**lsatTS: an R package for deriving vegetation greenness time series using Landsat satellite data**

Logan T. Berner, Jakob J. Assmann, Signe Normand and Scott J. Goetz

School of Informatics, Computing, and Cyber Systems, Northern Arizona University, USA

Department of Biology – Ecoinformatic and Biodiversity, Aarhus University, Denmark

# Abstract

Earth-observing satellites are crucial for assessing and monitoring global ecosystems. The Landsat satellites provide near global surface reflectance measurements since the early 1980s and are thus a corner stone of remotely-sensed ecological assessments. Landsat surface reflectance measurements are commonly used to derive spectral indices (e.g., NDVI) that can provide insight into annual to multi-decadal changes in ecosystem biophysical properties such as vegetation greenness. Nevertheless, multiple factors impede multi-decadal assessments of spectral indices using Landsat satellite data, including ease of data access and cleaning as well as challenges with cross-sensor calibration and irregular timing of cloud-free acquisitions. The R package *lsatTS* was developed to overcome these problems, it facilitates point sample-based time series analysis of spectral indices derived from Landsat surface reflectance measurements. The package includes functions that enable full data record extraction for point sample sites or study regions using the Google Earth Engine accessed from R. Moreover, the package includes functions for (1) rigorous data cleaning, (2) cross-sensor calibration with machine learning, (3) phenological modeling, and (4) other aspects of data analysis. For an example application, we show how *lsatTS* can be used to assess changes in vegetation greenness since the 1980s across a long-term monitoring area in the Arctic. Overall, this software provides a suite of functions to enable broader use of Landsat satellite data for the assessment and monitoring of vegetation greenness over the past four decades across local to global geographic extents.

# Background

## Ecological assessment and monitoring using the Landsat satellites

Satellite remote sensing is crucial for assessing and monitoring changes in Earth’s land surface over the last four decades (refs). The Landsat satellites are particularly value in this regard as they were designed for land surface monitoring at moderate spatial resolution (30 m). The first Landsat satellite (Landsat 1) was launched in 1972 as a partnership between NASA and the U.S. Geological Survey (USGS) and since that time a series of additional satellites have been launched, culminating in the recent launch of Landsat 9 in late 2021 (Wulder et al. 2019). The Landsat satellites carry multi-spectral sensors that have been used, for instance, for regional to global monitoring of forest cover (Hansen et al. 2013), surface water (Pekel et al. 2016), and wetlands (refs). These satellites observations have also been used to assess climate change impacts on Earth’s terrestrial ecosystems (refs), such as recent greening in the Arctic tundra biome (Berner et al. 2020). For a recent review of the Landsat program, science, and applications see Wulder et al. (2019).

## Impediments to long-term assessments using the Landsat satellites

* Data access and processing
  + Traditionally from USGS, but now made available through GEE
* Data cleaning
  + …It’s important but is hard
  + FMask
  + Residual water
* Cross sensor calibration
  + There are systematic differences in individual bands and spectral indices among Landsat 5’s Thematic Mapper (TM), Landsat 7’s Enhanced Thematic Mapper Plus (ETM+), and Landsat 8’s Operational Land Imager (OLI).
  + These differences can introduce spurious trends into time series generated from multiple sensors.
  + For instance, these biases can lead to spurious increases in NDVI (‘greening’) (Sulla-Menashe et al. 2017).
  + Existing approaches focus on linear corrections, but not all relationships are linear
* Irregulating timing of observations
  + Each Landsat satellite passes over a location about once every 16 days.
  + Clouds can obscure the land surface and lead to irregular acquisition surface reflectance measurements made under clear-sky conditions.
  + This makes it challenging, for instance, to assesses vegetation greenness at a desired phenological stage (e.g., maximum summer greenness).

## The lsatTS package

We developed the R package *lsatTS* to facilitate sample-based time series analysis of spectral indices derived from surface reflectance measured by the Landsat satellites. *lsatTS* grew out of recent research projects that assessed changes in vegetation greenness across the Arctic tundra and boreal forest biomes since the early 1980s using Landsat satellite data (Berner et al. 2020, Berner and Goetz In Review) and has been used in other research projects focused on specific aspects of Arctic and boreal ecology (Boyd et al. 2019, Verdonen et al. 2020, Boyd et al. 2021, Gaglioti et al. 2021, Mekonnen et al. 2021, Walker et al. 2021). *lsatTS* provides novel functions that facilitate Landsat data extraction, preparation, and analysis within the free, open-source, and widely-used R software environment (R Core Team 2020). The R software environment for statistical computing runs on multiple computing platforms (UNIX, Windows, MacOS) and provides state-of-the-art tools for data analysis visualization developed by a global user community (R Core Team 2020). Several R packages currently exist for accessing and processing Landsat data, including *landsat* (Goslee 2011), *landsat8* (dos Santos 2017), and *rLandsat* (ref). The *landsat* package provides functions for radiometric and topographic correction of Landsat scenes; *rLandsat* provides functions for searching and downloading Landsat 8 scenes from the USGS; and *landsat8* provides functions for computing top of atmosphere reflectance, radiance, and/or brightness temperature on Landsat scenes.

Nevertheless, there does not currently exist….

The new *lsatTS* package is unique in that it provides a coherent framework for sample-based time series analyses of spectral indices derived from surface reflectance measured by the Landsat satellite series. *lsatTS* includes functions for sample-based extraction of full data records from Landsat 5, 7, and 8 that is accomplished by querying the Landsat Collection 2 data set on GEE (Gorelick et al. 2017) using the *rgee* package in R (Aybar et al. 2020). Moreover, *lsatTS* includes functions that facilitate (1) data cleaning, (2) cross-sensor calibration with machine learning, (3) characterization of growing season conditions using phenological modeling, and (4) other aspects of vegetation greenness time series analysis (Table 1). Unlike wall-to-wall analyses, this sample-based framework is conducive to error propagation using Monte Carlo uncertainty analyses (Berner et al. 2020, Berner and Goetz In Review). Overall, this software provides a suite of functions to enable broader use of Landsat satellite data for assessment and monitoring of vegetation greenness over the past four decades in a sample-based framework suitable for local to global geographic extents.

Table 1. Function names and descriptions. These are listed in the order typically used.

|  |  |  |
| --- | --- | --- |
| **Step** | **Function** | **Description** |
| Data extraction | lsat\_get\_pixel\_centers | (*Optional*) Get point coordinates of all Landsat 8 pixel centers that fall within a polygon. |
|  | lsat\_export\_ts | Export full Landsat surface reflectance time series for a set of point coordinates using GEE accessed from R. |
| Data preparation | lsat\_general\_prep | Prepare data exported from GEE, including parsing satellite names and renaming and scaling bands. |
|  | lsat\_clean\_data | Filter out measurements based on presence of clouds, water, shadows, oblique view angles, and other criteria. |
|  | lsat\_summarize\_data\_avail | (*Optional*) Summarize data availability at each site, such as total number and years of observations. |
|  | lsat\_neighborhood\_mean | (*Optional*) For buffered sites, compute band-wise mean surface reflectance across grid cells within the buffer. |
|  | lsat\_calc\_spec\_index | Calculate a variety of widely used spectral indices, such as the Normalized Difference Vegetation Index (NDVI). |
|  | lsat\_calibrate\_rf | Cross-calibrate bands or spectral indices from Landsat 5/8 to match Landsat 7 using Random Forests. |
| Data analysis | lsat\_fit\_phenological\_curves | Characterize seasonal land surface phenology at each site by iteratively fitting flexible cubic splines. |
|  | lsat\_summarize\_growing\_seasons | Estimate various phenological metrics from fitted cubic splines, such as annual maximum vegetation greenness. |
|  | lsat\_evaluate\_phenological\_max | (*Optional*) Evaluate estimates of annual maximum vegetation greenness with measurement availability. |
|  | lsat\_calc\_trend | Calculate temporal trends using non-parametric Mann-Kendall trend tests and Theil-Sen slope indicators. |

# Package installation

The R package *lsatTS* is publicly available through GitHub. Users will need to have installed the R software environment on their computer. The *lsatTS* package is operating system agnostic and can be installed from within R using the *install\_github()* function from the *devtools* package:

devtools::install\_github("logan-berner/lsatTS")

To use the data extraction and preparation functions, users will need an account on GEE and to have installed and configured the *rgee* package to assess GEE from R. Please see the GEE (<https://earthengine.google.com/>) and *rgee* (<https://r-spatial.github.io/rgee/>) websites for details on signing up for an account and configuring *rgee*, respectively.

# Data extraction

*lsatTS* provides functions for sample-based extraction of full Landsat data records from GEE accessed using *rgee*…

* The main way of accessing the Landsat data using lsatTS is based on point sample locations.
  + For a given sample point all Landsat pixels that overlap with the point are extracted from the combined collection of all Landsat 5,7 and 8 Collection 2 Surface Reflectance scenes available on the GEE.
* Optional: For an area of interest, all landsat 8 pixel centers within the area can be determined with the ls\_get\_pixel\_centers() function. These pixel centre coordinates can then be passed on for the point based extraction.

# Data preparation

## Prepare data for analysis using lsat\_general\_prep()

After exporting Landsat data from Earth Engine, it is then necessary prepare the data for analysis. First, read the exported data into R using the *fread()* function from *data.table* and then use the *lsat\_general\_prep()* function to parse necessary information, rename columns, and scale band values. Note that all *lsatTS* functions depend on there being a column called “site” that uniquely identifies each location. If this column is not called “site” in your dataset, then make sure to modify your column name accordingly.

## Clean surface reflectance data using lsat\_clean\_data()

Most analyses use high-quality surface reflectance measurements that were acquired under clear-sky conditions. You can filter surface reflectance measurements using *lsat\_clean\_data()*. This function allows you to filter measurements based on pixel quality flags and scene criteria. The USGS provides pixel quality flags based on the CFMask algorithm (Zhu et al. 2015) and information on each scene (e.g., cloud cover). The default settings for *lsat\_clean\_data()* will filter out measurements flagged as snow or water, as well as measurements acquired at high solar zenith angle (>60°), those with high geolocation uncertainty (>15 m), or those acquired as part of scenes with extensive cloud cover (>80%). Addition water masking is provided based on maxim surface water extent () from the Landsat-based JRC Global Surface Water Dataset (Pekel et al. 2016).

*Optional: Compute neighborhood mean surface reflectance using lsat\_neighborhood\_mean()*

If each of your sites were buffered to include a neighborhood of Landsat pixels (e.g., 3 x 3 pixels), then *lsat\_neighborhood\_mean()* will compute the mean reflectance across this neighborhood of pixels for measurements at each period in time:

*Optional: Summarize data availability for each site using lsat\_summarize\_data\_avail()*

The function *lsat\_summarize\_data\_avail()* creates a summary table that provides information on the period and number of observations available for each site. It also generates a figure showing the cross-site aggregate number of observations across years:

*Calculate spectral indices using lsat\_calc\_spec\_index()*

Calculate common spectral indices using the function *lsat\_calc\_spec\_index()*. This function includes ~15 spectral indices, including the Normalized Difference Vegetation Index (NDVI), 2-band Enhanced Vegetation Index (EVI2), and others (Table 2). Note the function can only compute one spectral index at a time:

Table 2. Spectral indices that can be computed using the *lsat\_calc\_spec\_index()* function.

|  |  |  |  |
| --- | --- | --- | --- |
| **Name** | **Abbreviation** | **Formula** | **Citation** |
| Enhanced Vegetation Index | EVI |  | (Huete et al. 2002) |
| Enhanced Vegetation Index (2-band) | EVI2 |  | (Jiang et al. 2008) |
| Moisture Stress Index | MSI |  | Rock et al. 1986 |
| Near Infrared Vegetation Index | NIRv |  | (Badgley et al. 2017) |
| Normalized Burn Ratio | NBR |  | (Key and Benson 1999) |
| Normalized Difference Infrared Index | NDII |  | Hardisky et al. 1983 |
| Normalized Difference Moisture Index | NDMI |  | (Gao 1996) |
| Normalized Difference Vegetation Index (red) | NDVI |  | (Rouse et al. 1974) |
| Normalized Difference Vegetation Index (green) | gNDVI |  | Gitelson and Merzlyak 1998 |
| Normalized Difference Vegetation Index (kernel) | kNDVI | )2) | (Camps-Valls et al. 2021) |
| Normalized Difference Water Index | NDWI |  | (McFeeters 1996) |
| Plant Senescence Reflectance Index | PSRI |  | Merzlyak et al. 1999 |
| Soil Adjusted Vegetation Index | SAVI | 1.5 \* | (Huete 1988) |
| Soil Adjusted Total Vegetation Index | SATVI |  | Marsett et al. 2006 |
| Wide Dynamic Range Vegetation Index | WDRVI |  | (Gitelson 2004) |

## Cross-calibrate spectral reflectance or index across sensors using lsat\_calibrate\_rf()

There are systematic differences in spectral reflectance and indices among Landsat sensors and thus when combining data from multiple sensors it is important to further cross-calibrate the data. The function *lsat\_calibrate\_rf()* will calibrate individual bands or spectral indices from Landsat 5/8 to match Landsat 7. Landsat 7 is used as a benchmark because it temporally overlaps with the other two sensors. Cross-calibration can only be performed on one band or spectral index at a time and requires having data from 100s to preferably many 1,000s of sample sites. The approach involves determining the typical reflectance at a site during a portion of the growing season using Landsat 7 and Landsat 5/8 data that were collected the same years. A Random Forest model is then trained to predict Landsat 7 reflectance from Landsat 5/8 reflectance. If your data include both Landsat 5 and 8, then the function will train a Random Forest model for each sensor. By default, *lsat\_calibrate\_rf()* will add a new column with the cross-calibrated data ([band].xcal); however, the function will overwrite the existing column if you set the option overwrite.col = T. The function will also create an output directory that contains (1) trained Random Forest models, (2) a spreadsheet with model evaluation metrics, and (3) a multi-panel figure comparing sensors pre- and post-calibration. If you use the default setting that adds a new column with the cross-calibrated data, then you’ll either want to use those data in the subsequent functions (e.g., ndvi.xcal) or, once satisfied, manually overwrite the uncalibrated data to simplify subsequent column names:

# Data analysis

## Fit phenological curves to vegetation greenness time series using lsat\_fit\_phenological\_curves()

The function *lsat\_fit\_phenological\_curves()* characterizes seasonal land surface phenology at each sampling site using vegetation greenness (e.g., NDVI) time series from Landsat satellite observations. The function was constructed as a steppingstone to estimating annual maximum

vegetation greenness (e.g., NDVImax). The function iteratively fits cubic splines to seasonal vegetation greenness time series and returns information about the timing and magnitude of individual vegetation greenness observation relative to a multi-year seasonal phenology at

each site. The function was designed for spectral indices that are typically positive (e.g., NDVI). If you are working with a spectral index that is typically negative (e.g., NDWI) then multiply your index by -1 before running the *lsat\_fit\_phenological\_curves()* and *lsat\_summarize\_growing\_seasons()* functions and then back-transform afterwards

## Derived annual growing season metrics using lsat\_summarize\_growing\_seasons()

The function *lsat\_summarize\_growing\_seasons()* estimates several annual growing season metrics from vegetation greenness time series derived from Landsat satellite observations. The metrics include annual mean, median, and 90th percentile vegetation greenness of observations during each growing season, as well as phenologically-modeled estimates of annual maximum vegetation greenness and the seasonal timing (Day of Year) of maximum vegetation greenness. This function relies on output from *lsat\_fit\_phenological\_curves()*.

## Optional: Evaluate estimates of annual maximum vegetation greenness using lsat\_evaluate\_phenological\_max()

Estimates of annual maximum vegetation greenness are sensitive to the number of observations available from a growing season. The function *lsat\_evaluate\_phenological\_max()* is a tool for assessing how the number of annual Landsat observations impacts estimates of annual maximum vegetation greenness derived from raw observations and after phenological modeling. The algorithm computes a “true” annual maximum vegetation greenness using site x years with a user-specific number of observations and then compares these with estimates derived when using progressively smaller subsets of observations. This lets you determine the degree to which annual estimates of maximum vegetation greenness are impacted by the number of available observations.

## Compute trends in annual vegetation greenness using lsat\_calc\_trend()

Th function *lsat\_calc\_trend()* computes a temporal trend in annual time series of vegetation greenness for each sampling site over a user-specified time period. This is a wrapper for the *zyp.yuepilon()* function from the *zyp* package. This function will iteratively pre-whiten a time series (i.e., remove temporal autocorrelation) and then compute Mann-Kendall trend tests and Theil-Sen slope indicators:

# Example application

Here we provide an example focused on changes in annual maximum vegetation greenness across XX from 1984 to 2021. We use NDVI as a metric of vegetation greenness that broadly correlates with tundra productivity and aboveground biomass (Street et al. 2007, Raynolds et al. 2012, Berner et al. 2018, Berner et al. 2020)….

Figure X.

Conclusion

# Literature cited

Aybar, C., Q. Wu, L. Bautista, R. Yali, and A. Barja. 2020. rgee: An R package for interacting with Google Earth Engine. Journal of Open Source Software **5**:2272.

Badgley, G., C. B. Field, and J. A. Berry. 2017. Canopy near-infrared reflectance and terrestrial photosynthesis. Science Advances **3**:e1602244

Berner, L. T., and S. J. Goetz. In Review. Vegetation greenness trends consistent with a boreal forest biome shift.

Berner, L. T., P. Jantz, K. D. Tape, and S. J. Goetz. 2018. Tundra plant aboveground biomass and shrub dominance mapped across the North Slope of Alaska. Environmental Research Letters **13**:035002.

Berner, L. T., R. Massey, P. Jantz, B. C. Forbes, M. Macias-Fauria, I. H. Myers-Smith, T. Kumpula, G. Gauthier, L. Andreu-Hayles, B. Gaglioti, P. J. Burns, P. Zetterberg, R. D'Arrigo, and S. J. Goetz. 2020. Summer warming explains widespread but not uniform greening in the Arctic tundra biome. Nature communications **11**:4621.

Boyd, M. A., L. T. Berner, P. Doak, S. J. Goetz, B. M. Rogers, D. Wagner, X. J. Walker, and M. C. Mack. 2019. Impacts of climate and insect herbivory on productivity and physiology of trembling aspen (Populus tremuloides) in Alaskan boreal forests. Environmental Research Letters **14**:085010.

Boyd, M. A., L. T. Berner, A. C. Foster, S. J. Goetz, B. M. Rogers, X. J. Walker, and M. C. Mack. 2021. Historic declines in growth portend trembling aspen death during a contemporary leaf miner outbreak in Alaska. Ecosphere **12**:e03569.

Camps-Valls, G., M. Campos-Taberner, Á. Moreno-Martínez, S. Walther, G. Duveiller, A. Cescatti, M. D. Mahecha, J. Muñoz-Marí, F. J. García-Haro, and L. Guanter. 2021. A unified vegetation index for quantifying the terrestrial biosphere. Science Advances **7**:eabc7447.

dos Santos, A. 2017. landsat8: Landsat 8 Imagery Rescaled to Reflectance, Radiance and/or

Temperature. R package version 0.1-10. <https://CRAN.R-project.org/package=landsat8>.

Gaglioti, B., L. T. Berner, B. M. Jones, K. M. Orndahl, A. P. Williams, L. Andreu‐Hayles, R. D’Arrigo, S. J. Goetz, and D. H. Mann. 2021. Tussocks enduring or shrubs greening: Alternate responses to changing fire regimes in the Noatak River Valley, Alaska. Journal of Geophysical Research: Biogeosciences **126**:e2020JG006009.

Gao, B.-C. 1996. NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sensing of Environment **58**:257-266.

Gitelson, A. A. 2004. Wide dynamic range vegetation index for remote quantification of biophysical characteristics of vegetation. Journal of plant physiology **161**:165-173.

Gorelick, N., M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore. 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment **202**:18-27.

Goslee, S. 2011. Analyzing remote sensing data in R: The Landsat Package. The Journal of Statistial Software **43**.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, S. V. Stehman, S. J. Goetz, T. R. Loveland, A. Kommareddy, A. Egorov, L. Chini, C. O. Justice, and J. R. G. Townshend. 2013. High-Resolution Global Maps of 21st-Century Forest Cover Change. science **342**:850.

Huete, A., K. Didan, T. Miura, E. P. Rodriguez, X. Gao, and L. G. Ferreira. 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sensing of Environment **83**:195-213.

Huete, A. R. 1988. A soil-adjusted vegetation index (SAVI). Remote Sensing of Environment **25**:295-309.

Jiang, Z., A. R. Huete, K. Didan, and T. Miura. 2008. Development of a two-band enhanced vegetation index without a blue band. Remote Sensing of Environment **112**:3833-3845.

Key, C. H., and N. C. Benson. 1999. The Normalized Burn Ratio (NBR): A Landsat TM radiometric measure of burn severity. United States Geological Survey, Northern Rocky Mountain Science Center.(Bozeman, MT).

McFeeters, S. K. 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. International Journal of Remote Sensing **17**:1425-1432.

Mekonnen, Z. A., W. J. Riley, L. T. Berner, N. J. Bouskill, M. S. Torn, G. Iwahana, A. L. Breen, I. H. Myers-Smith, M. G. Criado, Y. Liu, E. S. Euskirchen, S. J. Goetz, M. C. Mack, and R. F. Grant. 2021. Arctic tundra shrubification: a review of mechanisms and impacts on ecosystem carbon balance. Environmental Research Letters **16**:053001.

Pekel, J.-F., A. Cottam, N. Gorelick, and A. S. Belward. 2016. High-resolution mapping of global surface water and its long-term changes. Nature **540**:418-422.

R Core Team. 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna.

Raynolds, M. K., D. A. Walker, H. E. Epstein, J. E. Pinzon, and C. J. Tucker. 2012. A new estimate of tundra-biome phytomass from trans-Arctic field data and AVHRR NDVI. Remote Sensing Letters **3**:403-411.

Rouse, J., R. Haas, J. Schell, and D. Deering. 1974. Monitoring vegetation systems in the Great Plains with ERTS. NASA special publication **351**:309-317.

Street, L., G. Shaver, M. Williams, and M. Van Wijk. 2007. What is the relationship between changes in canopy leaf area and changes in photosynthetic CO2 flux in arctic ecosystems? Journal of Ecology **95**:139-150.

Verdonen, M., L. T. Berner, B. C. Forbes, and T. Kumpula. 2020. Periglacial vegetation dynamics in Arctic Russia: decadal analysis of tundra regeneration on landslides with time series satellite imagery. Environmental Research Letters **15**:105020.

Walker, X. J., H. D. Alexander, L. T. Berner, M. A. Boyd, M. M. Loranty, S. M. Natali, and M. C. Mack. 2021. Positive response of tree productivity to warming is reversed by increased tree density at the Arctic tundra-taiga ecotone. Canadian Journal of Forest Research.

Wulder, M. A., T. R. Loveland, D. P. Roy, C. J. Crawford, J. G. Masek, C. E. Woodcock, R. G. Allen, M. C. Anderson, A. S. Belward, W. B. Cohen, J. Dwyer, A. Erb, F. Gao, P. Griffiths, D. Helder, T. Hermosilla, J. D. Hipple, P. Hostert, M. J. Hughes, J. Huntington, D. M. Johnson, R. Kennedy, A. Kilic, Z. Li, L. Lymburner, J. McCorkel, N. Pahlevan, T. A. Scambos, C. Schaaf, J. R. Schott, Y. Sheng, J. Storey, E. Vermote, J. Vogelmann, J. C. White, R. H. Wynne, and Z. Zhu. 2019. Current status of Landsat program, science, and applications. Remote Sensing of Environment **225**:127-147.

Zhu, Z., S. Wang, and C. E. Woodcock. 2015. Improvement and expansion of the Fmask algorithm: cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. Remote Sensing of Environment **159**:269-277.