B. TECH PROJECT REPORT

On

FORECASTING THE INUNDATION EXTENT OF RAMSAR SITES USING CMIP6 PROJECTIONS

By

ERUMALLA SAI KUMAR

190004013

DISCIPLINE OF CIVIL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY INDORE

November 2022

FORECASTING THE INUNDATION EXTENT OF RAMSAR SITES USING CMIP6 PROJECTIONS

A PROJECT REPORT

Submitted in partial fulfillment of the requirements for the award of the degrees

of

BACHELOR OF TECHNOLOGY

In

CIVIL ENGINEERING

Submitted by:

Erumalla Saikumar

Guided by:

Prof. Manish Kumar Goyal

(Professor, Civil Engineering)

INDIAN INSTITUTE OF TECHNOLOGY INDORE
November 2022

CANDIDATE'S DECLARATION

I hereby declare that the project entitled "**FORECASTING THE INUNDATION EXTENT OF RAMSAR SITES USING CMIP6 PROJECTIONS**" submitted in partial fulfilment for the award of the degree of Bachelor of Technology in 'Civil Engineering' completed under the supervision of Prof. Manish Kumar Goyal, Professor, Civil Engineering, IIT Indore is an authentic work.

Further, I/we declare that I/we have not submitted this work for the award of any other

degree elsewhere.

E. SaiKumar
21/11/2022
Erumalla Sai Kumar

Signature and name of the student(s) with date

_____

CERTIFICATE by BTP Guide(s)

It is certified that the above statement made by the students is correct to the best of my/our knowledge.

Signature of BTP Guide(s) with dates and their designation

Preface

This report on "**FORECASTING THE INUNDATION EXTENT OF RAMSAR SITES USING CMIP6 PROJECTIONS**" is pre-pared under the guidance of Prof. Manish Kumar Goyal, Professor, Civil Engineering.

Through this project, I have made some techniques to collect data of inundation areas of wetlands using Landsat images. Google earth engine platform is used for the process of making inundation maps. Inundation area values are extracted from the inundation maps which are used for the model training. Further I used different machine learning techniques to train the data and choose best of them as my model based on the obtained results like accuracy score, MSE. RMSE, MAE, R2 score. From the results of the model, trend analysis id done on the Ramsar sites for their inundation pattern.

I have tried to the best of our abilities and knowledge to explain the content in a lucid manner using tables, source codes, flow charts etc.

Erumalla Saikumar

190004013

BTech 4th Year

Discipline of Civil Engineering

IIT Indore

<u>Acknowledgements</u>

Saikumar Erumalla

B.Tech. IV Year

Discipline of Civil Engineering

IIT, Indore

<u>Contents</u>

List of Figures

LIST OF TABLES

Abstract

Wetlands are essential for preserving numerous natural cycles and providing habitat for a wide variety of wildlife. They defend our coastlines, act as a natural sponge against flooding and drought, and contribute to the fight against climate change. Wetland also provide environmental benefits like Water purification, Flood protection, Shoreline stabilization. They are essential to maintaining the appropriate balance of the ecosystem. The degree of submergence of these wetlands has a significant effect on this balance. Because of this ecological significance, it is essential to assess how flooding has evolved and offer suitable management and conservation methods. The Ramsar Convention categorises wetlands as Ramsar Wetlands based on certain parameters which of national importance. In this study 15 wetlands which were declared as Ramsar sites are studied for their inundation pattern from 1991 to 2060. The study makes use of pre-processed Landsat imageries (1991-2020) for the extraction of inundation area of each site. The climate data i.e., precipitation and temperature data are collected from CMIP6. Some machine learning models are used to forecast the inundation area to future i.e., from 2021 to 2060 based on the climate data extracted from CMIP6. Trend analysis is done on the inundation area and climate trends for both past and future data. The pattern in the trends indicated us that most of the sites for which inundation trend is decreasing shows decreasing trend in precipitation and an increasing trend in min and max temperature. Also, the sites which has increasing inundation trend shows increasing and decreasing trend in precipitation and temperature data respectively. It is also observed that there are 7 decreasing sites in the past out of which 3 were significantly decreasing. But in the future, there are 8 wetlands that showed decreasing trend out of which 2 are significantly decreasing.

1 Introduction

1.1 Overview

Wetlands contribute only about 2.6 percent of the land area yet play an important role in hydrology and source of organic carbon for over 20 percent of the globe (Scheyer et al., 2013). A diverse spectrum of animals relies on wetlands for their survival, and they sustain a diverse range of biological niches as well as flora and fauna (Valenti et al., 2020). Wetlands, on a larger scale, contribute to controlling regional climate by sequestering carbon and providing vital habitation for national and international migrant species. Wetland ecosystems are altering and vanishing because of agricultural and industrial expansion, water deviation, increased temperatures, and changes in rainfall patterns (Corcoran et al., 2013). Wetlands are projected to have lost and shrunk over the last several decades (Creed et al., 2017), threatening important ecosystem functions such as water refinement, control of flooding, protection of biodiversity, and supply of food(Amani et al., 2021). Since 1700, natural wetland area has decreased by 87 percent and wetland destruction has been 370 percent faster in the twentieth and early twenty-first centuries (Boretti and Rosa, 2019). Inland wetlands have suffered greater losses, which are currently occurring at a quicker rate, Asia is now seeing the highest rate of loss. The COVID-19 outbreak has altered our perceptions of healthiness and the environment, with a greater appreciation of the value of nature for overall health (Bavel et al., 2020). The pandemic has impacted almost every part of our life and caused the enormous loss. Additionally, the pandemic has forced us to appreciate nature more highly because of the benefits of wetlands to people's health and well-being (Shirzaei and Bürgmann, 2018). Wetlands that are effectively managed are critical to the public's well-being and livelihoods. Wetland ecological system approaches can provide health advantages to everyone. The Ramsar Convention was signed in 1971 to encourage the protection and responsible use of wetlands across the globe and it is the only international agreement focusing on wetlands (Xi et al., 2021). Many wetlands had been classified as Wetlands of International Importance based on the Ramsar site criteria. These wetlands are representative, uncommon, or unique, and have a special significance for biodiversity conservation (Xi et al., 2021). Thus, the conservation actions paired with restoration initiatives can improve the function of wetlands as natural climate solutions, allowing mankind to meet the objectives of the Paris Climate Accords and the UN Decade on Ecosystem Restoration (Cook-Patton et al., 2021; Temmink et al., 2022). Analysis of the past and future inundation trend of these sites help us understand their inundation pattern and appropriate steps can be taken. As per the

recent article ("Valuing wetlands," 2021), wetlands play a role in 75 indicators of SDGs by the United Nations.

Water drainage, pollution, unsustainable use, invasive species, disturbed flow patterns from dams, and sediment dumping from upstream deforestation and soil erosion pose threats to the world's surviving wetlands. The survival of humans and the world depends on wetlands. Due to the importance of the wetlands in maintaining balance in the ecosystem, it is necessary for us to study the inundation pattern of the wetlands. Studying the inundation pattern of the wetland help us to understand the variation in water level and identify wetlands which are in path to extinction. Also knowing the inundation pattern of the wetlands for the future help us to plan safety measures that will improve the wetland and its ecosystem.

1.2 Literature Review

Various classification methods, such as unsupervised and supervised band thresholding, band ratios, indices, various regression trees, and combinations of these methods, have been proposed in the past(Inman and Lyons, 2020). Band thresholding has emerged as the most efficient and accurate of these approaches. Murray-Hudson et al., 2015 proposed a method for thresholding the Short-Wave Infrared (SWIR) band and used MODIS data to generate high accuracy results (Murray-Hudson et al., 2015; Wolski et al., 2017). SWIR band can differentiate between inundated areas with thick vegetation and dryland vegetation with extreme accuracy.

These methods are used in many previous studies for the extraction of inundation area of waterbodies. I have used the same methos on the Ramsar wetlands to extract the inundation area. Many of the previous studies are only about the previous area. But in this project, I have used the past inundation area pattern to forecast the area to future and tried to find the inundation patterns of wetlands in future. Climate data has been extracted using CMIP6 projections.

2 Data and Methods

2.1 Data set

To obtain the dataset of the complete time series from 1991 to 2020, we utilized all three Landsat sensors (Landsat 5 TM, Landsat 7 ETM+, Landsat 8 OLI). Landsat 5 (availability: 1984-01-01 - 2012-05-05) and Landsat 7 (availability: 1999-01-01 - present) images have four visible and near-infrared (VNIR) bands, two short-wave infrared (SWIR or B7) bands, and one thermal infrared (TIR) band processed to orthorectified brightness temperature, while Landsat 8 (2013-04-11 - present) images have eleven bands (Murray-Hudson et al., 2015; Wolski et al., 2017). The resolution of the SWIR bands in all Landsat sceneries is 30 meters per pixel. All of the above datasets are available in the Google Earth Engine data catalogue and are ready to use (Gorelick et al., 2017; Zurqani et al., 2020). The details of Ramsar sites analysed in the study are mentioned in the Table 1 and location has been shown in Fig.1.

2.1.1 Data set for Creating Inundation maps of Wetland

 The wetlands analyzed in the study are the 15 recently declared Ramsar sites of India i.e., Karikili Band Sanctuary, Pallikaranai Marsh Reserve Forest, Pichavaram Mangrove, Sakhya Sagar, Pala Wetlands, Koonthankulam Bird Sanctuary, Satkosia George, Nanda Lake, Gulf of Mannar Marine Biosphere, Ranganathittu Bird Sanctuary, Vembannur Wetland Complex, Vellode Bird Sanctuary, Sirpur Wetland, Vedanthangal Bird Sanctuary, and Udhayamarthandapuram Bird Sanctuary. The majority of the wetland shapefiles used in the study came from official Ramsar Website. Karikili, Pallikaranai, Nanda Lake, Vellode Bord Sanctuary shapefiles were not available. As a result, using the coordinate (longitude and latitude) information available on the official website, the shapefiles for these places are manually produced in Google Earth Pro. To cover all of the characteristics, the shapefiles included the wetlands as well as portions of the surrounding region. Individually, these shapefiles were uploaded to the GEE platform for further processing.

Table 1: Details of the Ramsar sites

S. No.	Ramsar Site No.	Site name	Country	Area (ha)	Latitude	Longitude
1	2470	Satkosia Gorge	India	98196.72	20.57241	84.83235
2	2471	Nanda Lake	India	42.01	15.23687	74.1075
3	2472	Gulf of Mannar Marine Biosphere Reserve	India	52671.88	9.115879	78.78691
4	2473	Ranganathittu Bird Sanctuary	India	517.7	12.40785	76.68214
5	2474	Vembannur Wetland Complex	India	19.746	8.181778	77.37623
6	2475	Vellode Bird Sanctuary	India	77.185	11.25186	77.65191
7	2476	Udhayamarthandapuram Bird Sanctuary	India	43.767	10.45057	79.55471
8	2477	Vedanthangal Bird Sanctuary	India	40.348	12.54649	79.85602
9	2478	Sirpur Wetland	India	161	22.69951	75.81228
10	2479	Koonthankulam Bird Sanctuary	India	72.04	8.49569	77.75398
11	2480	Karikili Bird Sanctuary	India	58.442	12.59893	79.84242
12	2481	Pallikaranai Marsh Reserve Forest	India	1247.537	12.92785	80.22048
13	2482	Pichavaram Mangrove	India	1478.642	11.43817	79.78659
14	2483	Sakhya Sagar	India	248	25.43427	77.707
15	2484	Pala Wetland	India	1850	22.1971	92.9022
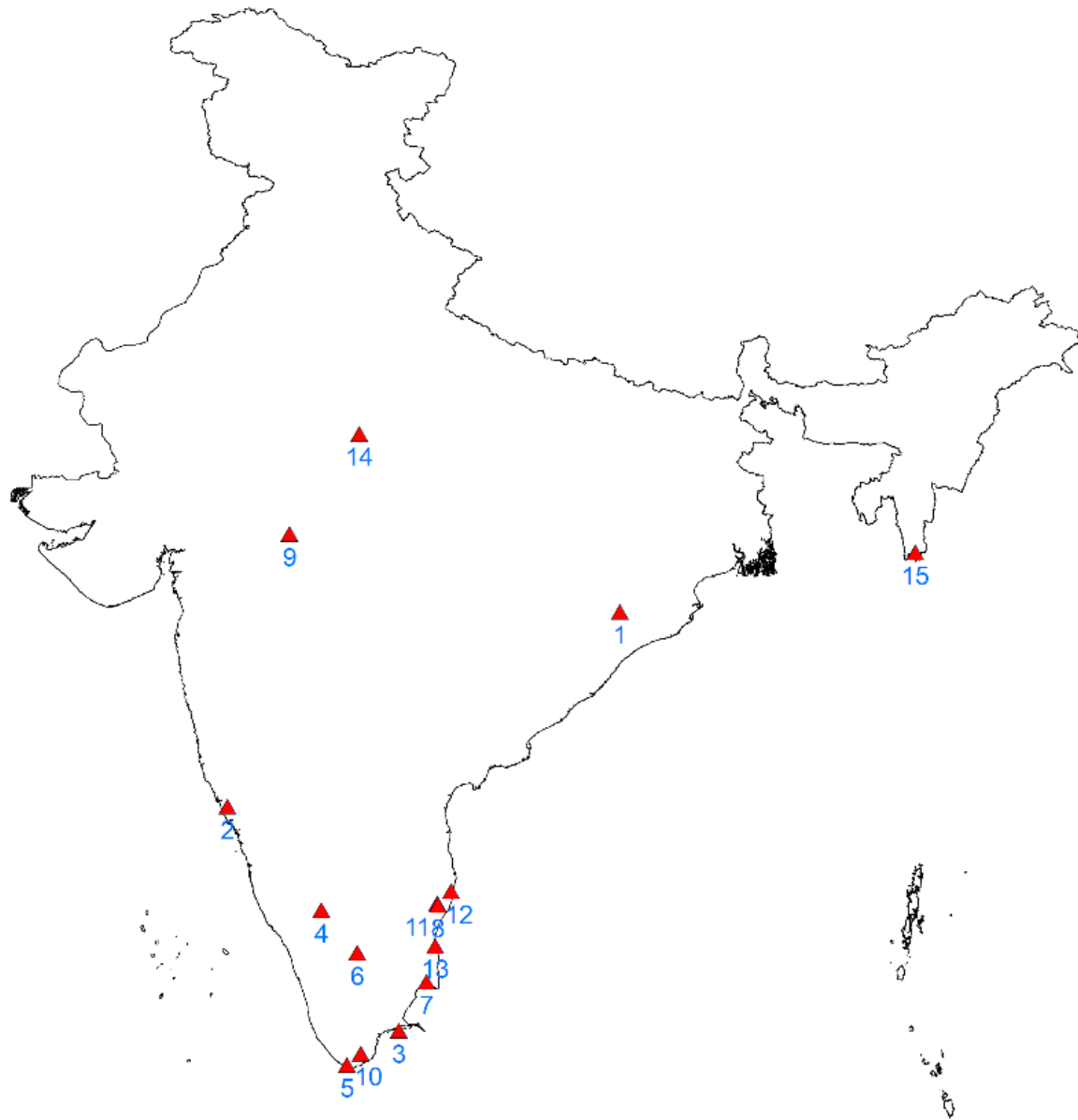
Figure 1: Map representing the location (by red triangle) of wetlands taken into the study and the label below these triangles represents the wetland by serial number mentioned in the Table 1

2.1.2 Data for Predicting the Inundation using CMIP6

Climate data i.e., maximum temperature, minimum temperature and Average precipitation of each wetland for each year has been extracted from CMIP6 using the coordinate information available on the Ramsar website. The data has been extracted from 1991 to 2060. The data from 1991 to 2020 has been used for training and 2021 to 2060 is used for the prediction.

2.2 Methodology

The majority of wetlands in the South Asian subcontinent experience yearly flooding events that coincide with the southwest monsoon season and are inundated to the greatest extent between June and September. During this season of southwest monsoon rains, they obtain the majority of their water intake (Bassi et al., 2014; IMD, 2020; Kumar et al., 2010). However, the melting of glaciers provides a substantial amount to some of wetlands. During the months of March-April and July-August, the majority of this water enters the marsh (Corcoran et al., 2013; Gallant, 2015; Maurya and Singh, 2016; Murray-Hudson et al., 2015). In July and August, intake from glacier melt combines with inflow from southwest monsoon precipitation, resulting in maximum discharge inflow into the wetlands and maximum inundation (Kumar et al., 2010; Pal and Al-Tabbaa, 2009). As a result, the current study could be condensed to four months to reduce computational costs and simplify the understanding of variations in inundated areas (Gouda et al., 2020).

Regression is a statistical method to determine the strength and character of the relationship between one dependent variable (usually denoted by Y) and a series of other variables (known as independent variables) Machine Learning Regression Model performs the task to predict a dependent variable(target) based on the given independent variable(s). So, this regression technique finds out a relationship between a dependent variable and the other given independent variables. We aim to predict the inundation area (target) of the wetlands based on the dependent variable such as the max temperature, min temperature, Site area, Site Id, and average precipitation. So different Machine Learning Regression Models have been assessed to predict the Inundation area and the model that gives high accuracy is used for future forecasting. Detailed methodology is represented in the form of flow chart in Fig.2.
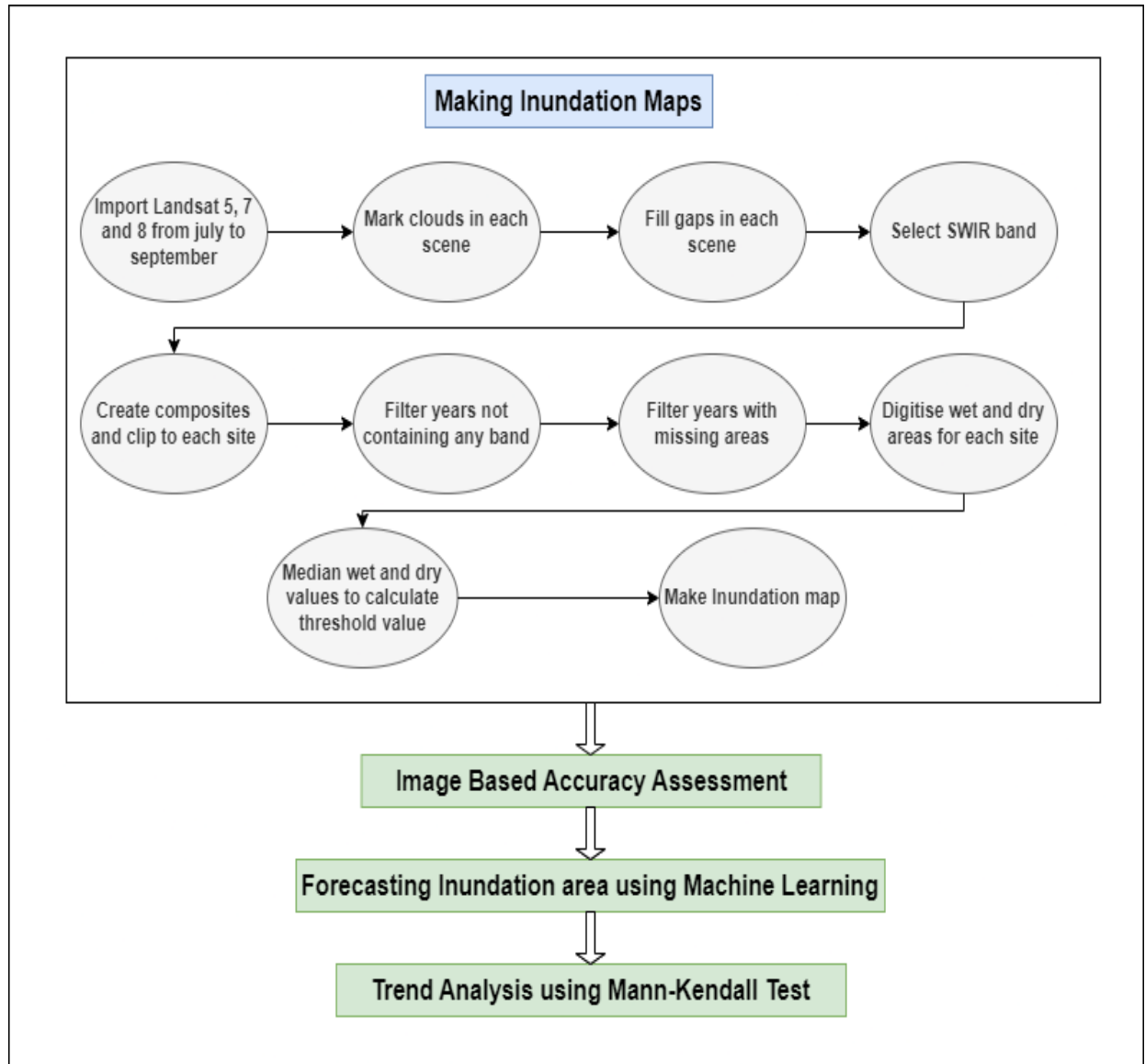
Figure 2: The schematic flowchart of the process of Methodology

2.2.1 Creating Inundation maps of wetlands

a) Cloud-masking

Clouds and cloud shadows are present in the Landsat sceneries, and they need to be masked to generate proper composites to enhance the accuracy of the inundation maps(Murray-Hudson et al., 2015; Wolski et al., 2017). The pixels classified as cloud or cloud shadow on the Landsat cloud mask band were masked for each scene using a gap-filling method (Yin et al., 2016). The pixels were then filled with the median value for the pixel from a year before or after the scene's date.

b) Landsat composites

A gap-filling method was afterward used to cloud masked images. The SWIR band (B7) was chosen for each scenario. Each year's composites are made using all the scenes available from June to September (Pal and Al-Tabbaa, 2009). A median of the corresponding pixel values from all the scenes of that year is evaluated and designated as the value of the corresponding pixel in the composites to be generated for each pixel in the study area.

c) Filtering bad composites

The SWIR band was missing for some of the composites created. Those composites were manually filtered by removing them from the image collection. Most of the cloud masking is done via the cloud masking algorithm (Corcoran et al., 2013; Gallant, 2015). However, there were some regions in almost all the coastal sites where the pixels were classified as a cloud most of the time. The masking algorithm made the pixels transparent in these circumstances. A filtering algorithm was used to filter these pixels. This resulted in a set of composites without even a single masked pixel (Hird et al., 2017). From site to site, the number of final faultless composites varied.

d) Creating inundation maps from the composites

The inundation maps are made from composites by thresholding the SWIR band pixel values. We manually assessed and digitized permanent wet (for example, the lake's central region and permanent channels) and dry areas (like barren land or hill region near the wetland) for each site.

The median SWIR values for wet ($SWIR_{wet}$) and dry ($SWIR_{dry}$) inundated areas for each composite were calculated using these digitized areas and a composite-specific $SWIR_{threshold}$ (to account for the dynamic seasonal and annual nature of the inundation patterns in wetlands) value was calculated using equation (1) (Wolski et al., 2017).

$$SWIR_{threshold} = SWIR_{wet} + 0.3 \left(SWIR_{dry} - SWIR_{wet} \right) \hspace{4cm} (1)$$

The classifier compares each pixel's SWIR value to its $SWIR_{threshold}$ for each composite. The pixels having SWIR values less than or equal to $SWIR_{threshold}$ are classified as inundated, while pixels having SWIR values larger than or equal to $SWIR_{threshold}$ are classified as dry (Milzow et al., 2009). As a result, each pixel with a particular SWIR value is classified and transformed into one of two values: 0 for dry pixels and 1 for inundated pixels, and an inundation map is generated using these changed pixels (Corcoran et al., 2013; Gumbricht et al., 2004).

e) Validation / Image-Based accuracy assessment

Validation of the thresholding method was required to determine its applicability, particularly for remote sensing applications in the Indian subcontinent. Comparing the inundation maps with historical imageries available in Google Earth Pro (GEP) was the simplest method. Based on the availability of historical imageries in GEP, a random set of five years has been chosen for each Ramsar site. A set of 50 random points was generated for each of the five years using GEE's random points function and exported as a KML file. The pixel values were extracted from the inundation maps at each of these points using GEE's Sample Region function and exported as a CSV file with only two values: 1 for inundated pixels and 0 for dry pixels. The KML file was then imported into GEP and each point was examined visually, and classed as dry (i.e., 0) or inundated (i.e., 1). Due to the lack of in-situ data, we had no choice but to use this as our reference dataset. For each site, this technique was repeated for each random point in each year of the five years. For each site, a collection of 250 random points was obtained, with their pixel values retrieved from the maps and reference data from the imagery. A total of 3,750 (250*15=3750) points were validated. Each site's error matrix was created separately. Overall Accuracy was calculated as the sum of the diagonal elements (correctly classified) in the error matrix divided by the total sampled points, Producers Accuracy was calculated as the diagonal entry of each column in the error matrix

divided by its respective column total, and User's Accuracy was calculated as the diagonal entry of each row in the error matrix divided by its respective row total.

2.3 Trend Analysis

The Mann-Kendall (MK) test was used on each site individually to discover trends in the variation of inundation extent (Blain, 2013; Mann, 1945). It was carried out under the assumption that a p-value of less than 0.05 indicated a significant trend (also represented by an absolute Zc score greater than 1.96). Since every site had a varied number of maps, the analysis was done based on the information available, and the trends discovered were projected across a 30-year period to allow for a uniform comparison of all the sites. The trends are classified on basis of Z-value. Z values is user as an indicator for Classification of the trends. A Z value less than 0 indicates a decreasing trend whereas value less than -1.96 signifies significant decreasing trend. Z value greater than 0 indicates increasing trend whereas a value greater than 1.96 indicates significantly increasing trend.

2.4 Predicting the Inundation using CMIP6

After getting the inundation area of the sites for all the available years (1991 to 2020), the data is merged into a single file for the training of the ml models. Along with the inundation data the maximum yearly temperature, minimum yearly temperature, and average yearly precipitation were taken as parameters for the training of the model. Different machine learning regression models like Linear Regression, Lasso Regression, Ridge Regression, SVM, Random Forest Regressor, and XGB Regressor were trained based on the data. The accuracy of each of the models has been checked and the model that performs better is used for future inundation forecasting.

a) Linear Regression

Linear Regression is a very commonly used machine learning regression algorithm that can be imported from the Linear Regression class. Assuming that the input variable isn't correlated with each other, a single input variable which is the significant one is used to predict one or more output variables (Su et al., 2012). Linear regression is represented as:

$$y = b*x + c$$

where y is dependent variable, x is independent variable, b is slope of the best fit line that could get accurate output and c is its intercept. It is possible for there to be a loss in output, which is usually measured as the square of the difference between the predicted and actual output, i.e., the loss function, unless there is a perfect line connecting the dependent and independent variables.

b) Ridge Regression-The L2 Norm

Ridge regression extension of a linear regression that tries to minimize the loss and also uses multiple regression data. In situations when the independent variables are highly correlated, ridge regression is a technique for estimating the coefficients of multiple-regression models. Its coefficients are estimated using ridge, a biased estimator with lower variance than ordinary least squares (OLS), which results in shrinking of the coefficients (Zhang et al., 2022). We can also reduce the complexity of the model by using this form of model.

The cost function for ridge regression: $\text{Min} \left(\|Y - X(\theta)\|^2 + \lambda \|\theta\|^2 \right)$

Lambda is the penalty term. λ given here is denoted by an alpha parameter in the ridge function. So, we can regulate the penalty term by varying the values of alpha. The penalty is greater with larger alpha values, which reduces the magnitude of coefficients. The regularization strength of Ridge is determined by λ value. Large λ cause the coefficients to contract, flattening the model and reducing its variance. As a result, regularisation techniques are frequently employed to avoid model overfitting.

c) Lasso Regression -The L1 Norm

LASSO stands for **L**east **A**bsolute **S**hrinkage and **S**election **O**perator. Lasso regression is a regularisation technique. For a more accurate prediction, lasso is preferred over other regression techniques. Shrinkage is used in this model. When data values shrink toward the mean, it is known as shrinkage. Simple, sparse models are encouraged by the lasso approach (i.e., models with fewer parameters) (Pereira et al., 2016). When a model exhibits a high degree of multicollinearity or when you wish to automate some steps in the model selection process, such as variable selection and parameter removal, this specific sort of regression is ideally suited. L1 regularisation is used in Lasso Regression. Because it does feature selection automatically, it is employed when there are more features. Lasso regression's mathematical formula is

Residual Sum of Squares + λ * (Sum of the absolute value of the magnitude of coefficients)

The quantity of shrinkage is indicated by λ. A predictive model is constructed using simply the residual sum of squares when λ = 0, which denotes that all features are taken into account. As λ gets closer to infinity, it eliminates more and more features, implying that no feature is taken into account. The bias increases with increase in λ whereas variance increases with decrease in λ.

d) Random Forest Regressor

Random forest is a Supervised Machine Learning Algorithm that is used widely in Classification and Regression problems. On various samples, it constructs decision trees and uses their majority vote for regression(Svetnik et al., 2003). The Random Forest Algorithm's ability to handle data sets with both continuous variables, as in regression is one of its most significant qualities. The steps involved in it is:

I.   Choose K arbitrary data points from the training set.
II.   Create a decision tree using these data points.
III.   Repeat the preceding steps for the number of trees you need to build (the number is provided as an input).
IV.   Make each tree for a new data point predict the values of the dependent variable given the input.
V.   Assign average value of the predicted values to the final output.

e) Support Vector Machines(SVM)

Support Vector Regression is a supervised learning algorithm that is used to predict discrete values. The SVMs and Support Vector Regression both operate on the same theory. Finding the best fit line is the basic aspect of SVR. The hyperplane with the most points on it is the best-fitting line in SVR. The SVR seeks to fit the best line within a threshold value, in contrast to other regression models that aim to reduce the error between the real and projected value(Pisner and Schnyer, 2020). The distance between the boundary line and the hyperplane is the threshold value. SVR is difficult to scale to datasets with more than a few ten thousand samples since its fit time complexity is more than quadratic with the number of samples.

SGD Regressor or Linear SVR are used for large datasets. While linear SVR merely takes into account the linear kernel, it offers a faster implementation than SVR. Because samples whose prediction is close to their objective are ignored by the cost function, the model created by Support Vector Regression only rely on a portion of the training data.

f) XGB Regressor

 XGBoost is a popular and efficient implementation of the gradient boosted trees algorithm. Gradient boosting is a supervised learning process that combines the predictions of a number of weaker, simpler models to attempt to properly predict a target variable.

Regression trees serve as the weak learners when utilising gradient boosting for regression, and each one of them associates each input data point with a leaf that holds a continuous score. With a convex loss function (based on the difference between the predicted and target outputs) and a penalty term for model complexity, XGBoost minimises a regularised (L1 and L2) objective function (in other words, the regression tree functions) (Shehadeh et al., 2021). Adding new trees that forecast the residuals or errors of earlier trees, which are then integrated with earlier trees to produce the final prediction, is how the training process is carried out iteratively. Because the loss when introducing new models is minimised, the technique is known as gradient boosting.

2.5. Evaluation Metrics for ML Models

The skill or performance of a regression model must be reported as an error in the predictions. If we are predicting a numeric, we don't want to know if the model predicted the value exactly instead, we only want to know how close the predicted values are to the original values. Error will address exactly this and summarize on average how close predictions are to their expected values . Four-error metrics are used in the study for evaluating and reporting the performance of a regression model.

a) Mean Squared Error

MSE is popular error metric used for regression problems. It is an important loss function for algorithms fit. And, also optimized using the least squares framing of a regression problem. Least squares refer to minimizing mean squared error between predicted and expected value

$$\text{Mean Squared Error} = \frac{1}{N} \sum_{k=0}^{n} (y\_i - yt\_i)^2$$

b) Root Mean Squared Error

RMSE, is just an extension of the mean squared error. The square root of the MSE is calculated in RMSE. The units of the RMSE are same as the original units of the target. So, it may be common to use MSE loss to train a regression model, and to use RMSE for evaluation and report the model performance.

$$\text{Root Mean Squared Error} = \sqrt[2]{\frac{1}{N} \sum_{k=0}^{n} (y\_i - yt\_i)^2}$$

c) Mean Absolute Error

The Mean Absolute Error score is calculated as the average of the absolute error values. Absolute is a mathematical function that will make a number positive. The difference between an original and predicted value may be positive or negative but it is forced to be positive while calculating the Mean Absolute Error.

$$\text{Mean Absolute Error} = \frac{1}{N} \sum_{k=0}^{n} |y_i - yt_i|$$

d) R2_Score

 The coefficient of determination also called the R^2 score is used for the evaluation of the performance of regression model. R2 score is the amount of the variation in the dependent attribute which is predictable from the input independent variable(s). It is used to check how well-observed results are reproduced by the model, depending on the ratio of total deviation of results described by the model.

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

Where y_i is the i'th expected, yt_i is the i'th predicted value.

SS_{res} is the sum of squares of the residual errors.

SS_{tot} is the total sum of the errors.

3 Results and Discussions

The inundation area of different sites has been extracted for the available years from the interval of 1991 to 2020. The inundation maps generated through the thresholding technique showed annual variation in each site. Each site showed a different pattern of variation, in different amounts. The number of composites varied from site to site. A total of 337 composites were extracted for 30 years for the 15 sites. The inundation maps extracted from Google earth Engine using the methos mentioned in Methodology are shown in Figure 3.

3.1 Validation of detected Inundation maps

The thresholding method was found to be accurate enough to be trusted as a good method for generating inundation maps based on the accuracy results. An image-based accuracy assessment was performed. There was a total of 3750 points that were validated. Through visual inspection of historical imageries, about 2334 points were identified as dry, with 2164 points correctly classified as dry in the maps. 1414 points on visual assessment were found to be wet, with 1314 points in the inundation maps correctly classified as wet. There was a total of 274 points that were incorrectly classified (inundated points were marked as dry and vice versa). Because of several factors such as the nature of the site, digitizing areas, and the variation in spectral values of wet and dry areas, the accuracy of each site varied. Sites with a lot of variances in inundation extent, such as those where practically every patch of the area was inundated in at least one year, showed low accuracy. The different accuracy values of each site analysed in the study are shown in figure 4.

The overall accuracy ranged from 90 % at Gulf of Mannar Marine and Koonthankulam to 96.4 % at Pala Wetland (Figure 2). The sites overall accuracy was found to be 92.64 ± 2.03 %, with average dry and wet Producer's accuracies of 94.70 ± 3.63 % and 86.41 ± 5.62 %, respectively (see Figure 2 for details on each site). and the dry and wet User's accuracy of 91.77 ± 3.24 % and 92.53 ± 3.35 %, respectively.

3.2 Forecasting inundation maps

The inundation area value is extracted from the composite's year-wise data and has been exported into an excel file. Now the inundation area is merged with climatological data which is extracted
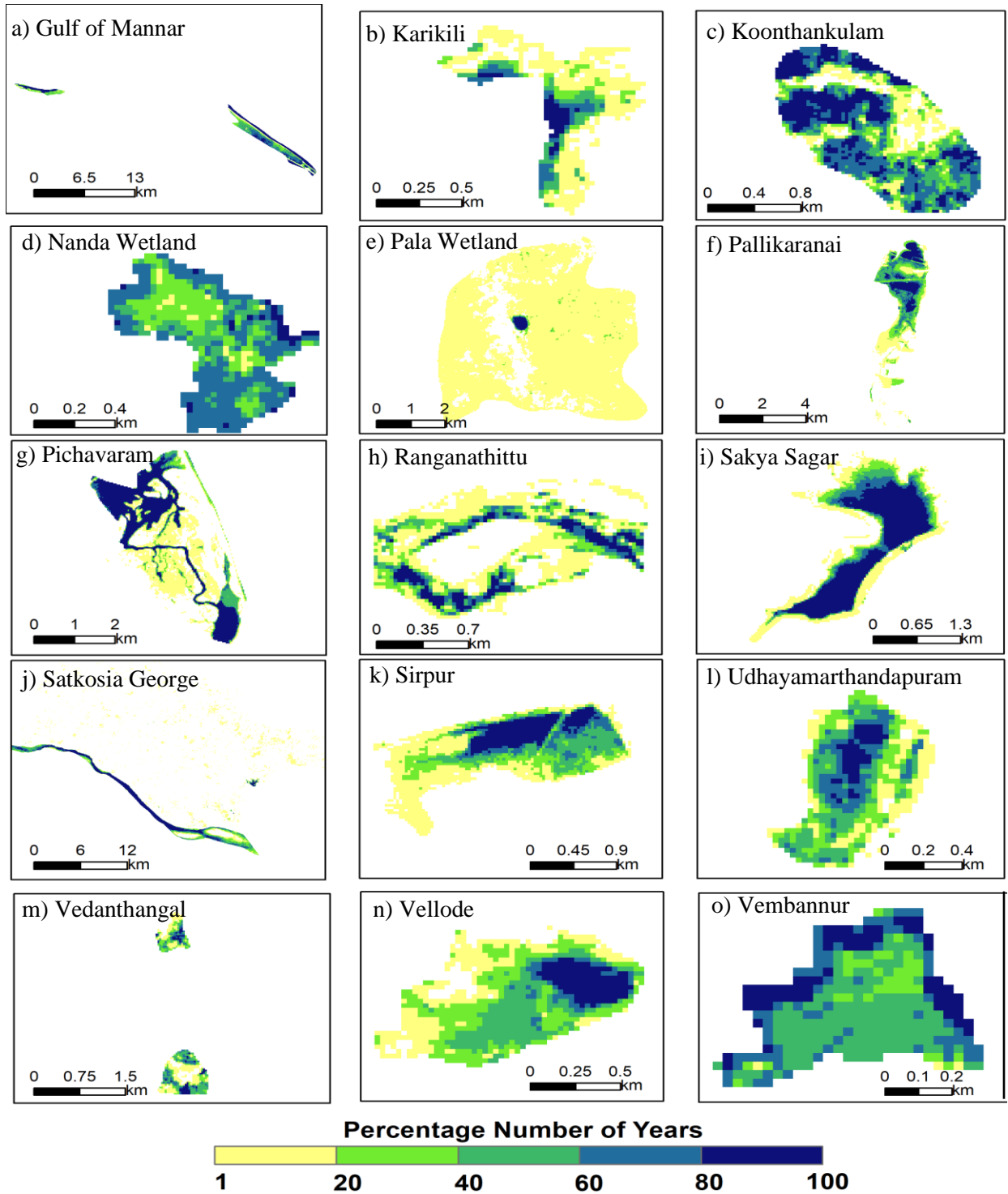
Figure 3: The inundation maps extracted using the GEE of all the 15 sites analysed in the study.
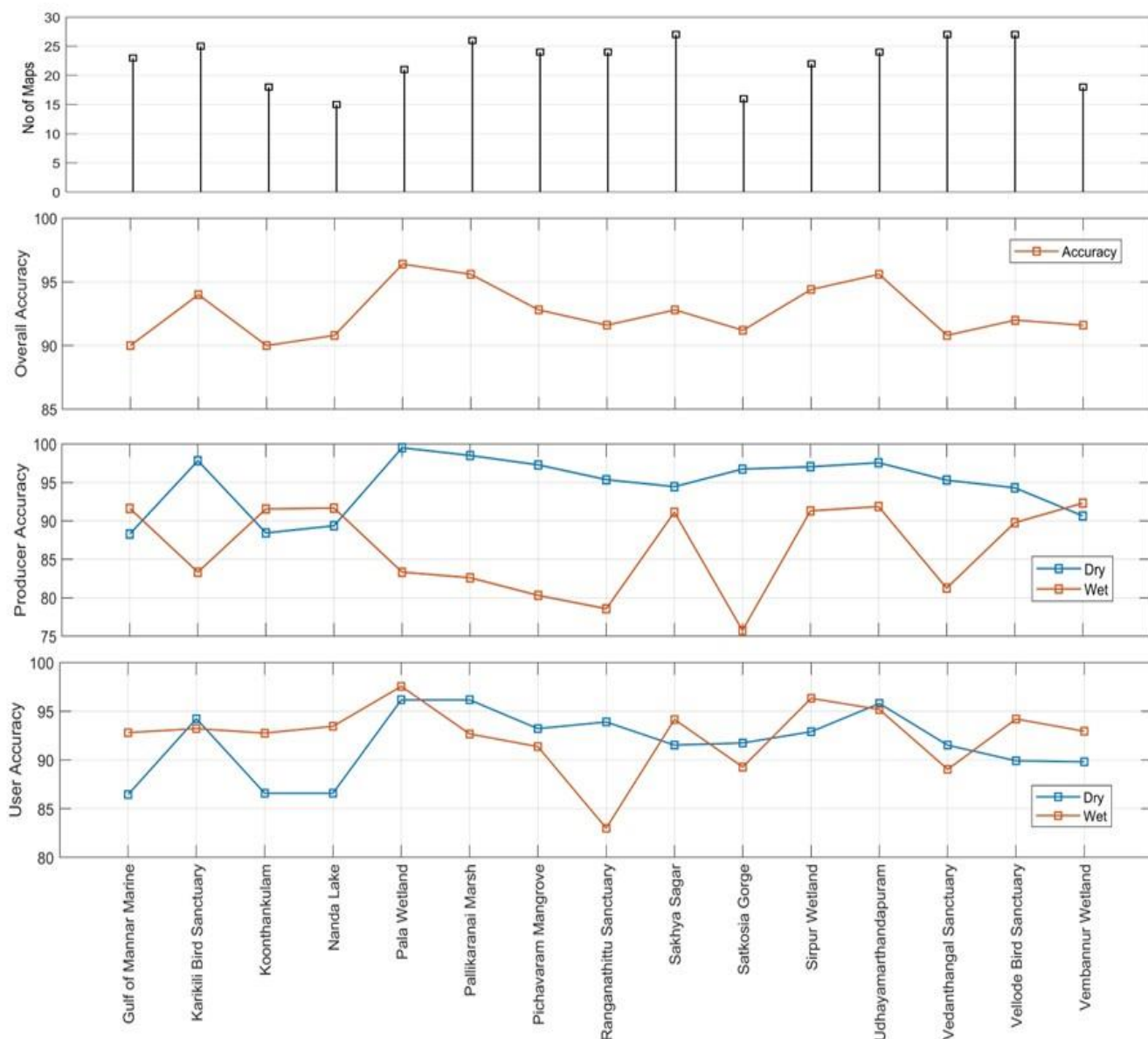
Figure 4: Figure showing Number of maps generated with the years when the data was available to generate the inundation during the period 1991-2020, Overall accuracy, Producer's accuracy, and User's accuracy obtained for each site during the validation of the inundation maps

from CMIP6, like the max temperature, min temperature, and average precipitation corresponding to each year and site respectively. The total area and the site id have also been added to the file to make the training data. Finally training data with 337 rows having columns i.e., Site id, total area, Min temperature, Max temperature, average precipitation, and year as independent variables and

the inundated area as the dependent variable that has to be predicted. Now the total data has been split into training and testing sets randomly in the ratio of 4:1 ratio respectively. Now different Machine learning Regression models have been used. The models have been fit on the training set and predictions were made on the testing set. Different models gave different error values. The models were evaluated based on Mean Squared Error, Mean Absolute Error, Root Mean Squared Error, and R2 score. Different values have been obtained for different evaluation metrics for different models. Prediction accuracy varied from model to model.

The dependence or correlation of the dependent variables i.e., Site id, total area, Min temperature, Max temperature, average precipitation, and year on the independent variable i.e., inundated area has been checked. The relation is shown in figure 5. The figure shows us that the inundated area depends mainly on-site area i.e., the total area of the site followed by max temperature, year, min temperature, and average temperature respectively. It depends least on the site id. So, some models were trained and predicted by including and excluding the site id and were checked for accuracy. The values for different evaluation metrics for different models are presented in table 2.

Table 2: The values of different evaluation metrics for different models used in the study. Here when the dependent variables are 5, it indicates that the site id is dropped from the parameters for the model.

Model	Dependent variables	RMSE	MAE	MSE	R2_score
Linear Regression	6	0.90433	0.8044	2.269	1.5063
Ridge Regression	6	0.73773	1.69542	6.22065	2.49412
Lasso Regression	6	0.86005	1.27525	3.31941	1.82192
Random Forest	6	0.94931	0.42471	1.20223	1.09646
Random Forest	5	0.97701	0.38795	0.77135	0.87825
XGB Regressor	6	0.96592	0.45201	1.14351	1.06935
XGB Regressor	5	0.94327	0.57364	1.90343	1.37965
Linear SVM	6	0.85651	1.58131	4.81457	2.19421
Linear SVM	5	0.91191	0.96134	2.9555	1.71917
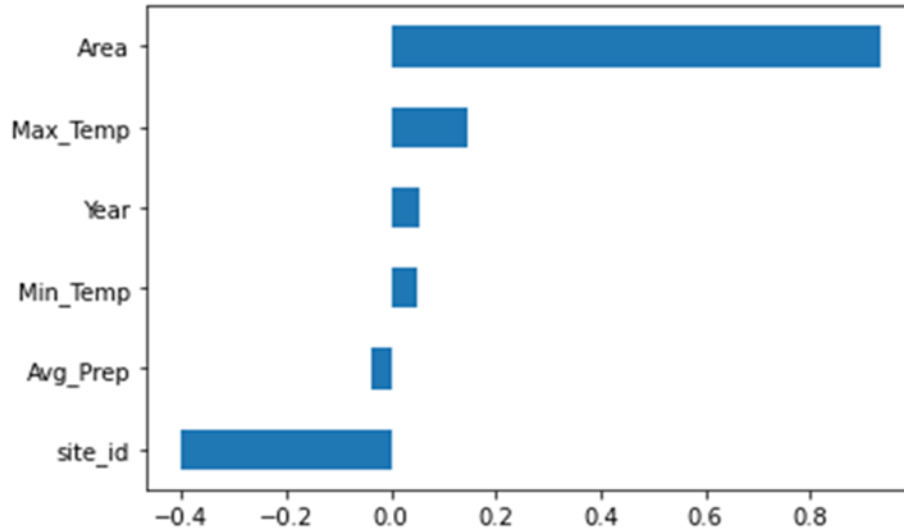
Figure 5: Graph representing the corelation of different parameters with the inundation area of site.

An observation is made in the prediction that when the number of parameters decreased to 5 I.e. when site id is dropped, the RMSE value of Random forest Regressor increased from 0.94931 to 0.97701, XGB Regressor got decreased from 0.96592 to 0.94327 and Linear SVM increased from 0.85651 to 0.91191. The Performance of different model for different evaluation metric id shown in figure 6. So, from the Error values obtained for different evaluation metrics for different models, we can see that Random Forest Regressor and XGB Regressor perform well in the prediction of inundation areas. One way to increase the accuracy in the prediction of these models is trying out Hyper Parameter Tuning. Hyper Parameter Tuning is performed on both XGB Regressor and Random Forest Regressor.

While the RMSE of Random Forest Regressor increased from 0.9493 to 0.97846, the RMSE of XGB Regressor barely got increased from 0.9659 to 0.9672. The comparision is shown in figure 7. Also, Random Forest Regressor gave the highest RMSE of 0.97, which indicates that it is the best model for the prediction of the inundation area of the wetlands.

Figure 6 : Graphs showing the comparison of Errors between different models on different evaluation metrics a) Comparision of Root Mean Squared Error b) Comparision of Mean Squared Error   c) Comparision of Mean Absolute Error   d) Comparision of R2 score e) Cumulative Comparision of all models.

Figure 7: Graph comparing the RMSE of both Random Forest Regressor and XGB Regressor before and after Hyper Parameter Tuning.

3.3 Trend in the inundation area of Ramsar sites

The predictions have been made using the Random Forst Regressor and have been saved. The Mann-Kendall (MK) test was performed on each site individually for the past and future trend. The trend analysis on past inundation data (1991 to 2020) revealed that 8 sites follow an increasing trend, with a positive MK test statistical value (Zc). 5 of the 8 sites (Karikili Bird Sanctuary, Koonthankulam, Pichavaram Mangrove, Vedanthangal Sanctuary and Vembannur Wetland) were found to be significantly increasing (Figure 6), with MK test statistical value (Zc) more than +1.96. With a negative MK test statistical value (Zc), 7 sites were found to be decreasing (Figure 6). Of these 7 sites, 3 sites (Ranganathittu Sanctuary, Vellode Bird Sanctuary and Udhayamarthandapuram Bird Sanctuary) were found to be significantly decreasing (Figure 6) with MK test statistical value (Zc) less than -1.96.

Similarly, Mann-Kendall (MK) test was performed on each site on the predicted future data. The trend analysis on predicted inundation data (2021 to 2060) revealed that 7 sites follow an increasing trend, with a positive MK test statistical value (Zc). 3 of the 7 sites (Gulf of Mannar Marine, Sirpur Wetland and Vedanthangal Sanctuary) were found to be significantly increasing. With a negative MK test statistical value (Zc), 8 sites were found to be decreasing (Figure 6). Of

these 8 sites, 2 sites (Pallikaranai Marsh, Pichavaram Mangrove) were found to be significantly decreasing. The past and future trends of the 15 sites analysed in the study are shown in figure 8.
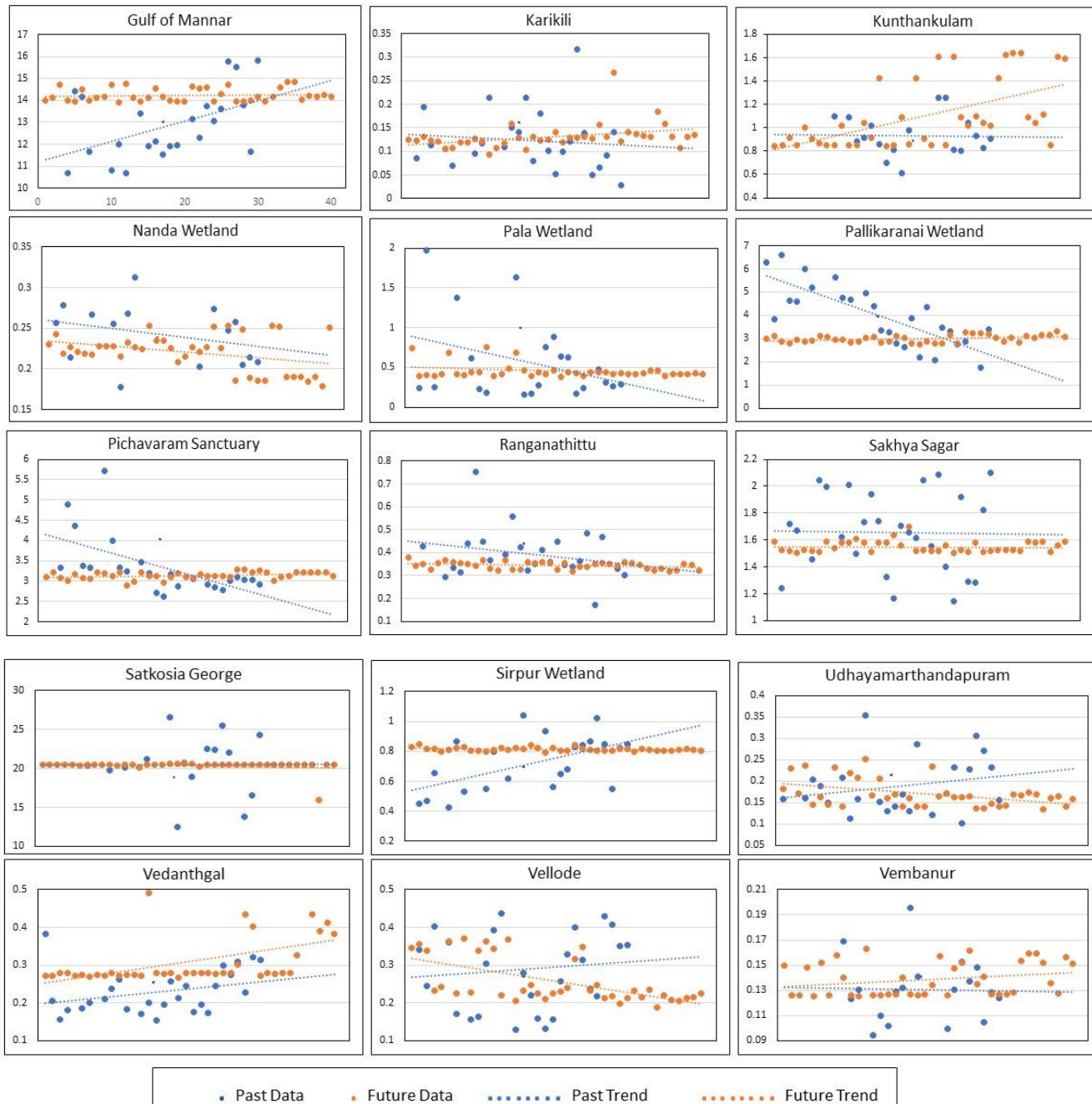
Figure 8: Graphs showing the past and future trends of the 15 sites analysed in the study. The blue line indicates the past trend whereas the red line indicates the future trend of each site.

There are only 3 sites namely Nanda Lake, Pala Wetland, Vedanthangal Sanctuary which did not show any trend change from past to future. Gulf of Mannar Marine and Sirpur Wetland which showed increasing and decreasing trend respectively in past showed significantly increasing trend in future. Two sites, Karikili Bird Sanctuary, Koonthankulam Bird Sanctuary found to be decreasing which were significantly increasing in past. Pallikaranai Marsh and Pichavaram Mangrove showed a significantly decreasing trend in future which were found to be increasing and significantly increasing respectively in the past. Ranganathittu Sanctuary and Sakhya Sagar are found to be decreasing in the future which were significantly decreasing and increasing respectively in the past. Udhayamarthandapuram Bird Sanctuary and Vellode Bird Sanctuary which were significantly decreasing in past showed an increasing trend in future. Satkosia George and Vembannur Wetland showed an increasing trend which were found to be decreasing and significantly decreasing in the past.

Some interesting patterns were found when we compared the inundation trends of the sites were compared with the trends of min temperature, max temperature and average precipitation. For the past data, most of the sites which showed decreasing and significantly decreasing trend for the inundation pattern also showed increasing or significantly increasing in the past max and min temperature. Also, all the decreasing sites in the past showed a decreasing trend in the average precipitation. This indicates us that the increase in the temperature and decrease in the precipitation may be a reason for their decrease in the inundation area of the wetlands. Also, most of the increasing sites showed increasing trend in precipitation and decreasing trend in minimum and maximum temperature.

Similarly, for the future data, the two significantly decreasing sites showed a decreasing trend in the average precipitation and significantly increasing trend in the min and max temperature. Most of the decreasing sites showed decreasing trend in the average precipitation. Also, most of the increasing sites showed the same trend as the past increasing wetlands for temperature and precipitation. The pattern in the trends indicate us that most of the sites for which inundation trend is decreasing shows decreasing trend in precipitation and an increasing trend in min and max temperature. Also, the sites which has increasing inundation trend shows increasing and decreasing trend in precipitation and temperature data respectively.

The major source of water for the wetlands is precipitation and inflows which ensures the source of water to wetlands. There could be many reasons for the sites to be significantly decreasing. Some of them could be an increase in the temperature which increases the evaporation from the wetland, decrease in the precipitation over time which is the main source of water inflow to the wetland. The huge reduction in water content could be linked to groundwater depletion. An increase in the human population around the wetlands could also be a reason as it increases the usage of water for drinking and domestic purposes and for irrigation purposes.

4 Conclusion

The inundation data that is required for the study has been extracted from the landsat images using the Google Earth Engine and the Climate Data i.e., Minimun temperature, Maximun Temperature and Average precipitation has been extracted from CMIP6. The inundation data from 1991 to 2020 has been merged with the climate data and area of wetland for the training of the model. The model that has given high accuracy i.e., Random Forest Regressor is trained on the training data(1991 to 2020) which is extracted forn google earth engine and CMIP6. For the prediction of future trends, future projections of the climate data(2021 to 2060) from CMIP6 is merged with area of wetland. Using the trained model, Inundation area forecasting has to be done for all the sites for the future i.e., from 2021 to 2060. To identify trends in the variation of inundation extent, Mann-Kendall (MK) test was performed on each site individually. The test has been performed on the past and future data individually for all the sites to understand their past and future trend of inundation areas. Mann-Kendall (MK) test has also been done on the climate data to understnd the trends in them for the past and future. Trends in the climate data and the inundation patters has been analysed for pattrens.

Each wetland showed different trend which can be attributed to different factors. The pattern in the trends indicated that most of the sites for which inundation trend is decreasing shows decreasing trend in precipitation and an increasing trend in min and max temperature. Also, the sites which has increasing inundation trend shows increasing and decreasing trend in precipitation and temperature data respectively. It is also observed that there are 7 decreasing sites in the past out of which 3 were significantly decreasing. But in the future, there are 8 wetlands that showed decreasing trend out of which 2 are significantly decreasing. The sites that are significantly decreasing in the past got improved and showed decreasing or increasing trend in the future whereas 2 wetlands which showed decreasing trend in past, showed a significantly decreasing trend.

References

Amani, M., Mahdavi, S., Kakooei, M., Ghorbanian, A., Brisco, B., DeLancey, E., Toure, S., Reyes, E.L., 2021. Wetland Change Analysis in Alberta, Canada Using Four Decades of Landsat Imagery. IEEE J Sel Top Appl Earth Obs Remote Sens 14, 10314–10335. https://doi.org/10.1109/JSTARS.2021.3110460

Bassi, N., Kumar, M.D., Sharma, A., Pardha-Saradhi, P., 2014. Status of wetlands in India: A review of extent, ecosystem benefits, threats and management strategies. J Hydrol Reg Stud 2, 1–19. https://doi.org/10.1016/j.ejrh.2014.07.001

Bavel, J.J. Van, Baicker, K., Boggio, P.S., Capraro, V., Cichocka, A., Cikara, M., Crockett, M.J., Crum, A.J., Douglas, K.M., Druckman, J.N., Drury, J., Dube, O., Ellemers, N., Finkel, E.J., Fowler, J.H., Gelfand, M., Han, S., Haslam, S.A., Jetten, J., Kitayama, S., Mobbs, D., Napper, L.E., Packer, D.J., Pennycook, G., Peters, E., Petty, R.E., Rand, D.G., Reicher, S.D., Schnall, S., Shariff, A., Skitka, L.J., Smith, S.S., Sunstein, C.R., Tabri, N., Tucker, J.A., Linden, S. van der, Lange, P. van, Weeden, K.A., Wohl, M.J.A., Zaki, J., Zion, S.R., Willer, R., 2020. Using social and behavioural science to support COVID-19 pandemic response. Nat Hum Behav 4, 460–471. https://doi.org/10.1038/s41562-020-0884-z

Blain, G.C., 2013. The Mann-Kendall test: the need to consider the interaction between serial correlation and trend - doi: 10.4025/actasciagron.v35i4.16006. Acta Sci Agron 35. https://doi.org/10.4025/actasciagron.v35i4.16006

Boretti, A., Rosa, L., 2019. Reassessing the projections of the World Water Development Report. NPJ Clean Water 2, 15. https://doi.org/10.1038/s41545-019-0039-9

Cook-Patton, S.C., Drever, C.R., Griscom, B.W., Hamrick, K., Hardman, H., Kroeger, T., Pacheco, P., Raghav, S., Stevenson, M., Webb, C., Yeo, S., Ellis, P.W., 2021. Protect, manage and then restore lands for climate mitigation. Nat Clim Chang 11, 1027–1034. https://doi.org/10.1038/s41558-021-01198-0

Corcoran, J., Knight, J., Gallant, A., 2013. Influence of Multi-Source and Multi-Temporal Remotely Sensed and Ancillary Data on the Accuracy of Random Forest Classification of Wetlands in Northern Minnesota. Remote Sens (Basel) 5, 3212–3238. https://doi.org/10.3390/rs5073212

Creed, I.F., Lane, C.R., Serran, J.N., Alexander, L.C., Basu, N.B., Calhoun, A.J.K., Christensen, J.R., Cohen, M.J., Craft, C., D'Amico, E., DeKeyser, E., Fowler, L., Golden, H.E., Jawitz, J.W., Kalla, P., Kirkman, L.K., Lang, M., Leibowitz, S.G., Lewis, D.B., Marton, J., McLaughlin, D.L., Raanan-Kiperwas, H., Rains, M.C., Rains, K.C., Smith, L., 2017. Enhancing protection for vulnerable waters. Nat Geosci 10, 809–815. https://doi.org/10.1038/ngeo3041

Gallant, A., 2015. The Challenges of Remote Monitoring of Wetlands. Remote Sens (Basel) 7, 10938–10950. https://doi.org/10.3390/rs70810938

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens Environ 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031

Gouda, K.C., Nahak, S., Goswami, P., 2020. Deterministic Seasonal Quantitative Precipitation Forecasts: Benchmark Skill with a GCM. Pure Appl Geophys 177, 4443–4456. https://doi.org/10.1007/s00024-020-02463-7

Gumbricht, T., Wolski, P., Frost, P., McCarthy, T.S., 2004. Forecasting the spatial extent of the annual flood in the Okavango delta, Botswana. J Hydrol (Amst) 290, 178–191. https://doi.org/10.1016/j.jhydrol.2003.11.010

Hird, J., DeLancey, E., McDermid, G., Kariyeva, J., 2017. Google Earth Engine, Open-Access Satellite Data, and Machine Learning in Support of Large-Area Probabilistic Wetland Mapping. Remote Sens (Basel) 9, 1315. https://doi.org/10.3390/rs9121315

IMD, 2020. Observed Monsoon Rainfall Variability and Changes during Recent 30 years (1989-2018), Government of India, Ministry of Earth Sciences (MoES), Climate Research and Services (CRS) Division , Pune Observed Monsoon Rainfall Variability and Changes during Rece 1–10.

Inman, V.L., Lyons, M.B., 2020. Automated Inundation Mapping Over Large Areas Using Landsat Data and Google Earth Engine. Remote Sens (Basel) 12, 1348. https://doi.org/10.3390/rs12081348

Kumar, V., Jain, S.K., Singh, Y., 2010. Analysis of long-term rainfall trends in India. Hydrological Sciences Journal 55, 484–496. https://doi.org/10.1080/02626667.2010.481373

Mann, H.B., 1945. Nonparametric Tests Against Trend. Econometrica 13, 245. https://doi.org/10.2307/1907187

Maurya, R.K.S., Singh, G.P., 2016. Simulation of present-day precipitation over India using a regional climate model. Meteorology and Atmospheric Physics 128, 211–228. https://doi.org/10.1007/s00703-015-0409-x

Milzow, C., Kgotlhang, L., Kinzelbach, W., Meier, P., Bauer-Gottwein, P., 2009. The role of remote sensing in hydrological modelling of the Okavango Delta, Botswana. J Environ Manage 90, 2252–2260. https://doi.org/10.1016/j.jenvman.2007.06.032

Murray-Hudson, M., Wolski, P., Cassidy, L., Brown, M.T., Thito, K., Kashe, K., Mosimanyana, E., 2015. Remote Sensing-derived hydroperiod as a predictor of floodplain vegetation composition. Wetl Ecol Manag 23, 603–616. https://doi.org/10.1007/s11273-014-9340-z

Pal, I., Al-Tabbaa, A., 2009. Trends in seasonal precipitation extremes – An indicator of 'climate change' in Kerala, India. J Hydrol (Amst) 367, 62–69. https://doi.org/10.1016/j.jhydrol.2008.12.025

Pereira, J.M., Basto, M., Silva, A.F. da, 2016. The Logistic Lasso and Ridge Regression in Predicting Corporate Failure. Procedia Economics and Finance 39, 634–641. https://doi.org/10.1016/S2212-5671(16)30310-0

Pisner, D.A., Schnyer, D.M., 2020. Support vector machine, in: Machine Learning. Elsevier, pp. 101–121. https://doi.org/10.1016/B978-0-12-815739-8.00006-7

Scheyer, T.M., Aguilera, O.A., Delfino, M., Fortier, D.C., Carlini, A.A., Sánchez, R., Carrillo-Briceño, J.D., Quiroz, L., Sánchez-Villagra, M.R., 2013. Crocodylian diversity peak and extinction in the late Cenozoic of the northern Neotropics. Nat Commun 4, 1907. https://doi.org/10.1038/ncomms2940

Shehadeh, A., Alshboul, O., al Mamlook, R.E., Hamedat, O., 2021. Machine learning models for predicting the residual value of heavy construction equipment: An evaluation of modified decision tree, LightGBM, and XGBoost regression. Autom Constr 129, 103827. https://doi.org/10.1016/j.autcon.2021.103827

Shirzaei, M., Bürgmann, R., 2018. Global climate change and local land subsidence exacerbate inundation risk to the San Francisco Bay Area. Sci Adv 4. https://doi.org/10.1126/sciadv.aap9234

Su, X., Yan, X., Tsai, C.-L., 2012. Linear regression. Wiley Interdiscip Rev Comput Stat 4, 275–294. https://doi.org/10.1002/wics.1198

Svetnik, V., Liaw, A., Tong, C., Culberson, J.C., Sheridan, R.P., Feuston, B.P., 2003. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. J Chem Inf Comput Sci 43, 1947–1958. https://doi.org/10.1021/ci034160g

Temmink, R.J.M., Lamers, L.P.M., Angelini, C., Bouma, T.J., Fritz, C., van de Koppel, J., Lexmond, R., Rietkerk, M., Silliman, B.R., Joosten, H., van der Heide, T., 2022. Recovering wetland biogeomorphic feedbacks to restore the world's biotic carbon hotspots. Science (1979) 376. https://doi.org/10.1126/science.abn1479

Valenti, V.L., Carcelen, E.C., Lange, K., Russo, N.J., Chapman, B., 2020. Leveraging Google Earth Engine User Interface for Semiautomated Wetland Classification in the Great Lakes Basin at 10 m With Optical and Radar Geospatial Datasets. IEEE J Sel Top Appl Earth Obs Remote Sens 13, 6008–6018. https://doi.org/10.1109/JSTARS.2020.3023901

Valuing wetlands, 2021. . Nat Geosci 14, 111–111. https://doi.org/10.1038/s41561-021-00713-4

Wolski, P., Murray-Hudson, M., Thito, K., Cassidy, L., 2017. Keeping it simple: Monitoring flood extent in large data-poor wetlands using MODIS SWIR data. International Journal of Applied Earth Observation and Geoinformation 57, 224–234. https://doi.org/10.1016/j.jag.2017.01.005

Xi, Y., Peng, S., Ciais, P., Chen, Y., 2021. Future impacts of climate change on inland Ramsar wetlands. Nat Clim Chang 11, 45–51. https://doi.org/10.1038/s41558-020-00942-2

Yin, G., Mariethoz, G., McCabe, M., 2016. Gap-Filling of Landsat 7 Imagery Using the Direct Sampling Method. Remote Sens (Basel) 9, 12. https://doi.org/10.3390/rs9010012

Zhang, Z., Chen, J., Mao, Y., 2022. Ridge regression and lasso regression based least squares algorithm for a time-delayed rational model via redundant rule. Int J Model Identif Control 40, 11. https://doi.org/10.1504/IJMIC.2022.124075

Zurqani, H.A., Post, C.J., Mikhailova, E.A., Cope, M.P., Allen, J.S., Lytle, B.A., 2020. Evaluating the integrity of forested riparian buffers over a large area using LiDAR data and Google Earth Engine. Sci Rep 10, 14096. https://doi.org/10.1038/s41598-020-69743-z