



Kierunek studiów: Geoinformacja

Nr albumu: 455828

Tomasz Matuszek

**Measuring impact of addition of Landsat 8
thermal band on supervised land cover
classification results**

*Ocena wpływu zastosowania kanału termalnego
Landsat na wyniki nadzorowanej klasyfikacji pokrycia
terenu*

Praca inżynierska napisana
w Instytucie Geoekologii i Geoinformacji
pod kierunkiem
dr. hab. Jakuba Nowosada

Poznań, 2022

Abstract

Abstrakt

Streszczenie powinno przedstawiać skrótowo główny problem pracy i jego rozwiązanie. Możliwa struktura streszczenia to: (1) 1-3 zdania wstępu do problemu (czym się zajmujemy, dlaczego jest to ważne, jakie są problemy/luki do wypełnienia), (2) 1 zdanie opisujące cel pracy, (3) 1-3 zdania przedstawiające użyte materiały (dane) i metody (techniki, narzędzia), (4) 1-3 zdania obrazujące główne wyniki pracy, (5) 1-2 zdania podsumowujące; możliwe jest też określenie dalszych kroków/planów.

Słowa kluczowe: (4-6 słów /zwrotów opisujących treść pracy, które nie wystąpiły w tytule)

Abstract

The abstract must be consistent with the above text.

Keywords: (as stated before)

Spis treści

| | |
|--|-----------|
| Abstract | 3 |
| 1 Introduction | 5 |
| 2 Source data | 7 |
| 2.1 Satellite imagery | 8 |
| 2.2 Land cover data | 9 |
| 3 Methods | 15 |
| 3.1 Machine learning | 15 |
| 3.2 R language environment | 16 |
| 4 Result of the model - land cover map | 17 |
| 5 Assessing model quality | 19 |
| 6 Evaluation of thermal band's impact on prediction results | 21 |
| 7 Conclusion | 23 |

Rozdział 1

Introduction

- applications and relevance of land cover maps
- machine learning and supervised classification of satellite images as a tool for creating land cover maps
- pointing out that thermal band is often omitted in land cover classification models, exact impact of thermal factor isn't fully clear
- goal of the thesis is to create land cover map of Poznań metropolitan area and measure the impact of thermal band on the model results

Wprowadzenie powinno mieć charakter opisu od ogółu do szczegółu (np. trzy-pięć paragrafów). Pierwszy paragraf powinien być najbardziej ogólny, a kolejne powinny przybliżać czytelnika do problemu. Przedostatni paragraf powinien określić jaki jest problem (są problemy), który praca ma rozwiązać i dlaczego jest to (są one) ważne.

Wprowadzenie powinno być zakończone stwierdzeniem celu pracy. Dodatkowo tutaj może znaleźć się również krótki opis co zostało zrealizowane w pracy.

Pisząc ten rozdział proszę pomyśleć o osobach, które zupełnie nie znają opisywanej tematyki. Należy tutaj krok po kroku wyjaśnić podstawowe koncepcje, istotność problemu,

wyniki poprzednich podobnych badań, itd. Ten rozdział obejmuje tylko kwestie, które już zostały wykonane przez inne osoby - nowe wyniki mają swoje miejsce w rozdziale **?@sec-wyniki**.

Każda kwestia opisana w tym rozdziale powinna być cytowana. Dodatkowo cytowania odbywa się poprzez uzupełnienie pliku `thesis.bib` zapisem w formacie BibTeX, a następnie dodanie nazwy referencji poprzedzonej znakiem `@`. Przykładowo, zacytowanie książki *Geocomputation with R* odbywa się poprzez (Lovelace et al., 2019).

W przypadku, gdy cytowanie zostało poprawnie wpisane oraz istnieje w pliku `thesis.bib` to bibliografia powinna się automatycznie wygenerować na końcu pracy.

W przypadku, gdy praca dyplomowa opisuje konkretny obszar to można po tym rozdziale stworzyć kolejny rozdział opisujący "obszar badań".

Ten i kolejne rozdziały mogą mieć także podrozdziały. Tworzenie podrozdziałów polega na stworzeniu nowej linii rozpoczynającej się od znaków `##` a następnie tytułu podrozdziału. Dodatkowo w postaci `{#sec - }` można dodać skrót nazwy rozdziału/podrozdziału umożliwiający odnoszenie się do niego używając operatora `[-@sec]`.

Rozdział 2

Source data

Satellite imagery used in our model was downloaded from Landsat ARD dataset, provided by GLAD laboratory at the University of Maryland (Potapov et al., [2020](#)). Training points were obtained from LUCAS dataset created by Eurostat (d'Andrimont et al., [2020](#)). This data was pre-processed and then used to train the model and validate its performance. Middle-West Poland was chosen as a training area for which satellite imagery and land cover data were downloaded (Figure 2.1).

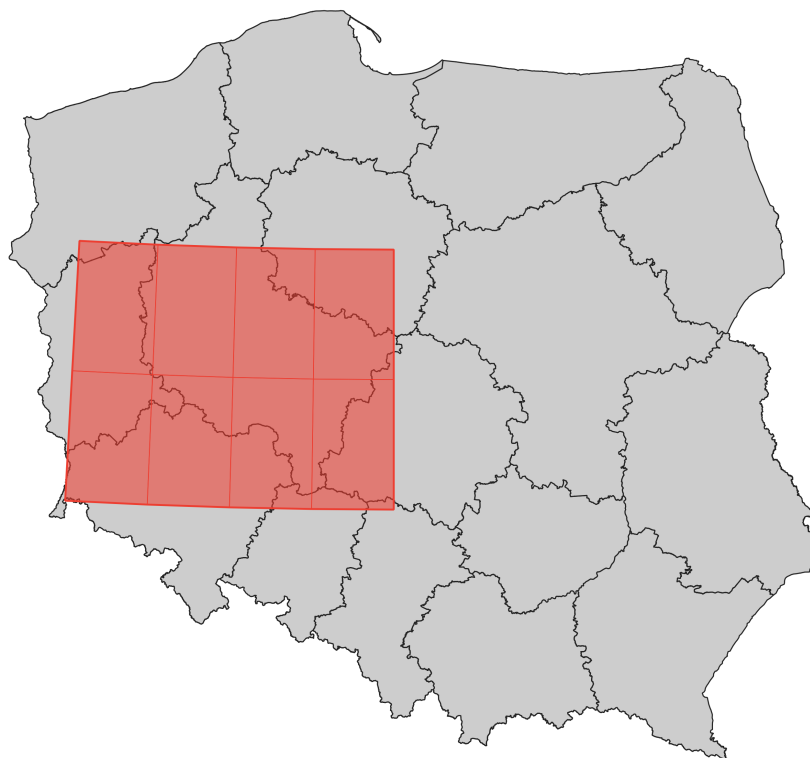


Figure 2.1: *Training area*

2.1 Satellite imagery

Satellite imagery from GLAD Landsat ARD is available in 16-day interval composites and is divided into 1 x 1 geographic degrees tiles. Processing of original Landsat images performed by GLAD team included conversion of spectral bands to top-of-atmosphere (TOA) reflectance, conversion of thermal band to brightness temperature (BT) in Kelvins, scaling the values of all bands as well as adding quality flag for every pixel (Potapov et al., [2020](#)).

Satellite images for eight 1 x 1 degree tiles (Figure 2.1) were downloaded using GLAD Tools v1.1 and PERL programming language. These images come from 10th interval of the year 2018, so downloaded mosaics consist of images created between 24.05.2018 and 8.06.2018. All downloaded images were merged and reprojected from WGS84 coordinate reference system (EPSG:4326) to UTM zone 33N (EPSG:32633). Every band was also resampled to 30 meters resolution. In addition, four spectral indices were derived: Normalized Difference Vegetation Index (NDVI), Modified Normalized Difference

Table 2.1: *Formulas of spectral indices dervied from Landsat data*

| | a | b |
|---|---|---|
| 1 | a | |
| 2 | b | |
| 3 | c | |
| 4 | d | |
| 5 | e | |

Water Index (MNDWI), Normalized Difference Moisture Index (NDMI) and Modified Bare soil Index (MBI). Formulas used to calculate these indices can be found in below table (Table 2.1).

- preprocessing of satellite imagery performed by GLAD team from University of Maryland
- process of downloading data using GLAD ARD software
- GLAD ARD dataset structure

2.2 Land cover data

Regarding land cover training set, data collected during LUCAS survey was chosen. It seems to be the most accurate and comprehensive dataset containing information about land use and land cover (Pflugmacher et al., 2019) due to the fact, that every point was either photo-interpreted by human or assessed during *in-situ* visit.

Survey consists of two phases. First phase is based on grid of points with 2km spacing covering whole territory of the European Union (which is above 1 million points). Each point of the grid is visually interpreted using ortho-photos or satellite images, and classified into one of seven land-cover classes. These are: arable land, permanent crops, grassland, wooded areas/shrub land, bare land, artificial land and water. In the second phase a subsample of grid points is selected and then visited by Eurostat surveyors. They classify each point according to full LUCAS land cover and land use classification. The survey takes place in the spring and summer in order to observe chosen places in high vegetation season.

Surveyor not only assign a land cover and land use classes to a point, but they also add auxillary information such as plant species present at the site, percentage of land coverage for chosen class, height of the trees and their maturity as well as information about water management and irrigation. If there are more than one land cover/land use types at the point, observer can also assign secondary class for every LUCAS point. TUTAJ BĘDZIE REFERENCJA DO LUCAS COPERNICUS REPORT GDY DOWIEM SIĘ, JAK GO CYTOWAĆ

Majority of the training points used to train our classification model were points from the second phase of LUCAS survey, also called LUCAS Micro Data. We downloaded a total of 4153 points for our study area. Pre-processing step included omitting records with missing data, excluding linear artificial land cover classes and excluding points that were surveyed more than 500 meters from their theoretical location. In the next step, detailed land cover classes were aggregated into eight main groups of land cover types. Then, we filtered some of the classes according to the percentage of land coverage or percentage of impervious surface coverage.

For the least frequent classes in the LUCAS Micro Data dataset - bare land, artificial land and water bodies - we also added points classified during the first phase of LUCAS survey. This step was necessary to ensure that every land cover class is represented by enough number of points. At the end of the pre-processing, we were left with 3778 training points.

Later in the analysis, after extracting values from Landsat ARD raster, LUCAS points were also filtered using quality flag provided. Only points with clear-sky quality flag were taken into account during the process of model training. We also excluded water bodies points in which NDWI was lower than 0. These two conditions excluded over 400 points in total.

- what is LUCAS programme and how its datasets are created
- LUCAS Grid and LUCAS Micro Data

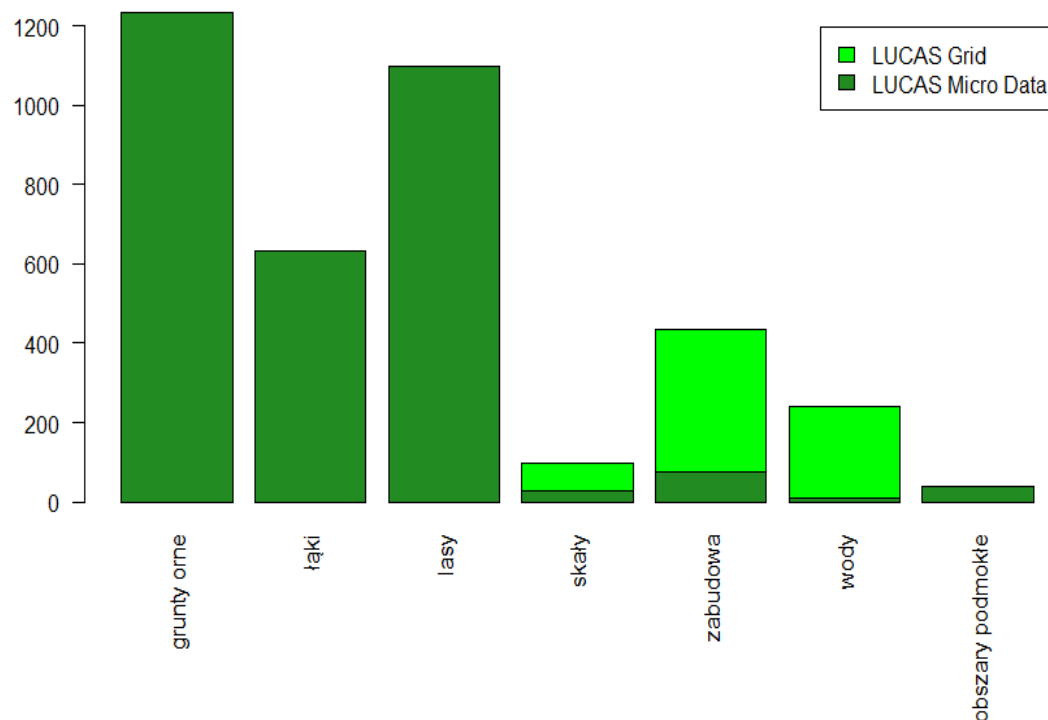


Figure 2.2: *Distribution of land cover classes after pre-processing*

- structure of LUCAS points dataset, description of relevant columns in addition to the land cover classification
- processing of data in order to choose reliable training points; reclassification to chosen set of land cover categories:
 - excluding railways, runways, roads and temporary grasslands from LUCAS Micro Data
 - reclassifying LUCAS Micro Data to seven main classes
 - filtering of training points based on auxiliary information (% of impervious surface or according land cover type)
 - adding LUCAS Grid data to three less frequent classes (with smaller weight)
 - choose only points where quality flag of Landsat ARD image was “clear sky”
 - choose only these points from water category where NDWI was above 0
- example of training points spatial distribution

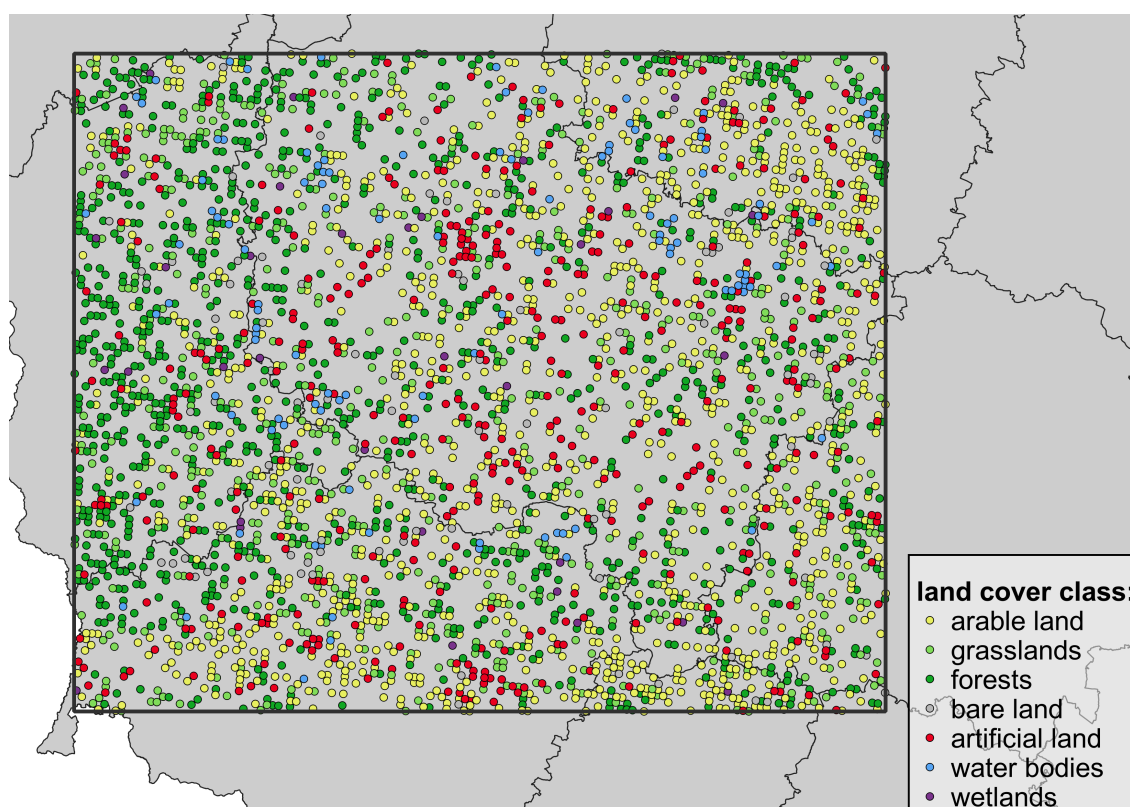


Figure 2.3: *Spatial distribution of LUCAS training points after pre-processing*

- table with land cover categories

Celem tego rozdziału jest przedstawienie użytych w pracy danych. Należy wyjaśnić jakie dane zostały użyte, jakiego są one rodzaju, dla jakiego okresu zostały pobrane/stworzone, co one zawierają, etc.

W tym rozdziale warto dodać ryciny i tabele przedstawiające użyte dane.

Zwróć uwagę, że poniższe bloki kodu mają parametr `#| echo: false;` oznacza to, że będą one niewidoczne w wynikowym pliku PDF. Każdy z bloków kodu musi mieć unikalną nazwę; w przypadku rycin powinna się ona rozpoczynać od prefiksu `fig-`. Dodanie podpisu pod rycinę odbywa się używając parametru `#| fig-cap:.` Następnie do tej ryciny można się odnieść używając operatora `[-@]`.

Podobnie wygląda odnoszenie się do plików graficznych. Tutaj wewnątrz bloku kodu należy użyć funkcji `knitr::include_graphics()` (Rycina). Dodatkowo możliwa

jest zmiana rozmiary obrazka używając parametrów takich jak `#| out-width: i #| out-height:`. Odnoszenie się do tabel odbywa się poprzez operator `[-@]` wraz z prefiksem `tbl-`. Natomiast tworzenie podpisu nad tabelą ma miejsce używając parametru `#| tbl-cap:`. Dodatkowo możliwe jest użycie pakietu **kableExtra** (Zhu, [2021](#)) do określenia szerokości kolumn (Tabela).

Rozdział 3

Methods

3.1 Machine learning

- what is machine learning and what are its applications
- classification vs regression algorithms
- supervised and unsupervised classification

3.1.1 Random forest algorithm

- what is a decision tree
- how random forest algorithm works

3.1.2 Model quality assessment

- idea of resampling
- measures and indices of classification model quality

3.1.3 Parameter tuning

- what is tuning of model's parameters
- idea of nested resampling

3.2 R language environment

Short description of R and RStudio environment. List of used libraries and packages.

Rozdział **Metody** zawiera opis użytych metod (np. statystycznych czy geostatystycznych) oraz technologii (np. pakiety R). Opis każdej z metod czy technologii powinien być zwarty i zawierać tylko najważniejsze informacje z punktu widzenia pracy dyplomowej.

Każda użyta metoda i technologia powinna być zacytowana. W przypadku pakietów R, wystarczy wypełnić poniższy blok kodu (zwróć uwagę, że ten blok kodu ma parametr `echo: false`; oznacza to, że będzie on niewidoczny w wynikowym pliku PDF).

... a następnie zacytować pakiet używając znaku @, po którym podać nazwę pakietu rozpoczynającą się od prefiksu R-. Przykładowe cytowanie języka R bez nawiasu to R Core Team (2021), a pakietu **kableExtra** w nawiasie to (Zhu, 2021). Więcej przykładów cytowania można znaleźć na stronie https://rmarkdown.rstudio.com/authoring_bibliographies_and_citations.html# citations.

W przypadkach, gdy cytowanie istnieje, ale nie jest pakietem R to należy dodać je do pliku `thesis.bib` i użyć powyższej składni ze znakiem @. W ostateczności, gdy dana technologia nie posiada cytowania, należy podać jej adres internetowy.

Rozdział 4

Result of the model - land cover map

- land cover map of Poznań metropolitan area
- probability map of model results

Część **Wyniki** może składać się z jednego lub więcej rozdziałów. Każdy z tych rozdziałów powinien mieć tytuł adekwatny do swojej treści.

Rozdziały wynikowe powinny korzystać z wiedzy opisanej w poprzednich rozdziałach (Rozdziały **?@sec-lit**, **?@sec-dane**, **?@sec-metody**). W przypadku prac analitycznych, ich treść powinna przedstawiać kolejne etapy eksploracji i analizy danych. W przypadku prac technicznych, treść tych rozdziałów powinna opisywać stworzone narzędzia, a następnie pokazywać ich zastosowanie/a.

W przypadku prac technicznych warto pokazywać fragmenty napisanego rozwiązania lub jego wywołania używając bloków kodu.

```
moja_funkcja = function(x){  
  cat(x, "rządzi!")  
}  
moja_funkcja("Autor tej pracy")
```

Autor tej pracy rządzi!

Rozdział 5

Assessing model quality

Table with quality indices:

- overall accuracy (OA)
 - classification error (CE)
 - producer's and user's accuracy (PA, UA)
 - Kappa coefficient
-

Część **Wyniki** może składać się z jednego lub więcej rozdziałów. Każdy z tych rozdziałów powinien mieć tytuł adekwatny do swojej treści.

Rozdziały wynikowe powinny korzystać z wiedzy opisanej w poprzednich rozdziałach (Rozdziały **?@sec-lit**, **?@sec-dane**, **?@sec-metody**). W przypadku prac analitycznych, ich treść powinna przedstawiać kolejne etapy eksploracji i analizy danych. W przypadku prac technicznych, treść tych rozdziałów powinna opisywać stworzone narzędzia, a następnie pokazywać ich zastosowanie/a.

W przypadku prac technicznych warto pokazywać fragmenty napisanego rozwiązania lub jego wywołania używając bloków kodu.

```
moja_funkcja = function(x){  
  cat(x, "rządzi!")  
}  
moja_funkcja("Autor tej pracy")
```

Autor tej pracy rządzi!

Rozdział 6

Evaluation of thermal band's impact on prediction results

- mean temperature for every predicted land cover class
- variable importance plots, variable profiles
- thermal band importance map (two methods: raster aggregation and interpolation of importance in LUCAS points)
- mean importance of thermal band on each land cover class
- difference raster map between prediction with and without thermal band included, transition matrix

Część **Wyniki** może składać się z jednego lub więcej rozdziałów. Każdy z tych rozdziałów powinien mieć tytuł adekwatny do swojej treści.

Rozdziały wynikowe powinny korzystać z wiedzy opisanej w poprzednich rozdziałach (Rozdziały **?@sec-lit**, **?@sec-dane**, **?@sec-metody**). W przypadku prac analitycznych, ich treść powinna przedstawiać kolejne etapy eksploracji i analizy danych. W przypadku

prac technicznych, treść tych rozdziałów powinna opisywać stworzone narzędzia, a następnie pokazywać ich zastosowanie/a.

W przypadku prac technicznych warto pokazywać fragmenty napisanego rozwiązania lub jego wywołania używając bloków kodu.

```
moja_funkcja = function(x){  
  cat(x, "rządzi!")  
}  
moja_funkcja("Autor tej pracy")
```

Autor tej pracy rządzi!

Rozdział 7

Conclusion

- land cover map of Poznań metropolitan area was created, impact of thermal band on classification results was measured
- despite thermal band having low overall impact on model results, there is a strong spatial auto-correlation for its importance
- land surface temperature was especially significant for land cover classification of urban areas, it helped in identify built-up areas
- it may mean that thermal band will become increasingly important in studies on urban sprawl and suburbanisation
- better land cover maps will help in better management of metropolitan areas growth and quantifying impact of urbanisation on natural environment more precisely

Podsumowanie pracy jest w pewnym sensie znacznie rozbudowanym abstraktem. Należy wyliczyć i opisać osiągnięcia uzyskane w pracy dyplomowej. Tutaj jednak (w przeciwieństwie do np. rozdziału **?@sec-wprowadzenie**) należy przechodzić od szczegółu do ogółu - co zostało stworzone/określone, jak zostało to zrobione, jakie ma to konsekwencje, itd.

Ten rozdział powinien też zawierać opis kwestii, których nie udało się rozwiązać w pracy dyplomowej (i dlaczego się nie udało) oraz pomysły na przyszłe ulepszenie uzyskanych wyników lub dalsze badania.

Bibliografia

- d'Andrimont, R, M Yordanov, L Martinez-Sanchez, B Eiselt, A Palmieri, P Dominici, J Gallego, HI Reuter, C Joebges, G Lemoine, and M van der Velde (2020). Harmonised LUCAS in-situ land cover and use database for field surveys from 2006 to 2018 in the European Union. *Scientific Data* 7(1), 352.
- Lovelace, R, J Nowosad, and J Muenchow (2019). *Geocomputation with R*. CRC Press.
- Pflugmacher, D, A Rabe, M Peters, and P Hostert (2019). Mapping pan-European land cover using Landsat spectral-temporal metrics and the European LUCAS survey. *Remote Sensing of Environment* 221, 583–595.
- Potapov, P, MC Hansen, I Kommareddy, A Kommareddy, S Turubanova, A Pickens, B Adusei, A Tyukavina, and Q Ying (29, 2020). Landsat Analysis Ready Data for Global Land Cover and Land Cover Change Mapping. *Remote Sensing* 12(3), 426.
- R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. <https://www.R-project.org/>.
- Zhu, H (2021). *kableExtra: Construct Complex Table with kable and Pipe Syntax*. R package version 1.3.4. <https://CRAN.R-project.org/package=kableExtra>.