**AUTHOR RESPONSE TO EDITOR AND REVIEWER FEEDBACK**

**LandsatTS: an R package to facilitate retrieval, cleaning, cross-calibration, and phenological modeling of Landsat time-series data**

**Subject Editor (Dr. Michael Borregaard) Comments**

The package is purely geographical, and thus somewhat on the edge of what Ecography publishes, but given the widespread use of Landsat data among ecologists it still seems useful. It does mean, however, that the quite complicated installation procedure becomes an issue. Very few of the readers of Ecography are likely to have rgee and google earth engine installed and have an active google earth account. They are not trivial to install, and the reviewer had multiple problems getting the install to work. These procedures should be spelt out much more clearly.

Author response: We thank the Subject Editor for not only considering our manuscript for publication with Ecography, but also providing helpful feedback on both the manuscript and software. Following the feedback obtained, we substantially revised and improved both the manuscript and software. This software helps make the indispensable Landsat satellite record more accessible to ecologists, land managers, and others who don’t specialize in satellite remote sensing.

We acknowledge that utilizing the *LandsatTS* package for R requires an account with Google Earth Engine (GEE) and installation of GEE and *rgee* but believe these are not too high of barriers to use. Landsat data are publicly available, but all data providers require an account, whether it’s GEE or the USGS. Obtaining a GEE account is free, fast, and provides access to a wide variety of data beyond Landsat. While most users find it straightforward to install GEE and rgee, there can be cases when installation proves to be challenging. Therefore, we point users to GEE and rgee documentation that can help guide them through the installation process. Overall, the *LandsatTS* software can help improve our understand of ecological dynamics around the world by enabling a broader community to utilize the unique Landsat data record.

In extension of this, it is hard to see whether your lsatTS install actually works, as the running example in the paper requires 2 days to download the data! I would definitely like the example in the paper to be one that is immediately runnable, so that a reader can read the paper with a working R session next to them, and also use the example to troubleshoot that their install of the package actually works correctly. Maybe one approach is to split the data acquisition and analysis example into two bits, make

the data acquisition smaller and supply the data needed for the example within the package.

Author response: We agree the original example application was a poor choice that took too long to run and so now provide a different example application that more quickly and effectively demonstrates software functionality. The new example application focuses on changes in vegetation greenness from 2000 to 2022 across a random sample of locations in the Noatak National Preserve in northern Alaska, USA. As suggested, we split the data extraction and analysis portions of the example. To demonstrate data export from GEE, the example application randomly selects and exports data for three random locations within the preserve, which takes about six minutes to complete. To demonstrated data analysis, the example application now relies on Landsat data included in the package (n = 100 random locations). The revised example application more efficiently demonstrates functionality of *LandsatTS*.

Much of the description in the paper is a point-by-point description of the functions, making this read less like an article and more like a printed manual page. It might make sense to leave the function descriptions in there, but I suggest having them a little later (maybe after the example) and focusing more on a description of the philosophy of the package, use case and workflow design. The readme of the github repository is useful and contains some good illustrations and could probably be reproduced in the supplementary materials.

Author response: We significantly restructured the manuscript to emphasize the philosophy, workflow, and use case of the package. This included fully restructuring the introduction, modifying the use case, and moving the function descriptions to end of the manuscript. This restructuring much improves the manuscript.

In general, Ecography prefers packages installable directly from CRAN. Are there strong technical reasons for not doing this?

Author response: While having written the code to a high standard and provided thorough documentation, we currently are not able to muster the further time and resources necessary to get this R package onto the CRAN. We did not originally plan to develop this package but rather did so opportunistically once it became apparent there was broader community interest in utilizing these tools. Developing this package has already taken time and resources that were considerably beyond the scope of our funded research projects. Besides resource limitations, there currently are a huge and growing number of R packages, many of which are only available through GitHub. GitHub makes it possible to readily implement software changes and updates, as well as easily install the package from within R, while the update and release process in CRAN is substantially more cumbersome.

In general, the github repo and code is well organized and documented, with a good set of unit tests. As a small comment maybe activating continuous integration on repo pushes would secure long-term consistency of the code base.

Author response: This is an excellent suggestion! For now, we have been running the tests locally before pushes to the main branch as setting up the rgee-GEE access on the GitHub runners provided a large hurdle given our limited knowledge of GitHub actions and how to securely handle the secrets that are required. However, we are keen to engage with this in the future and will aim to implement CI using GitHub actions in the long run – even if just for those functions that don’t required GEE access.

Secondly, I do have another concern about the package that I'd like you to address. It seems that the workflow is fairly fixed (some functions are marked as "optional"), and I partly get the impression that this package, especially the analytical part, is intended for a very particular pre-determined workflow, leading to a defined set of analyses/results. It would be good to see discussed how much the package lends itself to a broader set of use cases and frameworks, and how much creativity they allow the researchers using the package.

Author response: We primarily developed this package for generating and analyzing multidecadal time series of vegetation greenness using Landsat data and believe there is considerable interest among the ecological community in conducting similar analyses. That said, the package’s data extraction and processing tools also enable users to undertake other analyses that rely on carefully processed Landsat time series data for sample locations. Furthermore, we designed the software so there is a lot of flexibility within functions, such as how splines are fit when characterizing seasonal phenology, or choices between different cross-sensor calibration functions. We updated Figure 1 to better demonstrate different options for utilizing this software, and now also discuss several examples of other possible uses. For instance, part of the introduction now reads:

These tools have also been used to assess high-latitude vegetation responses to insect outbreaks (Boyd et al. 2019, Boyd et al. 2021), wildfires (Gaglioti et al. 2021), and permafrost degradation (Verdonen et al. 2020), as well as for syntheses focused on high-latitude disturbance regimes (Foster et al. 2022) and Arctic shrubification (Mekonnen et al. 2021). Among other applications, these tools could further be used to complement field-based ecosystems monitoring in protected areas, evaluate ecosystem impacts of extreme weather events (e.g., droughts), and improve local to global mapping efforts by enabling users to develop regression models for cross-sensor calibration. In summary, *LandsatTS* enables ecologists and other researchers to extract and process Landsat time series that can then be used to analyze vegetation phenology or for other user-defined applications.

In all these comments amount to quite a bit of restructuring, but I feel confident the authors should be able to meet the comments and submit a version that can eventually be accepted, so it's somewhere between a major and a minor revision.

Author response: We appreciate your feedback and have revised and restructured the manuscript as recommended. The revised manuscript and software is greatly improved.

As two small comments that I just mention for consideration:

1. would it be easier if the package was named LsatTS? This seems more consistent with the acronym

Author response: To improve clarity and discoverability, we changed the name of the package from *lsatTS* to *LandsatTS*.

2. It seems unnecessary to have all functions preceeded by `lsat\_`. Any user interested in such explicitness could always use `lsatTS::` instead.

Author response: We agree it is not entirely necessary for function names to be preceded by “lsat\_”, but this naming convention conveniently groups package functions while adding little to the length of function names. Further, we already have a userbase that is using the current naming convention. Therefore, we have opted to maintain the current naming convention to allow for continuity of the already existing code.

**Review 1 Comments**

The authors present a novel R package (lsatTS) that offers a range of functions for constructing, cleaning, and analyzing Landsat time series for phenology purposes. The integration with Google Earth Engine and related 3rd-party libraries (i.e., rgee) is a key feature. The authors have written a clear, well-organized overview of the package components and background rationale. There are only a few areas where I thought they could strengthen their description of their work, as described below, followed by comments regarding grammatical errors and minor edits.

Author response: We thank the reviewer for their positive, constructive feedback and have accordingly revised the manuscript and software.

The authors describe several existing R packages for processing Landsat data. However, they do not similarly review existing R packages for phenological analysis, such as “phenology”, “phenor”, and “phenofit”. Explaining how lsatTS complements those packages or provides additional functionality would highlight the novelty and utility of their effort.

Author response: We appreciate the reviewer’s suggestion and examined the packages they mentioned. The *phenology* package focuses on animal phenological count data, while *phenor* provides tools for evaluating plant phenology for several datasets. Since the manuscript is already rather long, we chose not to review those two packages; however, we now highlight the new *phenofit* package given its particular relevance. Part of the introduction now reads:

… To address this issue, *LandsatTS* includes tools to estimate annual maximum vegetation greenness based on site-specific phenological modeling that iteratively fits cubic splines to vegetation greenness time series. Users interested in other aspects of vegetation phenology (e.g., timing of spring onset or fall senescence) could extract and process Landsat data using *LandsatTS*, but then capitalize on tools provided by other R packages, such as the new *phenofit* package that provides state-of-the-art tools for fitting phenological models (Kong et al. 2022). More broadly, while *LandsatTS* provides tools focused on generating high-quality vegetation greenness times series, it also enables users to undertake other analyses that rely on cleaned and cross-calibrated Landsat data.

I think the random forest cross-calibration option is intriguing but a little puzzling. What is the benefit of the site-specific process over applying fixed band/index transformations (e.g., those in Roy et al. 2016)? If users do not have enough samples to train random forest models, the authors already provide the option of pre-processed data, which seems like a similar approach. Can the authors make some statement about the advisability of performing the RF step?

Author response: Further cross-sensor calibration is crucial for time series analyses, therefore *LandsatTS* includes tools that enable users to cross-calibrate spectral bands and indices from Landsat 5 and 8 with Landsat 7. Prior approaches for cross-sensor calibration focused on linear corrections for individual spectral bands and select spectral indices (e.g., NDVI) using regional data (e.g., continental USA) from Landsat Collection 1 (e.g., Ju and Masek, 2016;Roy et al., 2016). These published cross-sensor calibration models do not account for potential non-linearities, may not be suitable for other regions, and may not be appropriate for the newer Landsat Collection 2 dataset, thus new tools are needed to facilitate cross-sensor calibration. During revision, we added further rational to the Background section and also developed a new function called *lsat\_calibrate\_poly()* that enables users to cross-calibrate individual spectral bands and indices using polynomial regression models instead of random forest models. The new function yields results that are similar to the original function, but generates regression models that are more readily re-used and shared. Part of the Background section now reads:

Landsat time series analyses that use measurements from multiple sensors are hindered by systematic biases in spectral bands and indices among the Landsat 5 Thematic Mapper (TM), Landsat 7 Enhanced Thematic Mapper Plus (ETM+), and Landsat 8 Operational Land Imager (OLI) sensors (Ju and Masek 2016, Roy et al. 2016, Berner et al. 2020, Berner and Goetz 2022). If unaccounted for, these biases can introduce pronounced artificial trends into combined time series, such as spurious increases over time in spectral indices of vegetation greenness including the widely used Normalized Difference Vegetation Index (NDVI) (Sulla-Menashe et al. 2017). Prior approaches for cross-sensor calibration focused on linear corrections for individual spectral bands and select spectral indices (e.g., NDVI) using regional data (e.g., continental USA) from Landsat Collection 1 (e.g., Ju and Masek 2016, Roy et al. 2016). While valuable, these published cross-sensor calibration models do not account for potential non-linearities, may not be suitable for other regions, and may not be appropriate for the newer Landsat Collection 2 dataset. Therefore, *LandsatTS* includes functions to cross-calibrate spectral bands and indices among Landsat 5, 7, and 8 using either random forest machine learning or polynomial regression models. These models are fit using the user’s dataset. However, if the user’s dataset is too small to fit these models, then, if appropriate, the user can choose to fit models using pre-processed and staged Landsat data that were sampled from across the Arctic tundra and boreal forest biomes. Flexible implementation of cross-sensor calibration in the *LandsatTS* workflow enables the user to generate high quality time-series that are free from sensor-specific biases that can otherwise induce spurious trends.

Currently, the user is able to set a threshold deviation from the cubic spline curve for removing points (last\_fit\_phenological\_curves()). It would be helpful to have the option to specify thresholds that are distinguished by whether the point is above or below the curve, since typically a lower value is more suspect than a higher one.

Author response: This is a good idea, so we modified the function so users can now set separate thresholds to remove points that are above or below the fitted cubic spline.

It might be prudent to present readers with a small, manageable process instead of one that takes ~2 days to export the files.

Author response: We fully agree that a smaller, more manageable example application is needed. Therefore, we developed a new example application that focuses on changes vegetation greenness from 2000 to 2022 across a network of random sample points in the Noatak National Preserve in northern Alaska, USA. The data extraction portion of this example now takes about six minutes, while the data analysis portion of the example relies on data that are now provided with the package. This example now effectively demonstrates the package’s functionality in a far more reasonable amount of time.

I recommend commenting out (or at least drawing attention to!) the rm(list=ls()) command in the code, and setting setwd to a generic folder (see <https://www.tidyverse.org/blog/2017/12/workflow-vs-script/>)

Author response: We appreciate the reviewer highlighting the workflow vs script distinction and have revised the scripts accordingly.

**Minor edits and typos:**

L78: Remove hyphen between “widely-used”. Not necessary in compound adjectives when the first word is an adverb that ends in -ly. Other instances throughout paper.

Author response: Fixed here and elsewhere throughout the manuscript.

L98: “…provides integrated, sample-based framework…” Insert “an” before “integrated”.

Author response: Done

L132: This is a trivial request but it would be useful for the packages to be listed alphabetically.

Author response: Done

L210: Italicize “last\_general\_prep()”

Author response: Done

L216: “Each…were”. Change to “was”.

Author response: Done

L244: The Landsat sensors are listed correctly on page 2 (L5 TM, L7 ETM+, L8 OLI), but here L7 and L8 are

incorrectly referred to as ETM and ETM+, respectively. See also L248, Table 3.

Author response: Fixed

L278, 280: “moving-windows” incorrectly hyphenated here.

Author response: Fixed

L298: Change “(4) and” to “and (4)”

Author response: Done

L311: Missing period at end of sentence.

Author response: Done

L340: “The function extracts site x years with at least…” I’m unclear whether “site x years” is something I’m misinterpreting (x years of data at a site?) or if it’s a typo. Either way, the meaning could be clearer.

Author response: That is the correct interpretation, but to improve clarity, we modified the sentence to now read, “For each site, the function extracts years with at least the user-specified number of growing season observations…”

L354: “remove” -> “removes”

Author response: Done

L383: “was” -> “were” (Landsat data)

Author response: Done

L402: “…observations in the between…” Remove “in the”.

Author response: Done

L488: Italicize “last”

Author response: Done

L544: “…where there were temporally overlaps measurements from pairs Landsat satellites”. Couple of typos in there.

Author response: Corrected to, “… where there were temporally overlapping measurements from pairs of Landsat satellites.”

L566: “estimate” -> “estimates”

Author response: Done

L573: “…prior TO the turn…”

Author response: Done

L610: “dried” -> “drier”

Author response: Done

L611: “with defoliation” -> “to defoliation”

Author response: Done

Figure 1: Is there any significance to the fact that only some functions are italicized?

Author response: No, that was an accident. For the figure, we remove italics from all text.

Figure 2: (b) appears to show the location of Disko Island rather than the study area per se. A different color scheme for the positive NDVImax values might provide more contrast to the green background of the image.

Author response: This figure no longer appears in the manuscript because we now use a different example application.

Figure 4: Remove decimal from right-hand column “count” legend

Author response: Done.

Figure 5:

- Mention the time frame of the example in the caption.

Author response: Done. We now mention observations were made between 1985 and 2022.

- It’s unclear whether the Observation (pts) and Curve legends are supposed to match temporally; the color ramps are identical, but the years are clearly at unequal spacing. Is the reader meant to visually match the color of the points with a similarly colored curve? Otherwise, I’m not sure I understand the point of the color-coding.

Author response: We use these figures for a quick visual assessment of whether there are erroneous curves or observations, as well for visually highlighting long-term changes in phenology. Color coding help illustrate how individual curves are fit to observations. Curves are fit to observations from multiple years, but they are colored based on the focal year. We now also make these points in the figure caption.

- I’m a little confused about some of the curve fitting. In samples like pixel \_1838 (top row, middle column) many of the lower-NDVI pre-200 DOY points appear to be ignored, while curves are seemingly well-matched to the higher-NDVI points. Is there a weighting function in the curve-fitting routine that promotes points with higher NDVI?

Author response: The figure originally showed all observations, regardless of whether they were filtered out during the curve fitting routine. The curve fitting routine does not inherently promote observations with higher values; however, anomalous observations tend to have low values that get filtered out. We updated the figure so now it does not show observations that were filtered out as anomalies.

- There seems to be quite a lot of low-NDVI points, which makes me wonder about the noise reduction function in the library.

Author response: The curve fitting process involves iteratively removing points and refitting curves until the remaining points are all within a user-defined threshold of the final curve. However, the figure originally showed all points, regardless of whether they were filtered out during the curve fitting routine. As noted above, we updated the figure, so it now only shows observations that were used for curve fitting, while excluding those that were filtered. Nevertheless, we added another noise reduction step to the routine for initial outlier removal. The function now includes an initial step that, for each site, fits a curve using data pooled across all years and then filters out observations that differ from the curve by more than 100%.

- Overall, I find this figure hard to follow. Personally, I think I would prefer to see a sequence of years with individual phenology curves rather than the kind of consolidated representation shown here. This opinion is not a request to revise the approach! But it would be helpful for the authors to explain their justification for the 11-year aggregation of data.

Author response: We concede its challenging to display the seasonal distribution of observations, interannual changes, and curve fits across multiple decades for multiple sites. While imperfect, this figure does convey a lot of information. As we now further note in the function description, it is generally necessary to pool data across multiple years because there are typically few observations within an individual year. Part of the function description now reads:

… Often there are too few observations from an individual year to fit a reliable phenological curve, therefore the function enables users to pool observations over multiple years when fitting each cure. The default is a 7-year moving-window centered on the focal year, but the width of the moving window can be made shorter or longer if there are many or few observations in the data record…

Figure 7: Mention the number of sample locations. Please specify what “grouped by the concomitant temporal trends” means. Were the annual data for all sample locations with a particular trend (browning/greening/no-change) averaged?

Author response: After further consideration, we decided to remove this functionality.

**Reference**

Ju, J., and Masek, J.G. (2016). The vegetation greenness trend in Canada and US Alaska from 1984–2012 Landsat data. *Remote Sensing of Environment* 176**,** 1-16.

Roy, D.P., Kovalskyy, V., Zhang, H.K., Vermote, E.F., Yan, L., Kumar, S.S., and 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.