset. It is actually nowhere stated in the paper (and also not possible to guess from the reference Mote 2014) that it is based on the same records and an alternative method was applied.

Despite these concerns, the paper and dataset would be acceptable as a classical research paper (in a different journal) with following amendments/clarifications:

The primary purpose of the paper is to document a new FT-ESDR global data record. The FT-ESDR database described in the paper represents a substantial modification of an earlier FT-ESDR release, while the differences between the two datasets are described in the paper. The paper includes a global validation and performance assessment using a range of ancillary observational datasets, including global WMO weather station records, lake and river ice observations, and a Greenland surface melt record. Detailed algorithm quality assurance (QA) and quality control (QC) metrics are presented describing global product performance. The WMO station validation results and QA and QC metrics are included as part of the FT-ESDR database and were therefore considered appropriate for this journal. The cryosphere data comparisons were also considered appropriate for this journal because they illustrate FT-ESDR performance and synergy in relation to other cryosphere datasets.

Comparing the FT-ESDR derived non-frozen season metric with the Greenland melt season record allows for an alternative FT accuracy assessment over permanent snow and ice areas in Greenland, where validation is severely constrained by a sparse WMO station network. For better clarity, we added the text (Ln 292-294):

"*The Greenland melt record provides a synergistic geospatial database for FT-ESDR assessment over northern snow and ice dominant land areas where in situ weather station networks are extremely sparse.*" We also added additional information on the satellite data and algorithm underpinning the Greenland surface melt database, and differences from the FT-ESDR methods (See Reviewer 2, specific comment 3 below for our detailed response).

***1)* provide some graphics which illustrate the difference between the algorithm in this study and the one in Kim et al. 2011**

In accordance with the reviewer recommendations, we added the following additional statement and equation (1) in section 2.3 (Ln 162-172): "*A seasonal threshold algorithm (STA) was applied to classify daily FT state dynamics from SSM/I daily 37GHz (V-Pol) T_b time series to construct the prior FT-ESDR release (Kim et al., 2011). The STA used a spatial and seasonal scale factor,* $\Delta T_b(x, t)$*, defined for an observation acquired at location (x) and time (t) as:*

$$\Delta T_{bp}(x,t) = \frac{(T_{bp}(x,t) - \mathrm{FrozRef}(x))}{(\mathrm{ThawRef}(x) - \mathrm{FrozRef}(x))} \qquad (1)$$

where T_{bp} (x,t) [K] is the T_b retrieval acquired at location (x), time (t), and polarization (p); FrozRef(x) [K] and ThawRef(x) [K] define Tb retrievals under respective frozen and thawed landscape reference states. The STA threshold was derived annually on a grid-cell-wise basis using an empirical linear regression relationship between the satellite $\Delta T_b(x, t)$ *and daily SAT records from the National Centers for Environmental Prediction and National Center for Atmospheric Research reanalysis (Kanamitsu et al., 2001).*"

***2)* explain why you use different thresholds for each year**

The grid cell-wise MSTA FT threshold was defined annually using an empirical linear regression relationship between the satellite T_b retrievals and weighted ERA-Interim SAT. These T_b thresholds provide a consistent and continuous long-term global data record of landscape FT state dynamics and allow for mitigating FT classification degradation attributed from warming SAT by the use of different thresholds for each year. In accordance with the reviewer recommendations, we added the following additional statement in section 2.3 (Ln 201-211): "*The grid cell-wise* T_b *thresholds and annual calibration used for the MSTA reduce the potential influence of spatial and temporal variations in climate and land surface conditions on FT classification accuracy, and promote greater consistency in global product performance over the long-term record. The daily SAT record used for the MSTA calibration approximates FT related shifts in near surface air temperatures using a fixed 0.0°C temperature threshold. In contrast, The 37 GHz* T_b *retrievals used to construct the FT-ESDR are sensitive to land surface dielectric variations stemming from FT related shifts in surface soil, snow cover and vegetation elements within the sensor footprint. The SAT and* T_b *records represent different physical phenomena with different spatial and temporal characteristics; these factors may contribute uncertainty to the MSTA thresholds and classification accuracy, which are represented in the FT-ESDR accuracy quality assurance metrics (Section 2.4).*"

***3)* provide details on data type and algorithm of Mote (2014), discuss the difference in algorithm along the identified differences**

Greenland Melt onset is identified from SMMR, SSM/I, and SSMIS 37GHz (H-pol) T_b with dynamic thresholds. These thresholds are generated using a microwave emission model to simulate 37 GHz (H-pol) T_b variations associated with a melting snowpack. Scattering coefficients needed for the microwave emission model are empirically derived for each year from T_b observations prior to the onset of melt. The microwave emission model is used to simulate summer melt conditions by adding one percent volumetric liquid water to each grid cell. The resulting T_b values are used as thresholds to distinguish melt from non-melt conditions during the summer of a given year. For a detailed description of this approach, see Mote and Anderson, 1995 and Mote, 2007 (http://nsidc.org/data/docs/measures/nsidc-0533/index.html). In accordance with the reviewer recommendations, we added the algorithm of Mote (2014) in section 2.5 (Ln287-292): "*The Greenland melt record is derived using similar satellite 37GHz* T_b *retrievals as the FT-ESDR, but the* T_b *retrievals are used with a microwave emission model to distinguish either surface melt or no surface melt categorical conditions for each 25-km grid cell (Mote, 2014). The seasonal progression in proportional area and annual variation of surface melt and FT-ESDR defined NFS conditions were compared over the Greenland ice sheet area indicated from the Land-Ocean-Coastline-Ice database (Knowles, 2004).*" And we also added the following statement in section 3.6.2 (Ln 504-508): "*The FT-ESDR results also show a 4.3 percent larger non-frozen area relative to active melt areas; these discrepancies may reflect one or more factors, including different classification algorithms and environmental thresholds, and different sensitivities of the 37GHz V-pol and H-pol* T_b *retrievals used in the respective FT-ESDR and Greenland melt algorithms (Mote et al., 2014).*"

***4)* compare your results with those of Kim et al. 2011. Where/when does the new method provide better results?**

The new FT-ESDR showed a 0.6% and 1.9% mean annual classification accuracy improvement for respective PM and AM orbital crossings over the previous FT-ESDR version for the overlapping 1979-2012 record and vegetated land domain. Areas with enhanced accuracy include Central Asia, and North and Central Europe (Figure below). In accordance with the reviewer recommendations, we added the following additional explanation in the main text (Ln 322-326): "*These results showed general improvement in mean annual classification accuracy of 0.6% (PM overpass) and 1.9% (AM overpass) over prior FT-ESDR versions for the same period (1979-2012) and vegetated domain covered by the initial product release (Kim et al. 2011). Areas with improved accuracy include Central Asia, Northern and Central Europe, and South America (not shown).*"

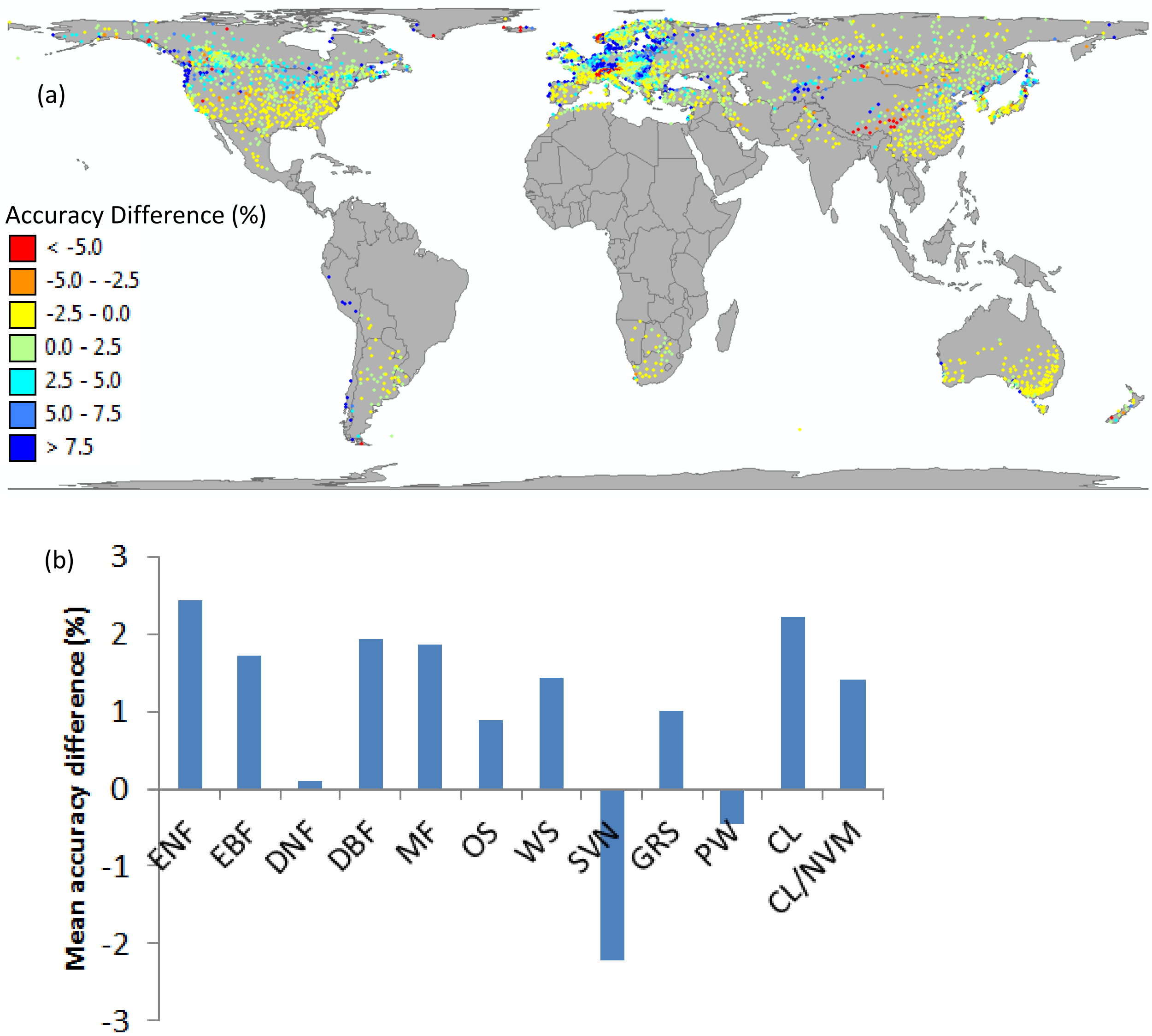

Figure. Spatial distribution of mean annual FT classification accuracy difference (a; New FT-ESDR accuracy - prior FT-ESDR accuracy) for PM overpass results for selected year 2008 and vegetated land areas. Mean annual FT classification accuracy difference for each land cover type (b). Land cover classes represented include evergreen needleleaf forest (ENF), evergreen broadleaf forest (EBF), deciduous needleleaf forest (DNF), deciduous broadleaf forest (DBF), mixed forest (MF), open shrubland (OS), woody savanna (WS), savanna (SVN), grassland (GRS), permanent wetland (PW), cropland (CL), and cropland/natural vegetation mosaic (CL/NVM) categories.

***5)* line 333: there seem to be differences actually all winter, not only transition period**
To address reviewer concerns we revised the text (Ln 376-378): "*The results of the FT classification error matrices suggest a mixture of both freeze and thaw classification errors during the seasonal transition periods and winter seasons.*"

***6)* Table 2: spell out abbreviations in the table, its difficult to read otherwise**
In accordance with the reviewer recommendations, we spell out all abbreviations in the table 2.

References

Kim, Y., J. S. Kimball, K. Zhang, K. Didan, I. Velicogna, and K. C. McDonald (2014a). Attribution of divergent northern vegetation growth responses to lengthening non-frozen seasons using satellite optical-NIR and microwave remote sensing, *International Journal of Remote Sensing*, 10.1080/01431161.2014.915595

Kim, Y., J. S. Kimball, J. Glassy, and K. C. McDonald (2014b). MEaSUREs Global Record of Daily Landscape Freeze/Thaw Status. Version 3. [1979-2011]. Boulder, Colorado USA: NASA DAAC at the National Snow and Ice Data Center. http://dx.doi.org/10.5067/MEASURES/CRYOSPHERE/nsidc-0477.003

Kim, Y., J. S. Kimball, K. Zhang, and K. C. McDonald (2012). Satellite detection of increasing northern hemisphere non-frozen seasons from 1979 to 2008: Implications for Regional Vegetation Growth. *Remote Sensing of Environment*, 121, 472-487

Kim, Y., J. S. Kimball, K. C. McDonald, and J. Glassy (2011). Developing a Global Data Record of Daily Landscape Freeze/thaw status using satellite passive microwave remote sensing. *IEEE Transactions on Geoscience and Remote Sensing*, 49 (3), 949-960

Watanabe, M., G. Kadosaki, Y. Kim, M. Ishikawa, K. Kushida, Y. Sawada, T. Tadono, M. Fukuda, and M. Sato. 2011. Analysis of the Sources of Variation in L-band backscatter from Terrains with permafrost, *IEEE Transactions on Geoscience and Remote Sensing*, 50 (1), 44-54