Characterizing Land Cover Change to Promote Sustainable Urbanization in Saskatoon, Saskatchewan: an Approach using Machine Learning

Lucas Lau

April 23, 2023

## Abstract

Saskatchewan is recognized for their consistent supply of agricultural foods and ingredients. Given their productive land and available resources, measures must be taken to protect and promote sustainable urbanization and sustainable use of their ecosystems to combat deforestation. This study goes beyond simply running a land cover classification, but rather assess the performance of machine learning using the random forest algorithm. The study area chosen is Saskatoon, Saskatchewan because they had the highest decrease in forest cover in Canada. The satellite imagery used were from Landsat-8 Level 2 Collection 2 Tier 1, primarily from the summer months to achieve the clearest images with the least cloud and snow cover. By manually collecting training points and polygons where appropriate, an average of 66 points were collected for each class: Forest, Grassland, Urban, Cropland, Water, and Soil. Results showed that vegetative areas decreased from 401.95 m2 to 220.18 m2 and urban areas increased from 245.05 m2 to 417.80 m2. To determine the performance of the machine learning model, a combination of resubstitution matrix, validation error matrix, training overall accuracy, and validation overall accuracy were used. These tests revealed that the model was overfitting with about a 99% accuracy. This study concludes with recommendations for how the analysis could have been improved if similar analysis were to be conducted in the future.

## Introduction

Urbanization is strongly associated with an increase in individual wealth, as land owners and developers choose to invest in and build up less developed cities for economic gain (Resnik, 2010). While this provides living spaces for the growing population, it does not necessarily serve the public good, as it negatively impacts the environment, such as deforestation, high water consumption, and displacement of wildlife. To address these issues, urban planning must prioritize and promote smart growth (Resnik, 2010). Urban development changes the functions of ecosystems, which can lead to increased air pollution and environmental hazards that pose risks to human health. Not to mention, through the transformation of natural landscapes, animals are impacted by the loss of habitat, food sources, and toxic substances (National Geographic, 2021). To evaluate the extent that urbanization impacts land features, it is important to categorize the land features over a period of time to compare long-term changes. Therefore, classifying land covers is crucial in assessing land use and ensuring the sustainability of terrestrial ecosystems.

One of the Sustainable Development Goals (SDGs) is focused on Life on Land: “To protect, restore and promote sustainable use of terrestrial ecosystems, sustainably manage forests, combat desertification, and halt and reverse land degradation and halt biodiversity loss” (The Global Goals, 2023). To achieve these goals, it is important to map and monitor land cover as it is essential for sustainable development planning and rational utilization of land resources (Jin et al., 2018). This process involves identifying land covers, such as forests, grassland, soil, cropland, urban, and water. By understanding how land is being used and how they are changing over time, information is gained about land use patterns and areas that are susceptible to risks from urbanization. With such information, it becomes possible to monitor sustainable forest management, deforestation, agriculture planning, and urban growth (Jin et al., 2018). For example, classifying forest features over a period of time can provide a clear timeline of deforestation and discourage further destruction to forests. This information can also support and encourage smart land use planning to avoid urban sprawl and mitigate environmental damage.

Remote sensing is commonly used in land cover mapping to differentiate and classify distinct surface areas. Remote sensing has advantages of being fast, macroscopic, and providing synchronous monitoring. It is able to have a consistent spatial context globally for accurate monitoring and reporting (Government of Canada, 2021). It allows one to acquire information about the Earth’s surface from a distance using various sensors, providing a wide range of different information that can be combined to create detailed models of the landscape of interest. Satellite imagery is highly effective and a primary source of data for mapping land cover (Gallego, 2004), and it can provide up-to-date information on land cover changes. However, the choice of algorithm for land cover classification is important because it determines how the data is analyzed and interpreted. Different algorithms have different strengths and limitations, so the decision should be based on what land covers are being mapped and the objectives of the study. The algorithm's efficiency and accuracy must be considered by how they are influenced by image resolution and atmospheric conditions during imaging (Jin et al., 2018). Moreover, the algorithm's ability to handle issues, such as data errors, complexity of area of interest, and the lack of training data should be considered (Jin et al., 2018). Although various land cover types can be monitored differently to promote sustainability, this study will focus on tracking the loss of vegetative land areas to urbanization over a period of time. Traditional classification algorithms often lack effectiveness and accuracy, and have, and are gradually being replaced by machine learning algorithms, such as random forest (RF). This study will be investigating that exact algorithm, RF, through machine learning to automate land cover classification in a region of Saskatchewan.

Machine learning has become popular in remote sensing, particularly in the field of land-use classification (Abdi, 2020). High classification accuracies have been demonstrated using advanced machine learning algorithms and it is likely to continue to be an important tool in the future of land cover classification (Abdi, 2020). Machine learning greatly reduces the amount of labour required to analyze large datasets and minimize human error. Despite that advantage, the quality of machine learning outputs relies on the quality of the training data, so low-resolution data will most likely have poor results. However, high-resolution satellite imagery is becoming more available so machine learning will become an increasingly valuable method for efficient and accurate land cover classification.

The random forest algorithm is a machine learning technique that is popular due to its interpretability and reliability in handling large and complex data and making accurate predictions (Zhang and Yang, 2020). RF works by constructing multiple decision trees, where subsets of features are randomly selected at training time and are aggregated. It then outputs a classification based on the mean of the individual trees. By using different samples of the training data, RF can reduce the outcomes of overfitting and improve generalization (Gao et al., 2018). RF is also relatively stable with noise and handles insufficient training data well compared to a single decision tree (Zhang and Yang, 2020).

Through the use of machine learning, supervised classification algorithm, and satellite imagery, this study aims to address the following research questions:

1. How should random forest be implemented in machine learning to calculate the loss of vegetative areas from 2013 to 2020 in Saskatoon, Saskatchewan?
2. What land covers should be included in the classification to effectively determine the loss of vegetation due to urbanization?
3. How does the spatial resolution of satellite imagery data affect the performance of land cover classification?
4. How does the performance of random forest classification through machine learning compare to traditional rule-based classification methods?

In view of such information, this study aims to:

1. Efficiently classify various land covers with minimal misclassification through machine learning,
2. Evaluate the loss of vegetation from 2013 to 2020 to propose smart land use planning in the area of study,
3. Reduce overfitting and increase robustness with random forest.

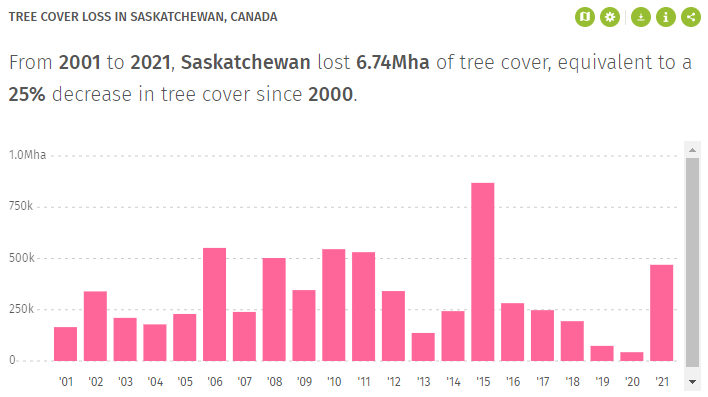
## Methodology

### Research Area

The research area will be the city of Saskatoon, Saskatchewan (Figure 1). From 2001 to 2019, Saskatchewan lost 62700 km2 of its tree land cover, equivalent to 23% of its total tree cover in 2001 (Hansen et al, 2019). Saskatchewan experienced an increasing tree cover loss from 2013 to 2015, and a decreasing tree cover loss from 2015 to 2020 (Figure 2). To identify this trend, we chose to focus on Saskatoon, the largest city in Saskatchewan in terms of both land area and population (Statistics Canada, 2021). In 2021, Saskatoon had a population of 266,141, a land area of 226.56 km2, and a population density of 1,174.7/km2 (Statistics Canada, 2021). Saskatoon’s population had a 10.6% increase from 2011 to 2016, and a 7.7% increase from 2016 to 2021 (Statistics Canada, 2021).



*Figure 1: Location of Saskatoon in Saskatchewan, Canada (Statistics Canada, 2021)*



*Figure 2: Tree Cover Loss in Saskatchewan, Canada (Hansen et al, 2021)*

### Dataset

To identify trends in tree cover change in Saskatoon, we will use satellite images taken over the city and compare vegetation area year over year from 2013 to 2020. Supervised machine learning will be used to identify vegetation and other land cover classification in satellite images.

Selection of satellite image source was based on spatial resolution and available time periods. It is necessary to use a medium to high spatial resolution that is suitable for machine learning to accurately identify multiple land classifications (Lee, Jung, 2019). The available datasets from the satellite image source must also include imagery before 2015 to identify significant tree cover changes. Sentinel-2B imagery was initially selected for its high resolution. However, its dataset starts at 2017 which does not provide a time period long enough to see significant land cover change (European Space Agency, 2013). Landsat 8 was selected instead for its acceptable medium resolution and its history of dataset which starts from 2013 (Pahlevan et al, 2017). To allow machine learning to accurately identify various land covers using pixel values, red, green, blue, and NDVI bands are used for this study.

To maximize the ability of machine learning, satellite images that are selected must have minimal interferences with the pixel data (Lee, Jung, 2019). To remove images that have significant cloud cover, only images that have a cloud cover of less than 3% are used for analysis. It is also important to use imagery from the same season when comparing images year over year in machine learning (Lee, Jung, 2019). This keeps pixel values similar for each land cover classification in each year, and allows the machine learning model to accurately predict land cover classes using training dataset from other images (Lee, Jung, 2019). To maximize the visibility of tree cover and remove snow cover, all images used for analysis are collected between the months June to July. However, there are certain years where Landsat 8 imagery over Saskatoon is not available during these months. If this is the case images from May, September, or October are used for analysis instead. To cover the full extent of the study area, it is necessary to use multiple satellite images for each year. To analyze multiple images at once, the mosaic function in Google Earth Engine Javascript is used to combine a set of images into one. The combined image is then classified in machine learning. All satellite images for this study are acquired using Google Earth Engine’s LANDSAT/LC08/C01/T1 image collection.

### Machine Learning

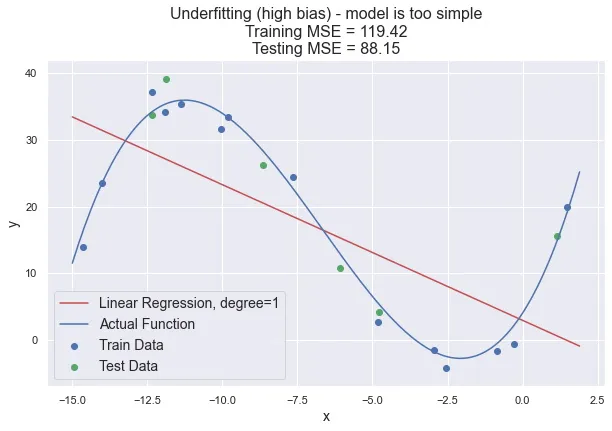
Once all satellite images are collected for each year, training data is to be collected for supervised machine learning using Google Earth Engine. During this step it is important to gather training data from each satellite image, and a similar quantity of training data from each land cover (Amoakoh et al, 2021). This reduces biases in the machine learning model and helps increase flexibility in the model to decrease overfitting (Amoakoh et al, 2021). The majority of training data is collected using points, however, it may sometimes be suitable to use polygons when there are large areas that have the same land cover (Amoakoh et al, 2021). For this study, approximately 66 data points are gathered for each land cover for each year.

A machine learning model is then built and trained using the training dataset gathered. Many algorithms of machine learning are suitable for remote sensing, such as SVM, random forest, and CNN. The most suitable algorithm or combination of algorithms depends on the quantity and complexity of data as well as the separability of our data (Lee, Jung, 2019). Support Vector Machine is a capable machine learning algorithm that performs well on small datasets that have high separability (Mountrakis et al, 2011). Our region of study is small, however pixel values in land cover classes are not highly separable. For example, soil and urban areas have similar NDVI values that are close to the value -1, while grassland and forest land areas have similar NDVI values close to 1. It is therefore not suitable to use SVM for this study. Convolutional Neural Network is another complex machine learning model that is commonly used in remote sensing (Kattenborn et al, 2021). However, CNN is only effective if a very large amount of data is collected (Kattenborn et al, 2021). Since our area of study is small, CNN is not suitable for this study. Random Forest was selected for this study since it is effective in smaller datasets and very capable in classifications (Amoakoh et al, 2021).

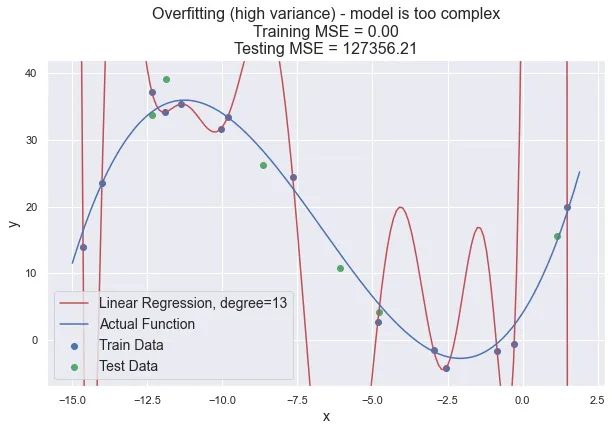
The random forest model is built using the smileRandomForest function in Google Earth Engine Javascript. Selecting the number of trees that are used in random forest depends on computing power. Generally, increasing the number of trees will increase the accuracy of the model at the cost of slower learning (Amoakoh et al, 2021). For this study, 100 trees were selected to balance accuracy with Google Earth Engine computing power. Once the machine learning model is built and trained, the model is used to classify the combined satellite images. The result is a layer of pixels where each pixel is classified as one of the four land covers. We can use this result to calculate the land area of each land cover using the pixelArea Javascript function. Calculating the land cover area for each year allows us to identify land cover changes and trends year over year from 2013 to 2020.

### Accuracy Assessment

It is highly unlikely that machine learning models are given the optimal parameters initially (Pettinger, 2982). It is therefore important to analyze accuracy metrics after results are outputted to determine how the model can be improved. In classification, resubstitution matrix, validation error matrix, training overall accuracy, and validation overall accuracy are used to determine the performance of the machine learning model (Pettinger, 2982). The most common issues in random forest classifications is model underfitting and overfitting (Pettinger, 2982). If both training error and validation error is high, then the model is underfitting (Barr, 2020). Underfitting occurs when the machine learning model is too simple for the dataset (Figure 3). This is also called high bias, which means the model is accurate in predictions but the initial assumption of the dataset is incorrect (Barr, 2020). If the training error is low but validation error is high, then the model is overfitting (Ying, 2019). Overfitting occurs when the machine learning model is too complex for the dataset (Figure 4). This is also called high variance, which means the model is inaccurate in predictions but the initial assumption of the dataset is correct (Ying, 2019). Changing the input data insignificantly in a high variance model will change the output significantly (Ying, 2019).



*Figure 3: Underfitting Example (Patil, 2019)*



*Figure 4: Overfitting Example (Patil, 2019)*

If the model is underfitting, then the machine learning model will need to be more complex. This can be achieved by adding more features in the dataset, such as adding additional bands from Landsat 8; adding more parameters to the model; and increasing the training time such as increasing the number of trees in random forest (Barr, 2020). Reducing noise in the dataset can also reduce underfitting, this can be achieved by further decreasing the percentage of cloud cover in satellite images (Barr, 2020).

If the model is overfitting, then the machine learning model will need to be simplified. This can be achieved by using K-fold cross validation and regularization techniques such as Lasso and Ridge (Ying, 2019). Overfitting can also be reduced by increasing the number of training data or using multiple machine learning models at once by adopting ensemble techniques (Ying, 2019).

## Results

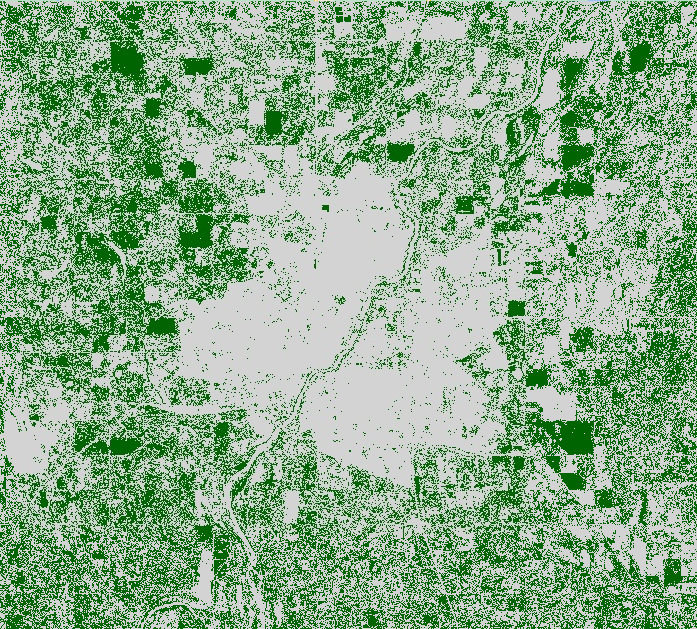
Land cover change is an important indicator when analyzing the environmental well-being and sustainability of urban environments. As urbanization continues to grow rapidly all over the world, it is essential to understand the trends and patterns of land cover change - specifically in Saskatoon, Saskatchewan - in order to successfully promote sustainable urban development and meet the Sustainable Development Goal: Life on Land. Over the last few years, the city of Saskatoon has undergone significant land cover change with its population steadily growing and urbanization expanding. This also indicates that there have been changes to land use as well. The land cover change in Saskatoon was characterized from 2013 to 2020, using the Random Forest machine learning algorithm to identify and classify the following land cover types: Water, Grassland, Urban, Forest, Cropland, and Soil. For the purpose of this study though, some land cover types with similar properties were combined, such as forest, cropland, and grassland. This is because it was very difficult to distinguish between these three types due to the fact that the NDVI values were quite similar between forest and other vegetative land cover types. All vegetation have similar NDVI values that are close to 1. For instance, a random pixel point in the forest land cover type had an NDVI value of 0.4685 and a random pixel point in the grassland land cover type had an NDVI value of 0.4041. By analyzing the land cover changes from 2013 to 2020, it would provide valuable insights into the trends and patterns of urbanization in Saskatoon, Saskatchewan, revealing both the opportunities and challenges that come with it.

In 2013, the land cover in Saskatoon, Saskatchewan was characterized with 401.95 m2 of vegetation area, 245.05 m2 of urban area, 327.94 m2 of water area, and 2.46 m2 of soil area. Based on these findings, the majority of the study region was characterized as vegetative area, which made up approximately 41.12% of the total area of the study region, while urban and water areas accounted for approximately 25.07% and 33.55% of the total area respectively. The soil land cover type however, only accounted for a very small part of the total area at approximately 0.25%. Since 2013 is the first year of the study, the analysis of land cover change in 2013 shows that the majority of Saskatoon was composed of vegetative land which implies that it had a relatively high level of natural vegetation cover at this time. This has several implications regarding the city’s environmental well-being, such as reducing humidity levels and temperatures by providing more shade, improving air quality by absorbing pollutants and releasing oxygen, and lastly, promoting more biodiversity by providing more vegetation for wildlife. According to figure 5, it is evident that the majority of Saskatoon is classified as vegetation in 2013.



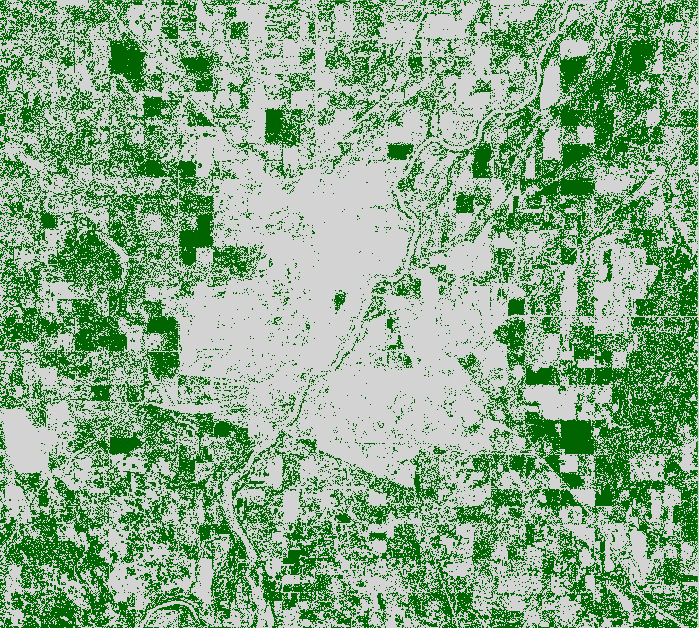
*Figure 5. Map showing vegetation and urban land cover types in 2013 in Saskatoon, Saskatchewan*

In 2014, the land cover in Saskatoon, Saskatchewan was characterized with 348.52 m2 of vegetation area, 316.28 m2 of urban area, 299.40 m2 of water area, and 1.85 m2 of soil area. Based on these findings, the majority of the study region was still characterized as vegetative area, which made up approximately 36.08% of the total area of the study region, while urban and water areas accounted for approximately 32.74% and 30.99% of the total area respectively. The soil land cover type however, still only accounted for a very small part of the total area at approximately 0.19%. The analysis of land cover change from 2013 to 2014 shows that the vegetation area decreased from 401.95 m2 to 348.52 m2, while the urban area increased from 245.05 m2 to 316.28 m2. The water area also slightly decreased from 327.94 m2 to 299.40 m2. The soil area remained relatively the same, still only taking up less than 1% of the total area. According to figure 6, it is evident that some of the vegetation has been converted to urban areas since 2013.



*Figure 6. Map showing vegetation and urban land cover types in 2014 in Saskatoon, Saskatchewan*

In 2015, the land cover in Saskatoon, Saskatchewan was characterized with 383.05 m2 of vegetation area, 283.17 m2 of urban area, 298.59 m2 of water area, and 1.23 m2 of soil area. Based on these findings, the majority of the study region was still characterized as vegetative area, which made up approximately 39.65% of the total area of the study region, while urban and water areas accounted for approximately 29.31% and 30.91% of the total area respectively. The soil land cover type however, still only accounted for a very small part of the total area at approximately 0.13%. The analysis of land cover change from 2014 to 2015 shows that the vegetation area increased slightly from 348.52 m2 to 383.05 m2 in 2015. The urban area actually decreased from 316.28 m2 to 283.17 m2 and the water area remained relatively the same, with a slight decrease from 299.40 m2 to 298.59 m2 in 2015. The soil area also remained relatively the same, still only taking up less than 1% of the total area. According to figure 7, it is evident that vegetation is still continuing to decrease in 2015 compared to figure 6.



*Figure 7. Map showing vegetation and urban land cover types in 2015 in Saskatoon, Saskatchewan*

In 2016, the land cover in Saskatoon, Saskatchewan was characterized with 265.50 m2 of vegetation area, 343.39 m2 of urban area, 238.76 m2 of water area, and 1.85 m2 of soil area. Based on these findings, Saskatoon underwent major changes toward urbanization as the majority of the study region was no longer vegetative area. Instead, the majority of the study region was characterized as urban area, which made up approximately 40.42% of the total area of the study region, while vegetation and water areas accounted for approximately 31.25% and 28.11% of the total area respectively. The soil land cover type however, still only accounted for a very small part of the total area at approximately 0.22%. The analysis of land cover change from 2015 to 2016 shows the vegetation area significantly decreased from 383.05 m2 to 265.50 m2 in 2016. The water area also decreased from 298.59 m2 to 238.76 m2. The urban area actually significantly increased from 283.17 m2 to 343.39 m2, becoming the majority land cover area of the study region after 2015. The soil area also remained relatively the same, still only taking up less than 1% of the total area. According to figure 8, it is evident that Saskatoon underwent major changes in regards to urban development as a lot of the vegetation that existed in the previous years are no longer there.



*Figure 8. Map showing vegetation and urban land cover types in 2016 in Saskatoon, Saskatchewan*

In 2017, the land cover in Saskatoon, Saskatchewan was characterized with 220.20 m2 of vegetation area, 442.05 m2 of urban area, 227.72 m2 of water area, and 1.85 m2 of soil area. Based on these findings, the majority of the study region was still characterized as urban area, which made up approximately 49.57% of the total area of the study region, while vegetation and water areas accounted for approximately 24.69% and 25.53% of the total area respectively. The soil land cover type however, still only accounted for a very small part of the total area at approximately 0.21%. The analysis of land cover change from 2016 to 2017 shows that the vegetation area continued to decrease from 265.50 m2 to 220.20 m2 in 2017. The water area also slightly decreased from 238.76 m2 to 227.72 m2. The urban area actually continued to significantly increase from 343.39 m2 to 442.05 m2, increasing by another 100 m2 compared to the previous increase from 2015 to 2016. The soil area also remained relatively the same, still only taking up less than 1% of the total area. According to figure 9, it is evident that there has not been much change in vegetation compared to the previous year; however, it is still decreasing.



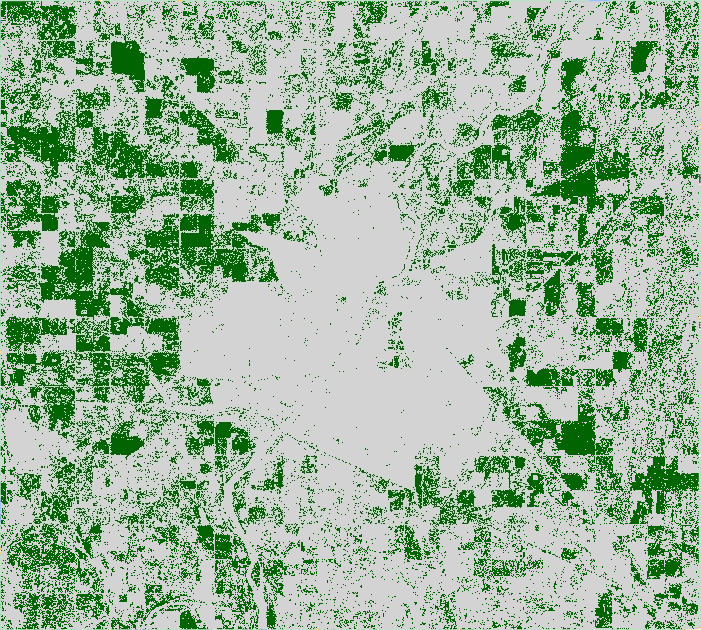
*Figure 9. Map showing vegetation and urban land cover types in 2017 in Saskatoon, Saskatchewan*

In 2018, the land cover in Saskatoon, Saskatchewan was characterized with 162.78 m2 of vegetation area, 422.79 m2 of urban area, 378.45 m2 of water area, and 1.85 m2 of soil area. Based on these findings, the majority of the study region was still characterized as urban area, which made up approximately 43.77% of the total area of the study region, while vegetation and water areas accounted for approximately 16.85% and 39.18% of the total area respectively. The soil land cover type however, still only accounted for a very small part of the total area at approximately 0.19%. The analysis of land cover change from 2017 to 2018 shows that the vegetation area continued to decrease from 220.20 m2 to 162.78 m2 in 2018. The urban area also slightly decreased as well from 442.05 m2 to 422.79 m2. The water area however, actually rose significantly since 2017, from 227.72 m2 to 378.45 m2. The soil area remained relatively the same, still only taking up less than 1% of the total area.



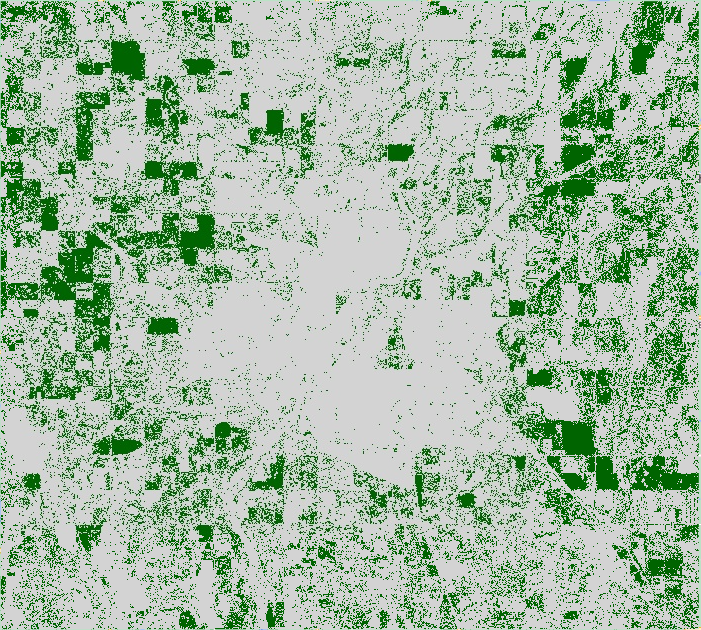
*Figure 10. Map showing vegetation and urban land cover types in 2018 in Saskatoon, Saskatchewan*

In 2019, the land cover in Saskatoon, Saskatchewan was characterized with 194.87 m2 of vegetation area, 403.04 m2 of urban area, 366.91 m2 of water area, and 1.23 m2 of soil area. Based on these findings, the majority of the study region was still characterized as urban area, which made up approximately 41.72% of the total area of the study region, while vegetation and water areas accounted for approximately 20.17% and 37.98% of the total area respectively. The soil land cover type however, still only accounted for a very small part of the total area at approximately 0.13%. The analysis of land cover change from 2018 to 2019 shows that the vegetation area started to increase from 162.78 m2 to 194.87 m2 in 2019. Based on the results from the previous years, vegetation area has been steadily declining every year from 2015 to 2018. The urban area slightly decreased from 422.79 m2 to 403.04 m2. The water area also slightly decreased since 2018, from 378.45 m2 to 366.91 m2. The soil area remained relatively the same, still only taking up less than 1% of the total area. According to figure 11, it is evident that some efforts have been made to preserve green spaces and/or promote urban forestry as vegetation slightly increased from the previous year.



*Figure 11. Map showing vegetation and urban land cover types in 2019 in Saskatoon, Saskatchewan*

In 2020, the land cover in Saskatoon, Saskatchewan was characterized with 220.18 m2 of vegetation area, 417.80 m2 of urban area, 326.83 m2 of water area, and 1.23 m2 of soil area. Based on these findings, the majority of the study region was still characterized as urban area, which made up approximately 43.25% of the total area of the study region, while vegetation and water areas accounted for approximately 22.79% and 33.83% of the total area respectively. The soil land cover type however, still only accounted for a very small part of the total area at approximately 0.13%. The analysis of land cover change from 2019 to 2020 shows that the vegetation area has continued to increase from 194.87 m2 to 220.18 m2 in 2020. The urban area also slightly increased as well, from 403.04 m2 to 417.80 m2. The results show that both vegetation and urban areas increased in the year 2020, but water areas decreased. The water areas decreased from 366.91 m2 to 326.83 m2. The soil area remained relatively the same, still only taking up less than 1% of the total area.



*Figure 12. Map showing vegetation and urban land cover types in 2020 in Saskatoon, Saskatchewan*

Comparing the results from 2013 to 2020, the land cover changes in 2020 show that the vegetation area decreased from 401.95 m2 to 220.18 m2 and the urban areas increased from 245.05 m2 to 417.80 m2. There was also an increase in water area from 327.94 m2 to 326.83 m2, while the soil area had relatively no change from 2.46 m2 in 2013 to 1.23 m2 in 2020. There was also a slight decrease in total land cover from 2013 to 2020 as the initial total land cover in 2013 was 977.40 m2, while in 2020 it was 965.04 m2. This could be due to the fact that not all land cover types were accounted for within the study area, resulting in different total land cover areas each year. To gain a better understanding of the changes in land cover types between 2013 to 2020, refer to the following tables: Table 1 provides a comparison of the areas of each land cover type during this period, and Table 2 provides a comparison of the percentage area occupied by each land cover type.

**Table 1.** Area of each land cover type from 2013 to 2020 in Saskatoon, Saskatchewan

|  | Vegetation | Urban | Water | Soil | Total |
| --- | --- | --- | --- | --- | --- |
| 2013 | 401.95 m2 | 245.05 m2 | 327.94 m2 | 2.46 m2 | 977.40 m2 |
| 2014 | 348.52 m2 | 316.28 m2 | 299.40 m2 | 1.85 m2 | 966.05 m2 |
| 2015 | 383.05 m2 | 283.17 m2 | 298.59 m2 | 1.23 m2 | 966.04 m2 |
| 2016 | 265.50 m2 | 343.39 m2 | 238.76 m2 | 1.85 m2 | 849.5 m2 |
| 2017 | 220.20 m2 | 442.05 m2 | 227.72 m2 | 1.85 m2 | 891.82 m2 |
| 2018 | 162.78 m2 | 422.79 m2 | 378.45 m2 | 1.85 m2 | 965.87 m2 |
| 2019 | 194.87 m2 | 403.04 m2 | 366.91 m2 | 1.23 m2 | 966.05 m2 |
| 2020 | 220.18 m2 | 417.80 m2 | 326.83 m2 | 1.23 m2 | 966.04 m2 |

**Table 2.** Percentage area of each land cover type from 2013 to 2020 in Saskatoon, Saskatchewan

|  | Vegetation | Urban | Water | Soil | Total |
| --- | --- | --- | --- | --- | --- |
| 2013 | 41.12% | 25.07% | 33.55% | 0.25% | 100% |
| 2014 | 36.08% | 32.74% | 30.99% | 0.19% | 100% |
| 2015 | 39.65% | 29.31% | 30.91% | 0.13% | 100% |
| 2016 | 31.25% | 40.42% | 28.11% | 0.22% | 100% |
| 2017 | 24.69% | 49.57% | 25.53% | 0.21% | 100% |
| 2018 | 16.85% | 43.77% | 39.18 | 0.19% | 100% |
| 2019 | 20.17% | 41.72% | 37.98% | 0.13% | 100% |
| 2020 | 22.79% | 43.25% | 33.83% | 0.13% | 100% |

## Discussion

The results of the land cover change analysis from 2013 to 2020 have provided valuable insights into the patterns and trends of urban development in Saskatoon, Saskatchewan. The results from the previous section suggests that the vegetation area declined over time, while the urban areas increased. The implications of the results for each year from 2013 to 2020 will be discussed in the following section, along with some of the limitations, issues, and constraints with the methods of this study.

The analysis of land cover change in Saskatoon from 2013 to 2014 revealed that there was a slight decrease in vegetation area, and an increase in urban and water areas. These changes have significant implications for the sustainability of urbanization in Saskatoon. The decrease in water area could have various implications and impacts on the ecosystem such as a decrease in precipitation and increase in water usage, or decreased water availability for plant and animal life in the area. It is important to investigate and understand the reasons behind this slight decrease, as it may have further implications regarding the city’s environmental and ecological well-being. The decrease in vegetation area and increase in urban area also implies that there was some degree of urbanization and land cover change in Saskatoon between 2013 and 2014. Normally, an increase in urban areas often comes at the expense of natural habitats and vegetation areas, which can have negative consequences on Saskatoon's life on land, but it is also possible for urbanization to have positive effects on other aspects such as economic growth or more job opportunities. Therefore, it is important to note both the positive and negative effects of urbanization when planning for smart land use in Saskatoon.

From 2014 to 2015, the results showed that there was a slight increase in vegetation area, a decrease in urban area, and little to no change in water area which have several implications for the study region. The decrease in urban area and increase in vegetation area may indicate some changes to land use and potential efforts to preserve and/or restore vegetation in Saskatoon since 2014. These efforts could have positive benefits for air quality and wildlife, improving the overall biodiversity. The water area having little to no change suggests that hydrology observed since 2014 have either stabilized or possibly corrected. As for the changes in land cover areas between 2014 and 2015, it suggests an increase toward urbanization and a possible stabilization of water areas. It also suggests the need for continued monitoring of land use changes and further strategies in order to promote sustainable urbanization in Saskatoon, Saskatchewan.

From 2015 to 2016, the results showed that the vegetation area significantly decreased, water area also decreased, while urban area significantly increased, becoming the majority land cover area of Saskatoon. These results imply that there may have been some land conversion or degradation, which in turn could have negative impacts on biodiversity and/or air quality which would negatively impact the sustainable development goal: life on land. The sudden increase in urban areas may be due to land use changes and/or an increase in population growth. The slight decrease in water area from 298.59 m2 to 238.76 m2 may indicate that there has been some changes in hydrology or water management practices from 2015 to 2016. The changes in land cover areas between 2015 and 2016 were much more prominent compared to the previous years, with urban areas now taking up the majority of the study region. This suggests a significant increase toward urbanization possibly due to urban sprawl.

From 2016 to 2017, the results showed a slight decrease in water area and a decrease in vegetation area, while urban areas continued to significantly increase. These results could implicate a continued trend of urbanization and a decrease in vegetative land cover which may have negative implications regarding the sustainable development goal: life on land in Saskatoon. For instance, a decrease in vegetation area could lead to a reduction in climate regulation or reduction in wildlife habitat. But on the other hand the increase in urban area could lead to increased stormwater runoff and flooding occurrences due to the implication that more urban area may lead to more impervious surfaces. The decrease in water area could also have negative implications for the environment as water bodies play an important role in supporting wildlife and changes in water quality and quantity can impact both human and ecological health.

From 2017 to 2018, the results showed that both the vegetation and urban areas decreased while the water area significantly increased which can have several implications. The decrease in vegetation area from 2017 to 2018 could be due to several factors such as urbanization, deforestation, or climate change. Since urban areas also slightly decreased from the previous year, there might not have been many changes implemented towards urbanization during this time. However, the increase in water area may suggest changes in precipitation patterns or increased development around water bodies.

From 2018 to 2019, the results showed that vegetation started to increase while urban and water areas decreased. This could indicate positive changes in the ecological health of the urban areas in Saskatoon, Saskatchewan and that there were efforts made to preserve green spaces and/or promote urban forestry. It could also implicate a decrease in urbanization or a shift in urban development patterns towards more sustainable practices. However, the decrease in water area suggests that there may have been some challenges in managing water resources during 2019.

From 2019 to 2020, the results showed that both the vegetation and urban areas continued to increase while water areas decreased. There are several implications to consider based on these results. An increase in vegetation area from 2018 to 2020 could imply that the efforts made to preserve green spaces and/or promote urban forestry have been successful which in turn could have positive effects on air and water quality. The slight increase in urban areas suggest that Saskatoon is becoming more urbanized and developed but it is also apparent that it has not increased by much over the last year, which could imply that there are still some efforts being made to ensure sustainable development.

### Limitations, Issues, and Constraints

Despite the valuable insights gained from using machine learning in land cover classification and change analysis, there are several limitations and constraints to consider. Firstly, the accuracy and reliability of a machine learning classification model depends heavily on the quality of the data being used. In this case, Landsat-8 imagery was used due to its long-term land cover changes and ability to provide images from earlier periods, enabling better analysis of changes over time, but at the cost of quality. Another limitation was that the training data points were all collected manually which introduced the potential for human error. This was problematic when having to distinguish between forest and grassland land covers as they share similar properties (i.e., NDVI value). In order to overcome this issue, land covers with similar properties, such as grassland, forest, and cropland were combined into a general term called vegetation. Furthermore, in order to collect one image for each year in order to perform a land cover change analysis, multiple images around the study region were mosaicked and clipped to the study region each year which meant that the classification model may not be a true representation of the area. Finally, although we used a random forest model, which is a popular and effective machine learning algorithm for land cover change analysis, the results revealed that the machine learning model was overfitted with an accuracy percentage of about 99%. While 66 training points and polygons were used during the training process, it is acknowledged that the study could have benefitted from additional training data.

## Conclusion

Landsat-8 imagery has lower resolution compared to other satellites, however, it remains the ideal satellite for analyzing long-term land cover changes. While higher resolution imagery is more suitable to collect accurate training data, Landsat-8 is able to provide images from earlier periods, enabling better analysis of changes over time, crucial for this study. Other methods, such as false-colour composites, were used to better identify land covers under low resolution. Implementing machine learning in land cover classification is very promising because as remote sensing technology is advancing, the availability of high-quality satellite imagery is increasing. This means that long-term land cover changes will not have to solely rely on Landsat imagery because improved satellites will have access to higher quality images from earlier periods in the future.

The land cover classification using random forest produced an overall high training accuracy. This is likely due to the quality and quantity of the training data used for the algorithm. As mentioned, the low-resolution satellite imagery used could have impacted the training accuracy because high-resolution imagery provides a more detailed representation of the landscape which improves the accuracy of the classification. Not to mention, the training data was collected manually so the choice of features used for each class, or the class itself could have been irrelevant or perhaps, the important characteristics of the landscape were not captured properly.

Moving forward, it will be ideal to address the issue of overfitting to improve the accuracy and generalization of the land cover classification method. Because the model fits the training data too well, it will perform poorly on new data. As it simply replicates the training data, it could be improved by only selecting the most relevant features for the study. For example, eliminating cropland and soil classes because they are not the most important classes when analyzing loss of vegetative areas to urban areas. Moreover, some classes had up to 80 training points that were very specifically selected and could have been avoided to reduce overfitting. In addition, cross-validation could have been utilized to improve overfitting since one is able to train the model on a subset then test it on others, allowing performance evaluation of the model on new data. This could identify overfitting early and avoid facing the issue at the end of the entire analysis. Combining multiple algorithms is also an option to improve the analysis because the model can leverage the strengths of each algorithm. It can create a more robust and better generalized model, while reducing bias and increasing flexibility.

Machine learning serves to facilitate land classification through its’ automating capabilities. As it’s trained to recognize patterns and features, it can efficiently detect changes in land cover over time and become a useful tool for other similar studies with room for improvement. Despite its slight complexity, machine learning can become a powerful tool for land cover classification, ultimately encouraging users to make more informed decisions about land use and sustainable urbanization.

## References

Abdi. (2020). Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. GIScience and Remote Sensing, 57(1), 1–20. <https://doi.org/10.1080/15481603.2019.1650447>

Amoakoh, Aplin, P., Awuah, K. T., Delgado-Fernandez, I., Moses, C., Alonso, C. P., Kankam, S., & Mensah, J. C. (2021). Testing the Contribution of Multi-Source Remote Sensing Features for Random Forest Classification of the Greater Amanzule Tropical Peatland. Sensors (Basel, Switzerland), 21(10), 3399–. <https://doi.org/10.3390/s21103399>

Barr. (2020). A Novel Technique for Segmentation of High Resolution Remote Sensing Images Based on Neural Networks. Neural Processing Letters, 52(1), 679–692. <https://doi.org/10.1007/s11063-020-10280-1>

Carpenter, Gjaja, M. N., Gopal, S., & Woodcock, C. E. (1997). ART neural networks for remote sensing: vegetation classification from Landsat TM and terrain data. IEEE Transactions on Geoscience and Remote Sensing, 35(2), 308–325. <https://doi.org/10.1109/36.563271>

European Space Agency. (2013). Sentinel-2: Technical specifications. Retrieved from <https://esamultimedia.esa.int/docs/S2-Data_Sheet.pdf>

Gallego, F. J. (2004). Remote sensing and land cover area estimation. *International Journal of Remote Sensing*, *25*(15), 3019–3047. <https://doi.org/10.1080/01431160310001619607>

Gao, J., Nuyttens, D., Lootens, P., He, Y., & Pieters, J. G. (2018). Recognising weeds in a maize crop using a random forest machine-learning algorithm and near-infrared snapshot mosaic hyperspectral imagery. *Biosystems Engineering*, *170*(Complete), 39–50. <https://doi.org/10.1016/j.biosystemseng.2018.03.006>

Government of Canada, S. C. (2021, October 27). *Census profile, 2016 census division no. 18, Census Division [Census Division], Saskatchewan and Saskatchewan [province]*. Census Profile, 2016 Census - Division No. 18, Census division [Census division], Saskatchewan and Saskatchewan [Province]. Retrieved January 31, 2023, from <https://www12.statcan.gc.ca/census-recensement/2016/dp-pd/prof/details/page.cfm?Lang=E&Geo1=CD&Code1=4718&Geo2=PR&Code2=47&SearchText=Division+No.+18&SearchType=Begins&SearchPR=01&B1=All&GeoLevel=PR&GeoCode=4718&TABID=1&type=0>

Government of Canada, S. C. (2022, December 6). *Census profile, 2021 census of Population*. Profile table, Census Profile, 2021 Census of Population - Division No. 18, Census division (CDR) [Census division], Alberta. Retrieved January 31, 2023, from <https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/details/page.cfm?Lang=E&SearchText=Division+No.+18&DGUIDlist=2021A00034818&GENDERlist=1%2C2%2C3&STATISTIClist=1&HEADERlist=0>

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, J. R. G. Townshend. (2021). High-Resolution Global Maps of 21st-Century Forest Cover Change, Science 342, 850–53. <http://earthenginepartners.appspot.com/science-2013-global-forest>.

Jin, Y., Liu, X., Chen, Y., & Liang, X. (2018). Land-cover mapping using Random Forest classification and incorporating NDVI time-series and texture: a case study of central Shandong. *International Journal of Remote Sensing*, *39*(23), 8703–8723. <https://doi.org/10.1080/01431161.2018.1490976>

Kamusoko. (2019). Remote Sensing Image Classification in R (1st ed. 2019.). Springer Singapore. <https://doi.org/10.1007/978-981-13-8012-9>

Kattenborn, Leitloff, J., Schiefer, F., & Hinz, S. (2021). Review on Convolutional Neural Networks (CNN) in vegetation remote sensing. ISPRS Journal of Photogrammetry and Remote Sensing, 173, 24–49. <https://doi.org/10.1016/j.isprsjprs.2020.12.010>

Lee, & Jung, H.-S. (2019). Machine Learning Techniques Applied to Geoscience Information System and Remote Sensing. MDPI - Multidisciplinary Digital Publishing Institute.

Mountrakis, G., Im, J., Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66, 3, 247-259. <https://doi.org/10.1016/j.isprsjprs.2010.11.001>.

Pahlevan, Schott, J. R., Franz, B. A., Zibordi, G., Markham, B., Bailey, S., Schaaf, C. B., Ondrusek, M., Greb, S., & Strait, C. M. (2017). Landsat 8 remote sensing reflectance (Rrs) products: Evaluations, intercomparisons, and enhancements. Remote Sensing of Environment, 190, 289–301. <https://doi.org/10.1016/j.rse.2016.12.030>

Patil, S. (2019). Overfitting and underfitting principles. Towards Data Science. Retrieved from <https://towardsdatascience.com/overfitting-and-underfitting-principles-ea8964d9c45c#:~:text=Underfitting%20means%20that%20your%20model,model%20makes%20not%20accurate%20predictions>.

Pettinger. (1982). Digital classification of Landsat data for vegetation and land-cover mapping in the Blackfoot River watershed, southeastern Idaho : a case study, including step-by-step procedures for computer-assisted analysis of Landsat digital data, with emphasis on assessment of classification accuracy and generation of output products. U.S. Govt. Print. Off.

Russo, M. (2010). Examining the relationship between urban form and public health: A literature review. National Institutes of Health, 32(2), 13-21. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2936977/#:~:text=Urban%20sprawl%20has%20occurred%20largely,coincide%20with%20the%20public%20good.&text=Many%20community%20leaders%20have%20found,planning%20to%20promote%20smart%20growth>.

Statistics Canada. (2019, February 13). Environmental and Clean Technology Products Economic Account, 2017. Retrieved from <https://www.statcan.gc.ca/en/subjects/standard/environment/elc/2017-1>

Statistics Canada. (2021). Census Profile, 2021 Census of Population: Saskatoon, City [Data file]. Retrieved from <https://www12.statcan.gc.ca/census-recensement/2021/dp-pd/prof/details/page.cfm?Lang=E&GENDERlist=1&STATISTIClist=1&HEADERlist=0&DGUIDlist=2021A00054711066&SearchText=saskatoon>

The Global Goals. (2022, March 15). Goal 15: Life on land. Retrieved from <https://www.globalgoals.org/goals/15-life-on-land/>

Ulloa-Torrealba Y, Stahlmann R, Wegmann M, Koellner T. Over 150 Years of Change: Object-Oriented Analysis of Historical Land Cover in the Main River Catchment, Bavaria/Germany. *Remote Sensing*. 2020; 12(24):4048. <https://doi.org/10.3390/rs12244048>

Witze, A. (2019, February 25). Cities are the future. Is that a good thing? National Geographic. Retrieved from <https://www.nationalgeographic.com/environment/article/urban-threats>

Ying. (2019). An Overview of Overfitting and its Solutions. Journal of Physics. Conference Series, 1168(2), 22022–. <https://doi.org/10.1088/1742-6596/1168/2/022022>

Zhang, F., & Yang, X. (2020). Improving land cover classification in an urbanized coastal area by random forests: The role of variable selection. *Remote Sensing of Environment*, *251*(Complete). <https://doi.org/10.1016/j.rse.2020.112105>