

Seasonality and Aerosol Optical Thickness Affect Landsat 7 and 8 Harmonization Performance

Galen Richardson¹, Anders Knudby¹, Elisha Richardson², Wenjun Chen³

¹ University of Ottawa, 60 University Private, Ottawa, ON, Canada, galen.richardson@uottawa.ca | aknudby@uottawa.ca

² Carleton University, 1125 Colonel By Dr, Ottawa, ON, Canada, elisharichardson@cmail.carleton.ca

³ Canadian Centre for Mapping and Earth Observation, 580 Booth St, Ottawa, ON, Canada, wenjun.chen@nrcan-rncan.gc.ca

Keywords: Harmonization, Cross-sensor Calibration, Seasonality, Aerosol Optical Thickness, Landsat

Abstract

Sensor harmonization is required to produce consistent Landsat imagery for long-term change detection. This study investigated the effect of seasonality and aerosol optical thickness (AOT) on linear harmonization functions, which are frequently used to create consistent Landsat 7 ETM+ and Landsat 8 OLI time series data. We found that training harmonization functions with pixels that have low or average AOT can greatly reduce the difference between near-coincidental Landsat 7 and Landsat 8 observations, and that seasonally trained harmonization models outperform models trained on year-round data. We assessed the effect of ETM+/OLI sensor harmonization on forest type classification with a Random Forest model, and found that seasonally harmonized imagery provided more consistent classification maps than the alternatives. This study illustrates important details related to the creation of harmonized datasets and is a significant step toward creating more consistent Landsat data for long-term change detection analysis.

1. Introduction

1.1 Background

Sensor harmonization, the cross-calibration of imagery recorded from different image sensors, is essential for making a consistent time series of Earth Observation (EO) data that spans multiple sensors. One of the primary EO missions that has benefited from sensor harmonization is the Landsat program, with over 40 years of data collected by five sensors at a 30 m spatial resolution (Wulder et al., 2022). Landsat data has, among other things, been used for monitoring changes in land use and land cover, tracking forest dynamics, and mapping inland water extents over long periods of time (e.g. >25 years) (Wulder et al., 2022; Olthof & Fraser, 2024).

The Landsat program has used the Multi-Spectral Scanner (MSS), Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+), Operational Land Imager (OLI), and OLI-2 to capture images of the Earth's surface since 1972 (Wulder et al., 2022). Some of these Landsat sensors are considered broadly equivalent, such as the OLI and OLI-2 sensors, and the TM and ETM+ sensors (Roy et al., 2016; Richardson et al., 2025). However, the TM/ETM+ and the OLI/OLI-2 sensors have substantially different spectral response functions (SRFs), notably in the near-infrared (NIR) band, and the Level-2 products are produced with different atmospheric correction algorithms, Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) and Landsat Surface Reflectance Code (LaSRC) respectively. These differences need to be reconciled to create a consistent and long-term time series of Landsat data (Roy et al., 2016; Olthof & Fraser, 2024; Richardson et al., 2025).

Because of the matching resolution (30 m), and the long period during which ETM+ and OLI acquired partially overlapping images one day apart, creating harmonization models for ETM+/OLI data has been a focus in EO research, since these models allow for building a time series that include TM and ETM+ data, as well as OLI and OLI-2 data (Roy et al., 2016; Olthof & Fraser, 2024). The objective of most Landsat harmonization research has been to create regional models that

are optimized for making consistent Landsat data for a given study area (Roy et al., 2016; Scheffler et al., 2020; Olthof & Fraser, 2024). Compared to global models, regional harmonization models often perform better since they are sensitive to the surface types and atmospheric conditions of the region they are calibrated on (Scheffler et al., 2020; Richardson et al., 2025).

Given the unique environments present in Canada, creating harmonization models that are tuned to the landscape should allow for change detection studies to be more accurate and reliable. Additionally, Canada experiences substantial surface changes throughout the year, which could adversely affect the harmonization functions optimized for different change detection objectives. Many Canadian change detection studies only consider summer images that were captured +/- 30 days from August 1st to minimize potential seasonal differences due to vegetation phenology (Guindon et al., 2024). These studies could benefit from harmonization functions specifically trained for summer imagery.

Olthof & Fraser (2024) and Richardson et al. (2025) have both proposed harmonization functions over portions of Canada that consider the seasonal differences observed in Landsat imagery. For harmonization models optimized for open water detection, Olthof & Fraser (2024) created ETM+/OLI harmonization models over the Hudson Bay Lowlands that only considered ice-free images. Richardson et al. (2025) found that a regional model over Iqaluit, NU that considered only imagery from July and August outperformed a global model. While this comparison evaluated a seasonal-regional model and a multi-season-global model, understanding the impact of image date selection could improve harmonization model performance.

In addition to seasonality, another way to improve harmonization model performance is stricter quality control for pixels used to train harmonization algorithms. With the availability of large quantities of open-access Landsat data, many EO studies use image composites of high-quality pixels from multiple Landsat images to create more consistent data for analysis (Wulder et al., 2022; Olthof & Fraser, 2024). In addition to cloud filtering, ETM+ and OLI Landsat surface

reflectance data allow users to filter for pixels that have acceptable aerosol optical thickness (AOT). AOT is important for studies that use ETM+ and OLI data, since one of the primary reasons for the different SRFs on the OLI sensor is to avoid atmospheric features within the ETM+ SRFs (Wulder et al., 2022; Roy et al., 2016). Surface reflectance ETM+ data include the *SR_ATMOS_OPACITY* band, which contains an estimate of aerosol optical thickness AOT at 550 nm (Zhang et al., 2022). AOT values ≤ 0.3 are considered clear or average, while values > 0.3 are considered hazy (Zhang et al., 2022).

1.2 Study Objective

The objective of this study is to assess whether the selection of images captured only in the summer improves cross-sensor harmonization performance in that season, and how AOT affects harmonization performance. Additionally, this study will provide insights into the influence of seasonal and year-round harmonization models on classifying forest types in Canada using a Random Forest (RF) model.

2. Methodology

In this section, we first describe how we sampled ETM+ and OLI data using the LEOHS Python package (Section 2.1), then discuss how we built the datasets, selected the harmonization models, and assessed the influence of AOT models (Sections 2.2-2.4), and finally how we assessed the harmonization models (Section 2.5).

2.1 Landsat Imagery Sampling Using LEOHS

LEOHS is a Python package that can be used to sample ETM+ and OLI observations captured within one day of each other and create harmonization functions over user-defined regions with different sampling parameters (Richardson et al., 2025). We used LEOHS version 1.0.3 to create two datasets of one million ETM+/OLI Surface Reflectance (SR) observations across Canada. Both datasets were sampled from images with $< 50\%$ cloud coverage from 2014 to 2021. Intra-day aerosol and surface variability were minimized in the LEOHS tool through a percent difference blue band filter, where instances of greater than 100% difference between ETM+ and OLI pixels were excluded from the output datasets (Roy et al., 2016; Richardson et al., 2025). The year-round dataset consisted of observations from all months, while the summer dataset only included observations from July and August. After the launch of Landsat 8, the Landsat 7 sensor started to acquire data on a continental acquisition strategy that excluded data over northern Canada, Greenland, and oceanic islands, resulting in the image overlap pattern found in Figure 1.

2.2 Dataset Creation

Pixels from the year-round and summer datasets were divided into training and testing datasets (75/25 dataset split) according to a gridded-spatial blocking strategy used to reduce spatial autocorrelation and prevent data leakage when calibrating and assessing the models (Lovitt et al., 2023). The cell size of the grid was set to 100,000 km², and each cell was assigned a different dataset according to the row and column numbers (Figure 2). Cells containing both odd row and column numbers were assigned to the test dataset, while all remaining cells were assigned to the training dataset.

We then appended landcover classes from the SCANFI Landcover product V1.2 (Guindon et al., 2024) to each pixel.

The SCANFI Landcover product spans across Canada's non-arctic landmass at a 30 m resolution for the year 2020 and was created using multi-season Landsat data to achieve an overall accuracy of 82% (Guindon et al., 2024). This product includes several forest type classes: Treed Broadleaf, Treed conifer, and Treed mixed, which are useful for vegetation mapping.

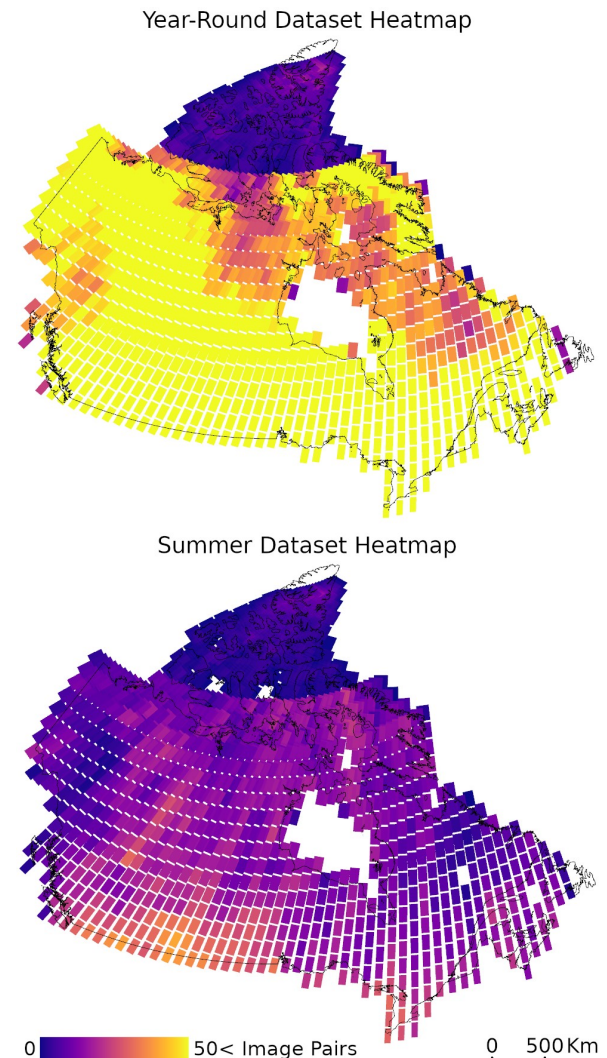


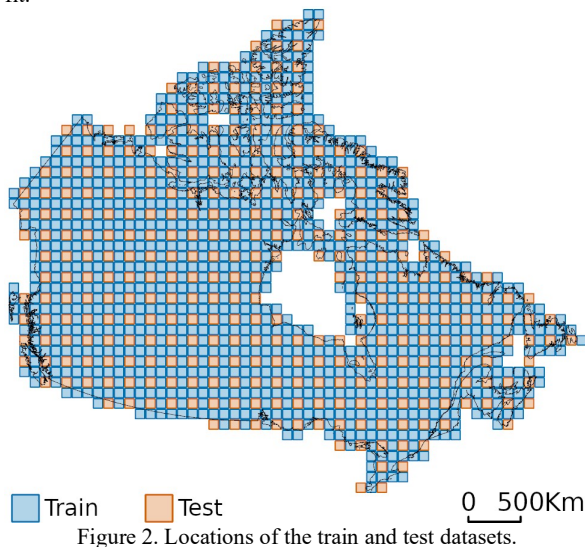
Figure 1. Number of available Landsat image pairs per WRS-2 overlap for the year-round and summer dataset.

2.3 Model Selection and Performance Evaluation

Ordinary Least Squares (OLS) regression is the most commonly used model type for Landsat harmonization, since it has been shown to outperform other models in comparison studies (Roy et al., 2016; Richardson et al., 2025). We therefore trained OLS harmonization models to predict OLI SR values from ETM+ SR data.

The performance of all models was assessed using the root mean squared deviation (RMSD) and the coefficient of determination (R^2), calculated for the harmonized SR value pairs, i.e. the OLI observations and the coincident harmonized ETM+ observations. RMSD quantifies the average magnitude of difference; lower RMSD values indicate a greater similarity of values and better model performance. R^2 indicates how well a

model captures the variance in the data under the assumption of a linear relationship; higher R^2 values reflect a stronger ability of the model to explain variability in the data through a linear fit.



2.4 Sensitivity Testing

To assess the influence of seasonally-specific model development, we trained separate models using summer-only data (July and August) and year-round data. To examine the influence of AOT, we compared training datasets with and without AOT filtering for clear and average AOT pixels ($AOT \leq 0.3$; Section 3.1). Then we compared OLS models trained with and without AOT filtering on the summer test dataset, which included only clear and average-AOT pixels ($AOT \leq 0.3$) to ensure an equal comparison (Section 3.2). As such, we produced a total of four different harmonization models and evaluated them on the one test dataset.

2.5 Harmonization Model Assessment

In addition to reporting overall model performance metrics, we applied the best-performing year-round and summer models to the summer dataset and summarized the average band-level performance across treed SCANFI land-cover classes to investigate land-cover-specific performance (Section 3.2).

We further assessed harmonization performance by determining how it influences forest type classification (Section 3.3). We used Landsat 7 and 8 scenes from July 2019 for an area east of Sudbury, Ontario (path 017, row 027). This region of Canada was selected since it contained all the different forest types in the SCANFI dataset. The images were captured within eight days of each other, both with low cloud coverage, and the Landsat 7 image had an average AOT of 0.09. We created calibration and evaluation datasets containing 75,000 and 25,000 points respectively. The points were selected using stratified random sampling and had a minimum spacing of 100 m between samples. For every point in the datasets, we ensured they had valid Landsat 7 and 8 pixel values and were not located in clouds or cloud shadows. We then trained an RF classifier using default parameters to predict forest and non-forest SCANFI land cover classes in a Landsat 8 scene (Guindon et al., 2024). Finally, we harmonized the Landsat 7 image using the best-performing year-round and summer models, applied the RF model to both harmonized images as

well as to the non-harmonized Landsat 7 image, and evaluated the accuracy of the resulting land cover maps.

3. Results

3.1 AOT Filtering Results

The summer and year-round training datasets consisted of 70% and 65% pixels with $AOT \leq 0.3$ respectively. Tables 1 and 2 show performance metrics for the year-round and summer training datasets with and without AOT filtering. As is evident from these tables, filtering out pixels with high AOT dramatically reduces the RMSD of all bands in both year-round and summer datasets. The effect is strongest in the blue band, which shows the greatest improvement in RMSD. When comparing the year-round and summer datasets, the summer dataset has substantially lower RMSD. A contributing factor to this result is the narrower ranges of ETM+ and OLI reflectance values present in the summer training dataset, which could make the magnitude of errors smaller (Figure 3). Additionally, R^2 values also increase when filtering for $AOT \leq 0.3$, with the summer dataset showing the biggest increase. The increase in R^2 is a further indication that filtering for low to average AOT improves model performance.

When plotting the per-band RMSD per 0.05 AOT bin, the magnitude of differences for all bands is relatively stable for AOT values ≤ 0.3 (Figure 4 A and C). However, beyond the AOT value of 0.3, the RMSD values increase in an inconsistent pattern, with the increase in RMSD being inversely related to wavelength; the blue band exhibits the largest increase in RMSD as AOT increases, and the SWIR1 and SWIR2 bands remain relatively unaffected by AOT (Figure 4 B and D). The metrics in Tables 1 and 2, together with the plots in Figure 4, show that restricting AOT to ≤ 0.3 substantially reduces discrepancies between ETM+ and OLI data and yields values whose variability is more effectively captured by a linear model.

Band	Year-round No Harmonization		Year-round ($AOT \leq 0.3$) No Harmonization	
	RMSD	R^2	RMSD	R^2
Blue	0.0878	0.8516	0.0359	0.8728
Green	0.0754	0.8697	0.0326	0.8872
Red	0.0752	0.8703	0.0330	0.8883
NIR	0.0678	0.8433	0.0379	0.8668
SWIR1	0.0377	0.8224	0.0310	0.8797
SWIR2	0.0274	0.8068	0.0218	0.8916
Mean	0.0619	0.8440	0.0320	0.8810

Table 1. Comparison of year-round training dataset RMSD and R^2 with and without $AOT \leq 0.3$ filtering.

Band	Summer No Harmonization		Summer ($AOT \leq 0.3$) No Harmonization	
	RMSD	R^2	RMSD	R^2
Blue	0.0304	-0.7304	0.0156	0.5512
Green	0.0255	0.1744	0.0145	0.7347
Red	0.0246	0.4847	0.0150	0.8184
NIR	0.0371	0.8565	0.0323	0.8876
SWIR1	0.0310	0.8477	0.0298	0.8645
SWIR2	0.0207	0.8789	0.0202	0.8932
Mean	0.0282	0.4186	0.0212	0.7916

Table 2. Comparison of summer training dataset RMSD and R^2 with and without $AOT \leq 0.3$ filtering.

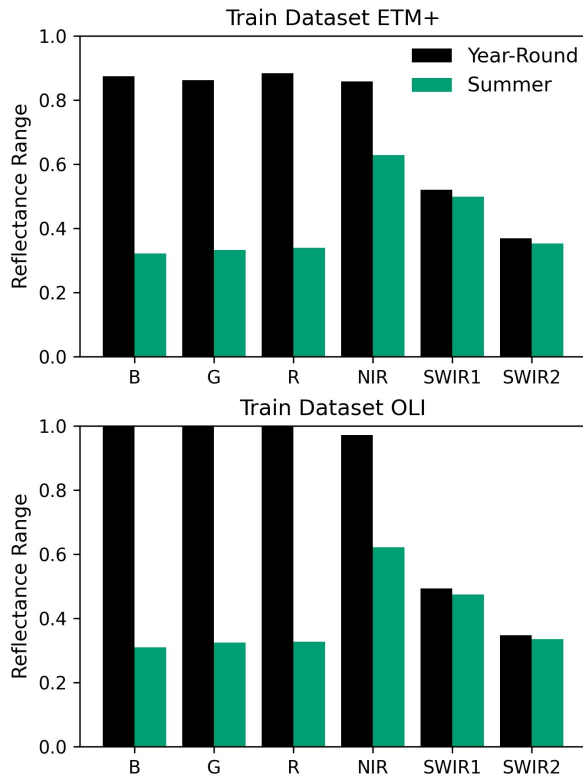


Figure 3. The reflectance range between the year-round and summer training datasets.

3.2 Summer Dataset OLS Evaluation

We evaluated harmonization models on the summer test datasets that were filtered with $AOT \leq 0.3$. We compared OLS models trained on all pixels (Table 3) and those that were trained on only pixels with $AOT \leq 0.3$ (Table 4). The summer models outperformed the year-round models, with the biggest improvements found in the visible wavelengths. Additionally, the models that were trained with $AOT \leq 0.3$ pixels outperformed those that were trained on all pixels, with lower average RMSD and higher R^2 values. The summer OLS model trained on $AOT \leq 0.3$ provides consistent improvements to RMSD and R^2 when compared to no harmonization and only AOT filtering.

When evaluating the harmonization performance per SCANFI class, the summer model trained with $AOT \leq 0.3$ had lower RMSD and higher R^2 than the year-round model (Table 5). The Treed Broadleaf class had the lowest overall RMSD, while the Treed Mixed class had the highest RMSD values.

Band	Year-round OLS		Summer OLS	
	RMSD	R^2	RMSD	R^2
Blue	0.0224	0.0804	0.0160	0.5313
Green	0.0183	0.5780	0.0155	0.6950
Red	0.0185	0.7256	0.0149	0.8207
NIR	0.0350	0.8678	0.0328	0.8837
SWIR1	0.0281	0.8802	0.0274	0.8856
SWIR2	0.0196	0.8996	0.0194	0.9018
Mean	0.0236	0.6719	0.0210	0.7863

Table 3. Comparison of year-round and summer OLS models trained on all pixels.

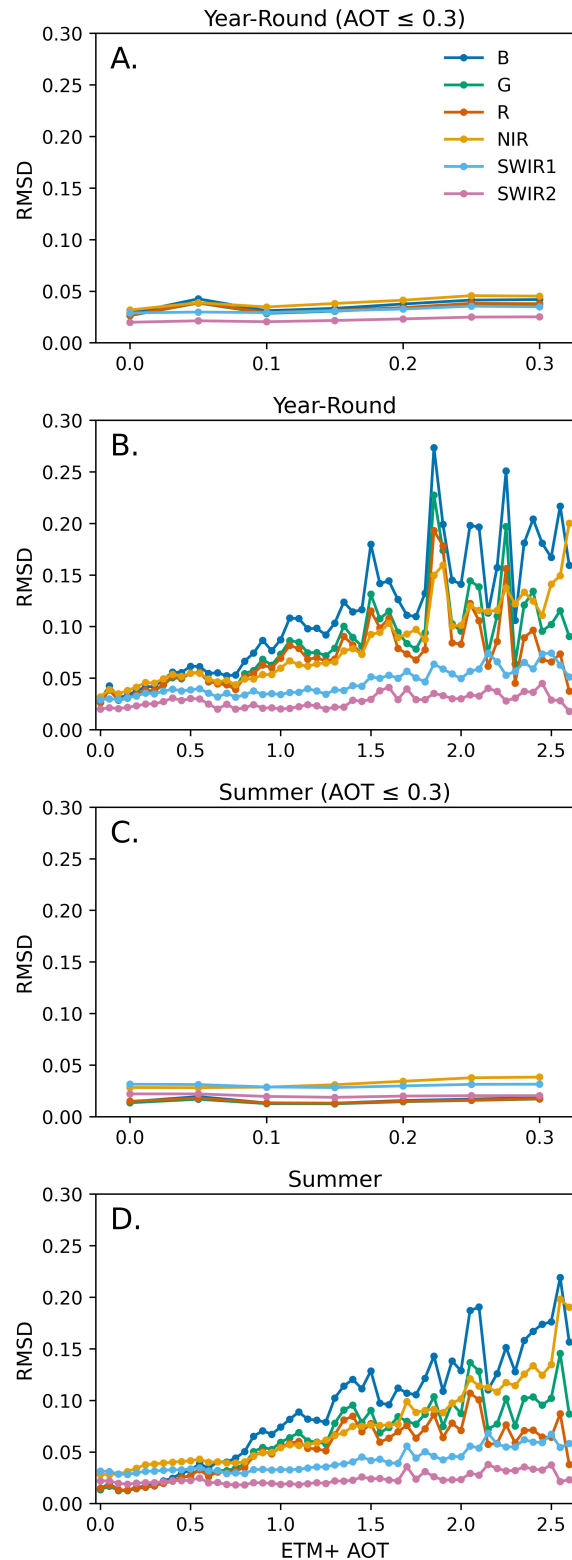


Figure 4. RMSD per 0.05 bins of AOT for ranges 0-0.3 (A/C) and the entire range of AOT (B/D) for the year-round and summer training datasets.

Band	Year-round OLS (AOT ≤ 0.3)		Summer OLS (AOT ≤ 0.3)	
	RMSD	R ²	RMSD	R ²
Blue	0.0164	0.5080	0.0122	0.7255
Green	0.0157	0.6881	0.0132	0.7818
Red	0.0157	0.8026	0.0132	0.8609
NIR	0.0323	0.8872	0.0321	0.8887
SWIR1	0.0273	0.8866	0.0273	0.8868
SWIR2	0.0193	0.9028	0.0193	0.9028
Mean	0.0211	0.7792	0.0195	0.8411

Table 4. Comparison of year-round and summer OLS models trained on pixels with AOT ≤ 0.3.

Land Cover	Year-round OLS (AOT ≤ 0.3)		Summer OLS (AOT ≤ 0.3)	
	RMSD	R ²	RMSD	R ²
Treed Broadleaf	0.0170	0.4221	0.0154	0.6406
Treed Conifer	0.0175	-0.0588	0.0152	0.5180
Treed Mixed	0.0201	0.6788	0.0192	0.7243
Mean	0.0182	0.3474	0.0166	0.6276

Table 5. The average harmonization performance for the treed SCANFI land cover classes.

3.3 SCANFI Classification

The RF model trained on the OLI image had an overall accuracy of 0.71 when classifying the three forest types and a class representing all non-forest landcovers (Table 6). The second-best-performing Landsat image was harmonized using the summer OLS model trained on only pixels with AOT ≤ 0.3, resulting in an overall accuracy of 0.66. Both no-harmonization and the year-round harmonization trained on AOT ≤ 0.3 had similar overall accuracies of 0.64.

When comparing the prediction maps from the different Landsat images, there is a substantial number of misclassified non-tree

land cover pixels in the year-round harmonized and non-harmonized images (Figure 5). Compared to the other variants of ETM+ data, the summer OLS harmonized image better represents the general pattern of the SCANFI map, despite the challenges of differentiating spectrally similar classes. The summer OLS harmonized image overclassified the Treed Broadleaf class and under classified the Treed Mixed class compared to the SCANFI map.

Landsat Image	Overall Accuracy
OLI	0.71
ETM+ No Harmonization	0.64
ETM+ Year-Round OLS (AOT ≤ 0.3)	0.64
ETM+ Summer OLS (AOT ≤ 0.3)	0.66

Table 6. The overall classification accuracy of the RF model on different Landsat images.

4. Discussion

In this study, we compared ETM+/OLI harmonization strategies and investigated the influence of AOT and seasonality on linear regression harmonization models. While the analysis was limited to Canada, the findings have global relevance for harmonization applications. Harmonization models created for data captured from a specific season (Summer) were shown to outperform year-round models. This result was consistent for the harmonization performance (Table 3) and the classification performance (Table 6). Additionally, limiting pixels to those with relatively clear atmospheric conditions (AOT ≤ 0.3) can improve harmonization results; the *SR_ATMOS_OPACITY* band should be considered for further use in cross-sensor harmonization studies and to create high-quality image composites. We believe that thresholding AOT can allow for more consistent ETM+/OLI data, since it can minimize the interference of atmospheric features in the ETM+ SRFs. Using

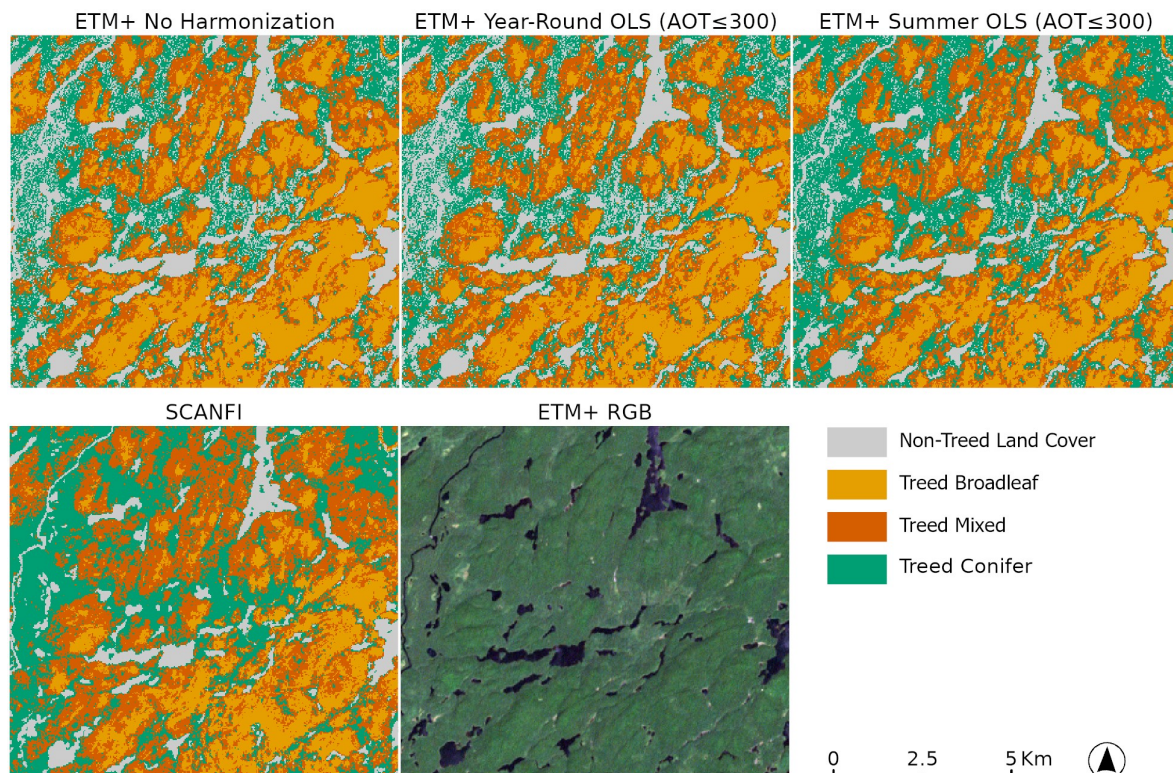


Figure 5. RF model predictions on different Landsat images.

seasonal data and limiting $AOT \leq 0.3$ provided the highest harmonization and classification performance in this study (Table 4 and Table 6). Table 7 contains the optimal harmonization models for Canadian vegetation studies that use summer Landsat ETM+ and OLI data with $AOT \leq 0.3$. Additionally, the pattern of RMSD across AOT bins in the summer dataset (Figure 4D) suggests that applying an AOT threshold of 0.5 could also yield relatively small differences between ETM+ and OLI band values.

Band	Summer OLS ($AOT \leq 0.3$)
Blue	$OLI = 0.7308ETM + 0.0036$
Green	$OLI = 0.8121ETM + 0.0076$
Red	$OLI = 0.8449ETM + 0.0036$
NIR	$OLI = 0.9697ETM + 0.0052$
SWIR1	$OLI = 0.9081ETM + 0.0052$
SWIR2	$OLI = 0.9162ETM + 0.0045$

Table 7. The optimal OLS harmonization models for the Canadian summer data.

The *SR_ATMOS_OPACITY* band on Landsat ETM+ and TM has limitations, since LEDAPS can fail to generate AOT values if there is no dark dense vegetation (DDV) available in the Landsat scene (Zhang et al., 2022). While trees are present in most of Canada, AOT thresholding for sensor harmonization is not possible in regions with limited or no DDV (e.g. Arctic or deserts). Harmonization models specific to these regions would be more suitable for such tasks and could be generated using the LEOHS tool (Richardson et al., 2025).

All Landsat images (including OLI) had accuracies < 0.75 for the SCANFI classification, which is partly attributable to the classes having very similar spectral reflectance. There are significant spectral overlaps between the Treed SCANFI classes, as illustrated by Figure 6. Nonetheless, this study illustrates that cross-sensor harmonization, using seasonally relevant pixels and filtering for AOT, can improve classification performance.

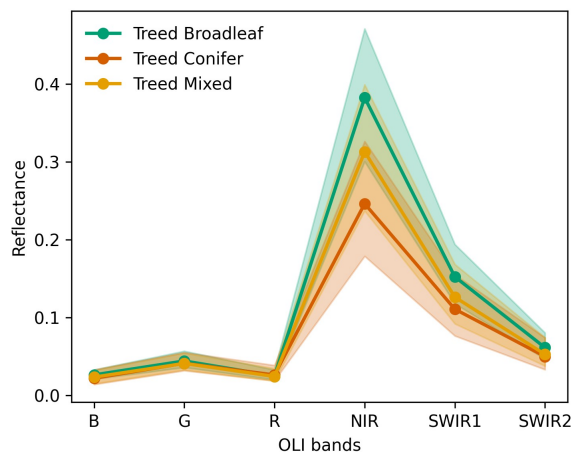


Figure 6. Mean and 5-95% ranges of OLI reflectance for the SCANFI classes.

Future work should further investigate the use of machine learning in sensor harmonization. While previous work has shown that RF models can cause artifacts when used for sensor harmonization, this could be caused by noisy high-AOT pixels being included in the training data (Scheffler et al., 2020). Machine learning models have the potential to benefit sensor harmonization, especially in the NIR band, since it often

appears to exhibit a non-linear relationship between ETM+, which could be attributed to the substantially different SRFs (Richardson et al., 2025).

5. Conclusion

Sensor harmonization is essential to producing consistent Landsat data for time series analysis. We assessed linear regression Landsat ETM+/OLI harmonization models that used data with different seasonality and AOT thresholds. Models trained on summer data and pixels with $AOT \leq 0.3$ outperformed other models in this study. Our study shows that harmonization models benefit from seasonally selected data, and these improvements translated into better classification performance. We suggest that future vegetation studies use pixels with $AOT \leq 0.3$ for change detection analysis, where applicable, and implement the harmonization functions published in this study. Although the results were derived from data over Canada, the findings have global relevance for harmonization applications.

Acknowledgements

The authors would like to acknowledge the SWEOL Lab at the University of Ottawa, the Canadian Centre for Mapping and Earth Observation, Koreen Millard, and Mickey Richardson.

References

- Guindon, L., Manka, F., Correia, D. L. P., Villemaire, P., Smiley, B., Bernier, P., Gauthier, S., Beaudoin, A., Boucher, J., & Boulanger, Y. (2024). A new approach for spatializing the Canadian National Forest Inventory (SCANFI) using Landsat dense time series. *Canadian Journal of Forest Research*, 54(7), 793–815. <https://doi.org/10.1139/cjfr-2023-0118>
- Lovitt, J., Richardson, G., Zhang, Y., & Richardson, E. (2023). Tree-CRowNN: A Network for Estimating Forest Stand Density from VHR Aerial Imagery. *Remote Sensing*, 15(22), 5307. <https://doi.org/10.3390/rs15225307>
- Olthof, I., & Fraser, R. H. (2024). Mapping surface water dynamics (1985–2021) in the Hudson Bay Lowlands, Canada using sub-pixel Landsat analysis. *Remote Sensing of Environment*, 300, 113895. <https://doi.org/10.1016/j.rse.2023.113895>
- Richardson, G., Knudby, A., Millard, K., & Chen, W. (2025). A tool for global and regional Landsat 7 and Landsat 8 cross-sensor harmonization. *Geocarto International*, 40(1), 2538108. <https://doi.org/10.1080/10106049.2025.2538108>
- Roy, D. P., Kovalsky, V., Zhang, H. K., Vermote, E. F., Yan, L., Kumar, S. S., & Egorov, A. (2016). Characterization of Landsat-7 to Landsat-8 reflective wavelength and normalized difference vegetation index continuity. *Remote Sensing of Environment*, 185, 57–70. <https://doi.org/10.1016/j.rse.2015.12.024>
- Scheffler, D., Frantz, D., & Segl, K. (2020). Spectral harmonization and red edge prediction of Landsat-8 to Sentinel-2 using land cover optimized multivariate regressors. *Remote Sensing of Environment*, 241, 111723. <https://doi.org/10.1016/j.rse.2020.111723>

Wulder, M. A., Roy, D. P., Radeloff, V. C., Loveland, T. R., Anderson, M. C., Johnson, D. M., Healey, S., Zhu, Z., Scambos, T. A., Pahlevan, N., Hansen, M., Gorelick, N., Crawford, C. J., Masek, J. G., Hermosilla, T., White, J. C., Belward, A. S., Schaaf, C., Woodcock, C. E., ... Cook, B. D. (2022). Fifty years of Landsat science and impacts. *Remote Sensing of Environment*, 280, 113195. <https://doi.org/10.1016/j.rse.2022.113195>

Zhang, Y., Woodcock, C. E., Arévalo, P., Olofsson, P., Tang, X., Stanimirova, R., Bullock, E., Tarrío, K. R., Zhu, Z., & Friedl, M. A. (2022). A Global Analysis of the Spatial and Temporal Variability of Usable Landsat Observations at the Pixel Scale. *Frontiers in Remote Sensing*, 3, 894618. <https://doi.org/10.3389/frsen.2022.894618>