# What Wets the Wetlands?

Reconstructing the Mara Wetland surface water dynamics through coupling satellite derived inundation patterns with hydrological field data

Juliette Marie Kool

# **WHAT WETS THE WETLANDS?** RECONSTRUCTING THE MARA WETLAND SURFACE-WATER DYNAMICS THROUGH COUPLING SATELLITE-DERIVED INUNDATION PATTERNS WITH HYDROLOGICAL FIELD DATA

A thesis submitted to the Delft University of Technology in partial fulfillment
of the requirements for the degree of

Master of Science in Civil Engineering

by

Juliette Marie Kool

March 2021

The cover of this report is a birds-eye-view of the South-West boundary of the Mara Wetland. The original high-resolution satellite image, derived from from ESRI World Imagery, is cropped and edited with the help of Joost Wendling.

*What Wets the Wetlands?*
*Reconstructing the Mara Wetland surface-water dynamics through coupling satellite-derived inundation patterns with hydrological field data* (2021)


TU Delft

Department of Water Resources
Faculty Civil Engineering & Geosciences
Delft University of Technology

Student number: 4343395
Supervisors: Prof. Michael McClain
Dr. Stef Lhermitte
Dr. Markus Hrachowitz

# ABSTRACT

The Mara Wetland in Tanzania has an important role in regulating the quality, timing and magnitude of the flow of water into Lake Victoria. In addition, the wetland provides natural resources for local communities and habitat for variety of species. The planned dam construction upstream of the wetland and projected changes in the local climate could affect the physical and ecological equilibrium of the system. Baseline information on seasonal inundation dynamics is necessary to sustainably manage these potential threats. The wetland is sparsely instrumented, which has hampered a thorough temporal and spatial understanding of the local water balance. In addition, the highly vegetated nature of the wetland, and relatively frequent cloud-coverage, motivates multi-source integration of remotely sensed data to capture flood patterns at a high resolution.

In this study, the spatiotemporal inundation pattern of the Mara Wetland in Tanzania is reconstructed using optical remote sensing data. The annual fluctuations in aerial wetland extent are analysed in parallel to the fluctuations of local water balance components: downstream water level of Lake Victoria, upstream discharge, direct precipitation and evaporation. The analyses aims to shed light on the underlying mechanisms and hydrological processes that control the hydric status of the wetland. Comparing the temporal changes in extent with surrounding physical processes provides insight on the responsiveness of the wetland to specific water balance components.

The intra- and inter-annual trends in inundation of the Mara Wetland are reproduced for the years 2017, 2018, 2019. The Random Forests (RF) algorithm is trained bi-seasonally (using bands and derived water and vegetation indices from Sentinel-2 data and a Digital Elevation Model (DEM) as input variables), and used to classify the land-covers of the wetland region in a semi-automated way for a total of 73 Sentinel-2 scenes. The scenes are classified into 7 individual land-cover classes; 3 wetland classes (open water, flooded vegetation, wet floodplain) and 4 dryland classes (dry floodplain, wet agriculture, dry agriculture, bare land). The overall classification accuracy achieved (based on an independent validation set, not used to train the classification algorithm) is 98.6 %. The spatiotemporal variability of the inundated area is used in combination with available hydrological field-data to reproduce the local water balance.

The seasonal expansion and contraction of the wetland follows a consistent bi-modal regime, and the results from the waterbalance affirm the importance of local precipitation in the seasonal expansion and contraction of the wetland. The base-flow supplied by the Mara River, together with the backwater from Lake Victoria appear to be at equilibrium at the extent of the permanent swamp during the dry season, insinuating the importance of the riverflow during these low-rainfall months. The occasional yet extreme flood events induced by high discharge rates are expected to play a specific ecological role in the wetland, and should be accounted for during future dam operations upstream.

# ACKNOWLEDGEMENTS

With this MSc thesis I officially conclude my studies in Delft. I would have never imagined to be writing my thesis during a pandemic, and looking back now I felt this timing has been both a blessing and a curse. The domino effect of cancelled plans and growing amount of limitations has left more time to think and work, but at the same time, the lack of a thrill and absence of social events was dispiriting when in need of distraction. In spite of everything, I have learnt a tremendous amount about life, friendship, emotions, communication and much more. An important takeaway from what I've learnt is expressed well in the following quote by Horace:

*"Mix a little foolishness with your serious plans. It is lovely to be silly at the right moment."* – Horace

To my main supervisor Michael McClain, thank you for always having confidence in my work and allowing me to be free to steer the project as I desired. To Stef Lhermitte, thank you for your moral support throughout the project and being there to answer any remote sensing-related questions. To Markus Hrachowitz, despite only meeting a few times, thank you for bringing a new perspective and for always taking a critical look at my work. I would like to thank Francesco Bregoli for his time, energy and openness to share his research and provide feedback. The talks we've had were always enjoyable and your enthusiasm for the project motivated me to continuously dedicate myself into improving my work.

Then to those closer to me. I've been privileged to be part of a family that is loving, open and supportive in every way. I am so grateful to my mother and father for letting me experience the world and for their "komt altijd goed" motto at times when it didn't feel that way. And even now, for *always* being there for me, nothing that is too crazy to ask for help with. Being the last child of four that graduates feels like the end of a chapter. To my parents; don't worry, I will keep coming by (probably too often). Then to my brother and sisters and to Juan and Farah, I am so happy to have you close and can't wait to see where the future brings us all.

Lastly to my people in Rotterdam. My roommates Floor, Celia, Marleen and Vera; I am so grateful to have spent this time with such supportive, open-hearted and strong individuals. Even after countless hours of nothingness during the days of lockdown, I've never felt a slight bit of boredom in our little palace. To Roos and Jente, thank you for the fun times and always being there to give advice, I'm very lucky to have friends like you and am very excited to continue sharing experiences with you both. Fien, you're a star, thank you for being such a good listener and great friend to me always. And last, but definitely not least, to my sister Dorith and her roommates Douwe, Joost, Max and Simon; thank you for providing a second home, letting me use your kitchen, fulfilling my dance deprivations and always making me laugh, even writing this has put a smile on my face.

# CONTENTS


1 INTRODUCTION 2
1.1 Research motive . . . 2
1.2 Problem statement and knowledge gaps . . . 3
1.3 Research objectives and scope . . . 3
1.4 Reader's guide . . . 4
2 THEORETICAL BACKGROUND 5
2.1 Wetland concepts . . . 5
2.1.1 Boundary definition . . . 5
2.1.2 Hydro-period and hydro-pattern . . . 6
2.1.3 Ecosystem services . . . 6
2.2 Physical principles of remote sensing . . . 7
2.2.1 Passive optical remote sensing . . . 7
2.2.2 Multispectral imagery . . . 8
2.2.3 Bands and indices . . . 8
2.2.4 Spectral, spatial and temporal resolution . . . 8
2.3 Image classification . . . 9
2.3.1 Random Forests (RF) algorithm . . . 9
2.4 Multitemporal remote sensing approaches for wetland mapping . . . 10
2.4.1 Optical remote sensing of wetlands . . . 10
2.4.2 Microwave remote sensing of wetlands . . . 10
2.4.3 Combined approaches . . . 11
3 STUDY AREA 12
3.1 General information . . . 12
3.2 Climate . . . 13
3.3 Drivers of change and future threats . . . 13
3.4 Vegetation and ecosystem functioning . . . 13
3.5 Hydrology of the Mara Wetland . . . 13
3.6 Population and contemporary livelihood . . . 14
4 DATA ACQUISITION AND PRE-PROCESSING 15
4.1 Google Earth Engine (GEE) processing platform . . . 15
4.2 Remote Sensing data . . . 15
4.2.1 Sentinel-2 (S2) . . . 15
4.2.2 Topographical data . . . 18
4.2.3 Orthophotography . . . 18
4.3 Hydro-meteorological data . . . 18
4.3.1 Discharge . . . 19
4.3.2 Lake Victoria (LV) stage levels . . . 19
4.3.3 Precipitation . . . 20
4.3.4 Evapotranspiration (ET) . . . 20
5 METHODOLOGY 21
5.1 Overall research structure . . . 21
5.2 Phase-I: Multitemporal image classification . . . 21
5.2.1 Land-cover analysis . . . 22
5.2.2 Random Forests classification . . . 26
5.2.3 Post-processing of classified imagery . . . 28
5.2.4 Evaluation of area estimates . . . 29
5.3 Phase-II: Hydrometeorological analyses . . . 30
5.3.1 Stage- and discharge-surface area correlations . . . 30
5.3.2 Waterbalance calculation . . . 31
6 RESULTS 33
6.1 Phase-I: Multitemporal wetland classification results . . . 33
6.1.1 Time-series of land-cover class area estimates . . . 33
6.1.2 Classification of sampled images . . . 33
6.1.3 Land-cover class occurrence frequency . . . 35

6.1.4 Accuracy assessment . . . 35
6.1.5 Prior area estimates . . . 37
6.1.6 Antecedent water fluxes . . . 38
6.2 Phase-II: Hydro-meteorological analyses results . . . 38
7 DISCUSSION 41
7.1 Classification (Phase-I) . . . 41
7.1.1 Classification procedure . . . 41
7.1.2 Classification accuracy . . . 42
7.2 Hydrological analyses (Phase-II) . . . 42
7.2.1 Approach, assumptions and limitations . . . 42
7.2.2 Wetland dynamics . . . 44
7.2.3 Waterbalance . . . 45
7.3 Additional sources of error . . . 45
7.3.1 Uncertainty of evapotranspiration data . . . 46
7.3.2 Uncertainty of precipitation data . . . 46
8 CONCLUSION 47
8.1 General conclusions . . . 47
8.2 Implications and recommendations for the Mara Wetland . . . 48
8.3 Potential applications of the methodology . . . 48
A APPENDIX A 55
A.1 Previous wetland area estimates . . . 55
A.2 Specifications of precipitation datasets . . . 55
B APPENDIX B 57
B.1 Index composite maps . . . 57
B.2 Unsupervised classification . . . 59
B.3 Multi-temporal spectral response . . . 60
C APPENDIX C 62
C.1 Spectral response of training dataset . . . 62
C.1.1 Wet season training data . . . 62
C.1.2 Dry season training data . . . 62
C.2 Land-cover class area estimates . . . 63
D APPENDIX D 66

# LIST OF FIGURES


Figure 2.1 Wetland boundary according to area of inundation and extent of waterlogging (C. A. Semeniuk & Semeniuk, 2004) . . . 6
Figure 2.2 Linkages between ecosystem services and human well-being (Millennium Ecosystem Assessment, 2005) . . . 7
Figure 2.3 The Electromagnetic Spectrum (Physics, 2016) . . . 7
Figure 2.4 Typical spectral signature of grassland pixels. Edited from (HSU, 2019) . . . 8
Figure 2.5 Example of Random Forests prediction model . . . 9
Figure 3.1 Map of the location of Lake Victoria (blue) and Mara River Basin (green) and the outline of the wetland region (red box) . . . 12
Figure 4.1 Illustration of a false color composites and the outline of the two the Sentinel-2 satellite image tiles 36MWD (blue) and 36MXD (yellow) and wetland delineation as set by (Bregoli et al., 2019) (black) . . . 16
Figure 4.2 Cloud mask coverage . . . 17
Figure 4.3 Digital Elevation Model (DEM) of wetland area . . . 18
Figure 4.4 Example of an ortho-photograph acquired on 02-11-2017 overlaying two satellite image scenes from 28-12-2018 (left) and 02-04-2019 (right) . . . 18
Figure 4.5 Discharge at Mara Mine for 2017-2020 . . . 19
Figure 4.6 Location of Mara Mine gauging station . . . 19
Figure 4.7 LV stage levels for the years 2017, 2018, 2019 . . . 19
Figure 4.8 Correlation chart showing the water levels measured at Kirumi Bridge against those measured in Mwanza. Data from the water level at Kirumi Bridge is provided by F. Bregoli (Bregoli et al., 2019) . . . 20
Figure 4.9 Regional mean monthly accumulated precipitation estimate from TRMM, PERSIANN-CDR, CHIRPS, GPM, GSMaP and ERA5 for the years 2017, 2018, 2019 . . . 20
Figure 5.1 Flow chart of the overall research structure . . . 21
Figure 5.2 Flow chart of the steps carried out for the land-cover class analysis . . . 22
Figure 5.3 Google satellite images of land-classes used as a reference for collecting training data. The top and bottom half of figures (c) and (d) are the labeled class in wet and dry season, respectively . . . 24
Figure 5.4 Spectral signature of land-cover classes sampled in June (left) and November (right) . . . 24
Figure 5.5 Spectral (index) response for land-cover classes sampled in June (left) and November (right) . . . 26
Figure 5.6 Flow chart of the steps carried out to collect training data and classify image collection . . . 26
Figure 5.7 Contour map of the wetland area. The black region contains the wetland and the remaining part of the image is masked as upland areas. . . . 27
Figure 5.8 Post-classification steps and accuracy assessment . . . 28
Figure 5.9 Main papyrus swamp delineation . . . 28
Figure 5.10 Confusion matrix . . . 29
Figure 5.11 The delineation of the total wetland area (dark green) and permanent swamp (light green) . . . 31
Figure 5.12 Time-series showing monthly accumulated evapotranspiration for the whole wetland area (dark green), the flooded vegetation swamp (light green) and the average of these (red) . . . 32
Figure 6.1 Time-series stacked area plot of the land-cover classes resulting from the classification procedure . . . 33
Figure 6.2 Link to a video showing the spatiotemporal dynamics of the land-covers of the Mara Wetland for all 73 classified scenes. Access by clicking or scanning the QR code. . . . 33
Figure 6.3 Classified map for June training image (11-06-2020) . . . 34
Figure 6.4 Classified map for November training image (03-11-2020) . . . 34
Figure 6.5 Open water covered area frequency map for 2017-2020 . . . 35

Figure 6.6 Flooded vegetation covered area frequency map for 2017-2020 . . . . . 35
Figure 6.7 Wet floodplain frequency map for 2017-2020 . . . . . 35
Figure 6.8 Confusion matrix representing re-substitution accuracy (training accuracy). OW = Open water, FV1 = Flooded vegetation 1, FV2 = Flooded vegetation 2, BA = Bare, AW = Agriculture wet, WF = Wet floodplain , AD = Agriculture dry, DF = Dry floodplain, PA = Producer's accuracy, UA = User's accuracy . . . . . 36
Figure 6.9 Confusion matrix for classification of independent validation data (testing accuracy). OW = Open water, FV1 = Flooded vegetation 1, FV2 = Flooded vegetation 2, B = Bare, AW = Agriculture wet, WF = Wet floodplain, AD = Agriculture dry, DF = Dry floodplain, PA = Producer's accuracy, UA = User's accuracy . . . . . 37
Figure 6.10 Relationships between the peak inundated area (2/year for the years 2017, 2018, 2019) and change in local water fluxes over the 3 months preceding the peak date. (a) cumulative precipitation (average between CHIRPS and PERSIANN-CDR and error indicates range between the data-sets), (b) cumulative discharge amount and (c) difference in Lake Victoria water level between the peak-flooding date and that of 3 months prior . . . . . 38
Figure 6.11 Correlation chart showing Lake Victoria water levels against inundated area . . 38
Figure 6.12 Discharge (2 days before date of area estimate) and estimated inundated area . 39
Figure 6.13 Discharge (2 days before date of area estimate) and estimated open water area 39
Figure 6.14 Monthly water balance of the wetland based on the net precipitation, discharge, and evaporative fluxes accumulated over one month. The uncertainty range is computed as the sum of the uncertainty from the precipitation and evaporation estimates. The vertical lines correspond to the dates of peak inundated area, estimated from the classification procedure. The red line is the water availability based on the assumption the outflow from the wetland into LV is 30% of the inflow from the Mara River (based on the estimation made by Bregoli et al., 2019). 40
Figure 7.1 Time-series chart of wetland area and waterbalance components. The grey lines correspond to the inundated area peaks. . . . . 43
Figure B.1 NDMI visualisation for 11-06-2017, adjusted with a new minimum and maximum 57
Figure B.2 NDMI visualisation for 03-11-2017, adjusted with a new minimum and maximum 57
Figure B.3 NDVI composite for 11-06-2017 from S2 image collection . . . . . 58
Figure B.4 NDVI composite for 03-11-2017 from S2 image collection . . . . . 58
Figure B.5 MNDWI composite for 11-06-2017 from S2 image collection . . . . . 58
Figure B.6 MNDWI composite for 03-11-2017 from S2 image collection . . . . . 58
Figure B.7 NDII composite for 11-06-2017 from S2 image collection . . . . . 59
Figure B.8 NDII composite for 03-11-2017 from S2 image collection . . . . . 59
Figure B.9 Unsupervised K-means clustering classification for 06/2017 . . . . . 60
Figure B.10 Unsupervised K-means clustering classification for 11/2017 . . . . . 60
Figure B.11 MNDWI training data timeseries . . . . . 61
Figure B.12 NDVI training data timeseries . . . . . 61
Figure D.1 Correlation chart showing Lake Victoria water levels against open water area . 66
Figure D.2 Correlation chart showing Lake Victoria water levels against open water and flooded vegetation area . . . . . 66

# LIST OF TABLES


Table 4.1 Sentinel-2 MSI satellite sensor spectral bands . . . 16
Table 4.2 Spectral indices used as input variables for the Random Forests classifier . . . 17
Table 5.1 Land-cover classes (from literature) . . . 22
Table 5.2 Image dates used for land-cover analysis . . . 23
Table 5.3 Land cover classes based on Google Satellite and UAV Imagery . . . 25
Table 5.4 Map and training data proportions . . . 27
Table 5.5 Land-cover class sampling date . . . 27
Table 6.1 Wetland class area statistics . . . 34
Table 6.2 Overall accuracy and Kappa coefficient for training data . . . 36
Table 6.3 Overall accuracy and Kappa coefficient for validation data . . . 37
Table A.1 Reported wetland area estimates from various studies . . . 55
Table A.2 Precipitation datasets specifications . . . 56
Table C.1 Median band and index response for training data samples obtained in the wet season . . . 62
Table C.2 Median band and index response for training data samples obtained in the dry season . . . 63
Table C.3 Land-cover classification area estimates (in km²) for all 73 Sentinel-2 scenes . . . 63

# ACRONYMS

| | |
|---|---|
| BOA | Bottom of Atmosphere |
| CDR | Climate Data Record |
| CHIRPS | Climate Hazards Group InfraRed Precipitation with Station data |
| CRS | Coordinate Reference System |
| DEM | Digital Elevation Model |
| ET | Evapotranspiration |
| GEE | Google Earth Engine |
| ITCZ | Intertropical Convergence Zone |
| LV | Lake Victoria |
| LiDAR | Light Detection and Ranging |
| LVBWB | Lake Victoria Basin Water Board |
| MA | Millennium Ecosystem Assessment |
| MM | Mara Mines |
| MNDWI | Modified Normalized Difference Water Index |
| MODIS | Moderate Resolution Imaging Spectroradiometer |
| MRB | Mara River Basin |
| MSI | MultiSpectral Instrument |
| NDII | Normalized Difference Infrared Index |
| NDMI | Normalized Difference Moisture Index |
| NDMI | Normalized Difference Moisture Index |
| NDVI | Normalized Difference Vegetation Index |
| NDWI | Normalized Difference Water Index |
| NIR | Near InfraRed |
| RADAR | Radio Detection and Ranging |
| RF | Random Forests |
| S2 | Sentinel-2 |
| SAR | Synthetic Aperture RADAR |
| SWIR | Short Wave InfraRed |
| TOA | Top of Atmosphere |
| UAV | Unmanned Aerial Vehicle |

# 1

# INTRODUCTION

Wetlands are considered to be amongst the most important ecosystems on the planet. Ecosystem services from wetlands include carbon storage, water purification, flood attenuation, natural resource provision and biodiversity support (Millennium Ecosystem Assessment, 2005; Tockner and Stanford, 2002). Despite their valuable services to the environment and us humans, wetlands are being threatened worldwide (Hu et al., 2017). The loss and degradation of wetlands has caused rapid decline in freshwater biodiversity (Dudgeon et al., 2006) and loss of their key provisional, regulatory and cultural services (Zedler & Kercher, 2005). Wetlands are often highly dynamic systems owing to the seasonal availability of water, and resultant ecological response. These natural variations often drive the primary productivity of plants (Ward et al., 2014), and govern the presence of terrestrial and aquatic species (Campbell et al., 2016). Variability in water availability thereby controls the structure of the local food-web and overall biological diversity (Poff et al., 1997).

Changes in wetland hydrology often correlate to human-driven changes such as dam construction, deforestation or agricultural activities (Ilyas et al., 2019). Upstream changes in land use can lead to alterations in river discharge and sediment transport, which may disrupt the physical and ecological equilibrium of a wetland (Bregoli et al., 2019), and eventually lead to irreversible degradation (of Tanzania, 2017). An understanding of the spatiotemporal inundation patterns of wetlands and its driving forces is therefore critical in informing policy for sound environmental management (Thomas et al., 2015).

Fortunately, the growing concern for the future of wetlands and the direct consequence of their loss has lead to the establishment several environmental management programs and research groups. Although the understanding of wetland hydrology has thereby improved, there are a number of common challenges that have hampered a deeper understanding of wetland ecosystem functioning. The wide variety of wetland types and complex dynamics between water, soils, vegetation and animals in these these ecosystems often makes it difficult to understand the driving mechanisms (Ilyas et al., 2019). Field measurements are often highly informative (Gallant, 2015), however, investigating wetland hydrology using field measurements is often too time-consuming and expensive (Gallant, 2015; Zimba et al., 2018; Halabisky et al., 2016). Additionally, field-data alone is often unable to capture the broad-scale heterogeneity of wetland landscapes, and the ecological importance of the spatial patterns of inundation often makes 2D hydrological modelling unsuitable. Remote sensing techniques have a high potential to estimate hydrological processes, especially for wetlands with a scarce amount of field measurements (Mohamed et al., 2004; Slagter et al., 2020).

## 1.1 RESEARCH MOTIVE

The Mara floodplain wetland exhibits seasonal inundation patterns, which govern the presence of species, and thereby control the functioning of the ecosystem. In addition, the seasonal inundation of the floodplain is essential for several local livelihood activities such as crop cultivation and fisheries (Bogers, 2007). Over the past 100 years the Mara Wetland has undergone significant morphological changes due to various natural and anthropogenic factors, of which the extensive deforestation in Kenya has been a major culprit (Bregoli et al., 2019). The future of the Mara Wetland ecosystem is now further under threat due to projected urban developments and climate change. The planned dam construction upstream of the wetland could systematically alter the natural fluctuations in river flow, and thereby also the functioning of the ecosystem and accessibility of wetland resources for local communities. The additional projected increase in average temperatures and increase in the amount and intensity of rainfall adds to the vulnerability of the wetland. Furthermore, continuous population growth around the wetland will increase water demand and thereby put pressure on the water resources essential for wetland ecosystem functioning.

The hydrodynamics of the wetland and the driving factors must be understood in order to attune dam operations and local climate adaptation strategies to meet both ecological and social requirements.

## 1.2 PROBLEM STATEMENT AND KNOWLEDGE GAPS

Past research on the Mara River Basin (MRB) has demonstrated the uncertainties that exist in the understanding of the hydrological and ecological dynamics of the wetland system (Dessu and Melesse, 2012; Bregoli et al., 2019; Subalusky et al., 2019; Hulsman et al., 2018; Bogers, 2007). The Mara Wetland is situated at the most downstream point of the Mara River, at the intersection between the river and Lake Victoria (LV). Although the common understanding exists that the Mara River, LV, local rainfall and groundwater play a role in governing the hydric status of the wetland, the relative contribution of each is still unknown. The difficulty of conventional hydrological modelling of the wetland system is in part due to the backwater effect from LV. The varying stage levels of the lake results in changing boundary conditions and distorted discharge measurements at the outlet. Additionally, the wide spatial distribution of water throughout the wetland means 2D hydrological modelling is an unsuitable method to capture the full complexity of the system.

A number of studies on the MRB have used aerial photographs and Landsat imagery to estimate the expansion of the wetland over the past 50-60 years (Bregoli et al., 2019; USAID, 2018; Mati et al., 2008; Mayo et al., 2013; Mutie et al., 2006; Mayo et al., 2013). The reported values for wetland area are presented as average yearly values, which does not capture the seasonal fluctuations in extent and the duration of inundation, also referred to as a wetland's 'hydroperiod' (Council, 1995). Field measurements of the Mara wetland are insufficient to explain the hydrological processes of the wetland; therefore, the hydrological dynamics of the Mara Wetland in Tanzania is analyzed using remote sensing techniques. Satellite imagery provides the opportunity to visualise and understand the spatial and temporal dynamics of the wetland system, including those parts that are otherwise inaccessible. A time-series of the aerial extent of expansion and contraction can provide a basis analysing the driving force behind these patterns. Coupling trends in inundation extent with stage field measurements been researched for the Inner Niger Delta (Ogilvie et al., 2015), the Zambezi River Basin (Zimba et al., 2018) and for a coastal wetland systems in Mexico (Wickel et al., 2020) and but has not been done for the Mara Wetland.

## 1.3 RESEARCH OBJECTIVES AND SCOPE

The purpose of this study is to investigate hydrological dynamics of the Mara Wetland by reconstructing the spatial and temporal inundation patterns of the wetland using optical remote sensing data. The seasonal fluctuations in aerial inundation extent are compared to the fluctuations of the (quantifiable) water balance components: LV stage levels, upstream Mara River discharge, local precipitation and evapotranspiration.

This research aims to shed light on the underlying mechanisms and hydrological processes that control the hydric status of the wetland. Comparing the temporal changes in extent with the corresponding changes in magnitude of the surrounding physical processes can provide insight on the contribution of the specific water balance components to the inundation process. The outcome of the study aims to support various stakeholders including decision-making parties, researchers and local authorities in the development of water-resource related activities upstream of the wetland and formulation of climate adaptation strategies for the wetland. Understanding the dynamics and processes of the water cycle in wetlands is a vital part of managing this system in a sustainable manner.

The objective of this study is thereby two-fold. On the one hand it is aimed at contributing to the body of knowledge on the water balance and dominant hydrological processes of the Mara Wetland. In the interem, the secondary objective is to gain an insight on the potential of using optical remote sensing techniques to reveal the physical dynamics and interactions of the hydrological components of a vegetated wetland ecosystem. This gives rise to the following research questions:

*How does the spatial extent of the Mara Wetland change intra- and inter-annually?*

and

*How accurate can the land-covers in and around the Mara Wetland be mapped using Sentinel-2 (S2) imagery?*

and

*How does the intra- and inter-annual trends in wetland inundated area relate to the discharge upstream, stage levels downstream and local precipitation?*

and

*What is the water-balance of the wetland on an average spatial, monthly temporal scale?*

## 1.4 READER'S GUIDE

The remainder of this report is as follows. After this introduction, the report contains a chapter describing the relevant theoretical background and concepts deemed necessary to follow the research purpose and methodology (Chapter 2). Following this, Chapter 3 provides an introduction of the study location (Mara Wetland). The next chapter, Chapter 4 provides a description of the data used, the data source and how the data is pre-processed to be used in the methodology. Following this, Chapter 5 contains a description of the research methodology, subdivided in two main research phases (Phase-I, Section 5.2 and Phase-II, Section 5.3). This chapter contains a number of flow-charts that present the full methodology in a concise way. Subsequently, the results are presented in Chapter 6 and discussed in Chapter 7. Overall conclusions and recommendations are given in Chapter 8.

# 2

# THEORETICAL BACKGROUND

This chapter contains the theoretical background information relevant for this study. A general introduction of wetland ecosystems is provided in Section 2.1, and the definitions of the terms used to describe wetland hydrology are given in Subsection 2.1.1 and Subsection 2.1.2. Furthermore, Section 2.2 provides a general overview what remote sensing entails. Lastly, Subsection 2.4.1, Subsection 2.4.2 and Subsection 2.4.3 provide a short literature review on previous studies that are have similarly used satellite remote sensing to map wetland hydrodynamics.

## 2.1 WETLAND CONCEPTS

The concept of what constitutes a wetland is rather ambiguous, but is broadly understood to be a habitat that has the unique properties of both water and land, or as the transitional zones between aquatic and terrestrial environments. Various definitions for the term 'wetland' exist in literature, many of which. however, are contradictory. A concise, generic definition that defines the term 'wetland' has not yet been agreed upon. Attempts to develop a global understanding have nonetheless been made. The Ramsar Convention on Wetlands is the intergovernmental treaty for the protection of wetlands and its resources, signed in 1971 in Ramsar, Iran and now joined by 171 nations as contracting parties. Article 1 of the convention defines wetlands as "areas of marsh, fen, peatland or water, whether natural or artificial, permanent or temporary, with water that is static or flowing, fresh, brackish or salt, including areas of marine water the depth of which at low tide does not exceed six metres" ("Ramsar Convention", 1971). Since the definition itself ambiguous terms such as 'marsh, fen, peatland', a more generic definition is used in this study:

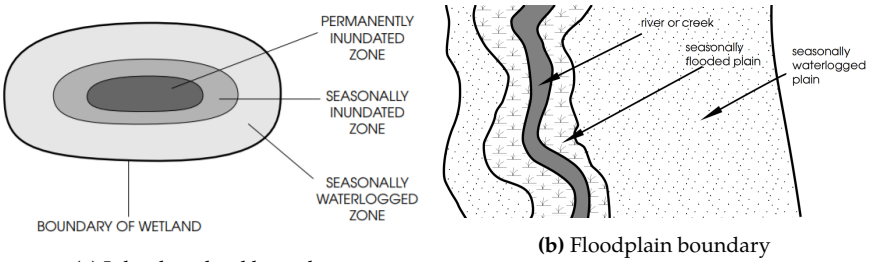
> *"Areas of permanently, seasonally, intermittently, or tidally waterlogged to inundated soils, sediments, or land, whether natural or artificial, fresh to saline"* (C. A. Semeniuk & Semeniuk, 2004)

There are five main natural wetland types; marine, estuarine, lacustrine, riverine and palustrine. Finding global agreement on a classification system to further subdivide these wetland types has proven to be a challenge. Different definitions and classifiers are used, which are often not relevant for wetlands in other climatic zones (C. A. Semeniuk & Semeniuk, 1995).

### 2.1.1 Boundary definition

Using this definition, the boundary that distinguishes wetland from upland is according to the extent to which the soil is temporarily or permanently waterlogged. Identifying the precise location of the boundary between wetland and upland is an important part of mapping wetlands for conservation or management purposes. This boundary can in many cases be defined according to the hydrology of the area, the hydric status of soils and the presence of aquatic vegetation. Difficulties in delineating wetlands can arise when wetlands have multiple centers or if the wetland is increasingly covered by alien vegetation (C. A. Semeniuk & Semeniuk, 2004).

Wetlands where the input, may it be rain, river discharge or groundwater, shows strong seasonal variation, exhibit correspondingly migrating boundaries in waterlogged land. These wetlands often have a central zone of permanent inundation and outer zones of waterlogged lands. For such inland wetland areas, such as the Mara Wetland, the boundary of the wetland is defined by the maximum extent of the area that is waterlogged. Figure 2.1a gives a schematisation of the how such a wetland is delineated. Similarly, the floodplain around the river that feeds the wetland is delineated in a similar manner, shown in Figure 2.1b.




**(a)** Inland wetland boundary

**(b)** Floodplain boundary

**Figure 2.1:** Wetland boundary according to area of inundation and extent of waterlogging (C. A. Semeniuk & Semeniuk, 2004)

For simplicity in distinguishing between 'dryland' and 'wetland' areas, 'dry' and 'wet' in this study actually refers to the degree of moisture. Dryland areas are thus characterised by the absence of the moisture, and wetland areas that are either moist, waterlogged or flooded.

### 2.1.2 Hydro-period and hydro-pattern

Hydro-period is defined as "the period in which a soil area is waterlogged" (Merriam-Webster, n.d.). In a study on the hydrology of prairie wetlands published in 2016, Hayashi points out the limitation of this definition, as it does not bring across the importance of subsurface moisture dynamics, which play an important role in the overall hydrology of the system (Hayashi et al., 2016). Wetland systems should be seen as highly integrated hydrological units that are governed by both surface and groundwater processes (Rains et al., 2016). Hydro-pattern is defined as "describing the variation of water levels over time and space" (U.S.EPA, 2008). The term hydropattern thereby also encompasses the drivers of water level variability; namely the changes in hydrologic inputs and outputs, as well other hydraulic controls such as physical changes within the wetland. The hydropattern is thus the term adopted in this study to describe the seasonal hydrology of the wetland.

### 2.1.3 Ecosystem services

It has long been recognized that humans depend on the Earth's natural resources for their well-being. The term "ecosystem services", has, however, only been introduced in 1970. The Millennium Ecosystem Assessment (MA) was called for at the United Nations General Assembly in 2000 and aims to examine the functioning of ecosystems, and specifically relating to the consequences of changing ecosystems on human well-being. Herein the term "ecosystem services" is defined as "benefits people obtain from ecosystems" (Millennium Ecosystem Assessment, 2005). Figure 2.2 shows an overview of the commonly encountered ecosystem services and how these link to human well-being as understood by the MA.

The main ecosystem services provided by wetlands that have been recognized in literature are amongst others; gas/climate/water regulation, erosion control, sediment retention, nutrient cycling, refugia, food production and cultural/recreational values (Costanza et al., 1997).

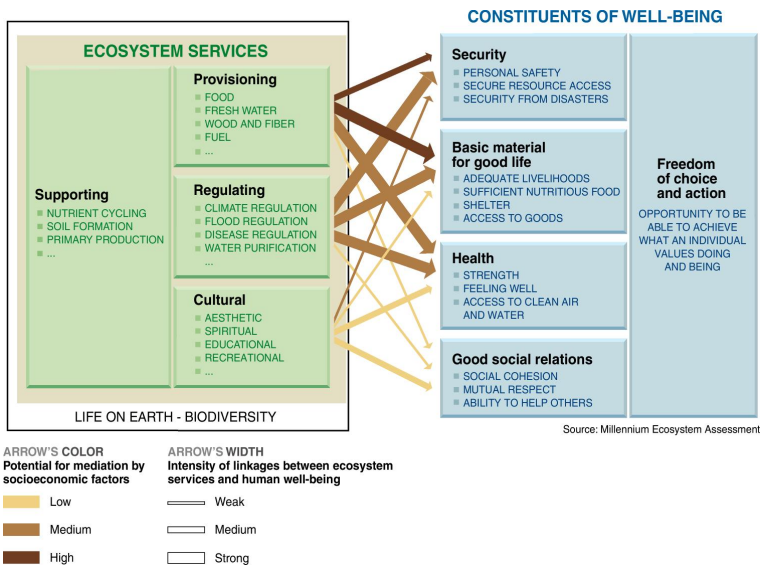



**Figure 2.2:** Linkages between ecosystem services and human well-being (Millennium Ecosystem Assessment, 2005)

## 2.2 PHYSICAL PRINCIPLES OF REMOTE SENSING

Remotely sensed data is a highly valuable source of information of the features on Earth's surface. Different types of satellite sensors, both passive and active, record information at different portions of the electromagnetic spectrum, the majority of which are invisible to the human eye. The sensors utilize specific characteristics of the land and water covered areas to detect differences between the features of Earth's surface. The sensor thereby relies on the principle that the physical properties of these features change the way in which the energy is reflected. The resultant satellite images are visualized by assigning colors to the, otherwise invisible, sensor response.

### 2.2.1 Passive optical remote sensing

This section provides a rudimentary explanation of the working principles of passive optical remote sensing. 'Passive' indicates the satellite is only a receiver and not also an emitter, contrary to 'active', which means the instrument both emits and receives radiation. Optical remote sensors depend on radiation from the sun. A fraction of the solar irradiance passes through the atmosphere and reaches the Earth's surface, where it is then absorbed, reflected or transmitted from target objects. The sensor detects radiation from the target objects at different regions (bands) of the electromagnetic spectrum (see Figure 2.3). The spectrum is organized according wavelength/frequency.

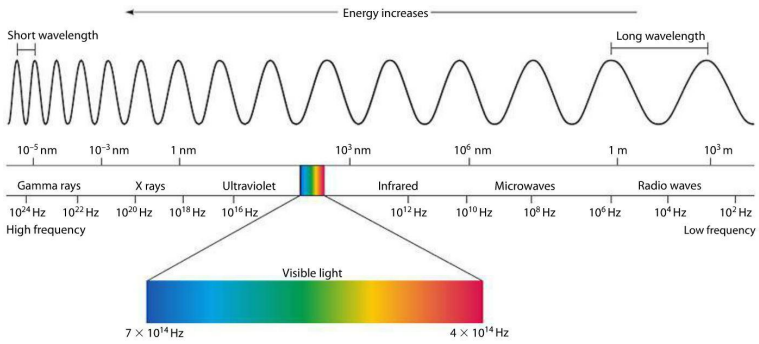



**Figure 2.3:** The Electromagnetic Spectrum (Physics, 2016)

*Spectral signature of land-cover types*

Depending on the physical characteristics of the target object, the wavelength at which the incoming radiation is reflected or absorbed differs. The precise percentage of radiation reflected or absorbed is unique for all target materials and results in a characteristic spectral signature as shown in Figure 2.4.

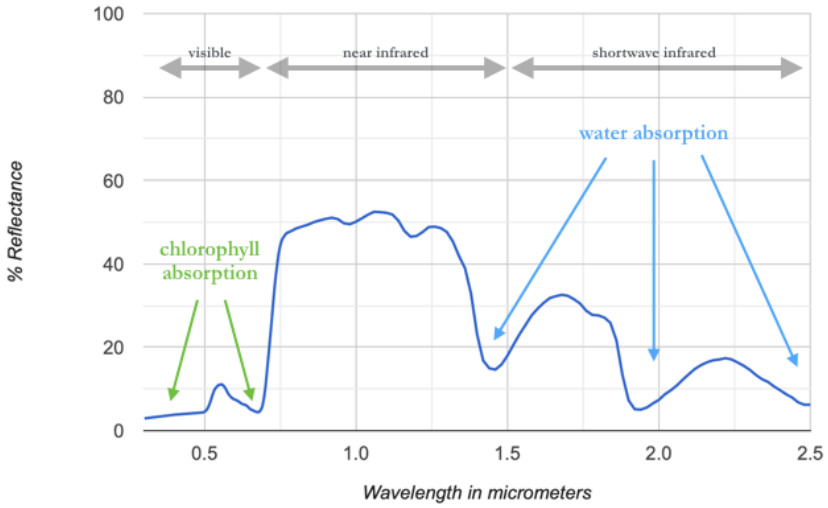



**Figure 2.4:** Typical spectral signature of grassland pixels. Edited from (HSU, 2019)

This principle is the basis of optical remote sensing, and allows us to distinguish between different types of land surfaces. The (spectral and spatial) resolution of a sensor determines which information is given in a pixel and the level of detail of this information.

### 2.2.2 Multispectral imagery

Optical remote sensing uses the visible, near infrared and shortwave infrared regions of the electromagnetic spectrum (Liew, 2001). Multispectral sensing is a specific type of optical remote sensing operating with a multi-channel receiver. Other types of optical remote sensors are panchromatic (one broad band) or hyperspectral (numerous very narrow bands). Each channel of the multispectral instrument is sensitive to radiation within a range of wavelengths, known as bands. The exact width of the bands (range of wavelengths) to which the sensors on the instrument are sensitive to differs depending on the properties of the satellite instrument. Examples of multispectral satellites are Sentinel-2 (S2), Landsat Multispectral Scanner and IKONOS.

### 2.2.3 Bands and indices

The response from different bands results in a multilayer image, whereby each layer can be analysed independently. The response of the pixels at each band, or image layer, can be visualised alone or in combination with other bands. Displaying a single band results in a panchromatic (gray-scaled) image. It is often much more useful to create a composite (by stacking the different bands) and assigning colors to each.

### 2.2.4 Spectral, spatial and temporal resolution

As mentioned before, the spectral and spatial resolution are important factors in determining the type and quality of the information contained within a pixel. The spectral resolution is the width of the wavelength intervals (bands), an instrument with a high spectral resolution has narrow bandwidths. The spatial resolution is the size of the pixels in the satellite image, and determines at what level of detail the Earth's surface represented in the image.

The temporal resolution, also known as the revisit time, is the frequency at which a satellite covers the same area. There is often a trade-off between spatial and temporal resolution, because high spatial resolution sensors have a smaller field of view and thus take more time to cover the same area as a satellite with a lower spatial resolution.

## 2.3 IMAGE CLASSIFICATION

It is often useful to classify a region according to the different land-cover types present. This can be done by applying an image classification algorithm that uses the unique spectral features of land-cover types. The classification procedure can be 'supervised' or 'unsupervised'. A supervised classification algorithm requires input in the form of user-defined 'training areas'. The algorithm then classifies the remaining image pixels according to the similarity of its spectral features to these training areas. The 'unsupervised' approach requires no user-input other than the number of groups it should generate, the image is then automatically segmented according to similar pixels, forming 'clusters'. An example of a map of the the Mara Wetland produced using a supervised classification algorithm is provided in Section B.2 in Appendix B.

There are advantages and disadvantages for both methods. The main disadvantage of supervised classification is the user-input that is required, which can be time-consuming and there must be a certain amount of prior knowledge of the area. For unsupervised the main disadvantage is that the grouping may not be in line with the goal of the classification and there is no control over the classes. Common image classification algorithms are Classification and Regression Trees (CART), Random Forests (RF), Support Vector Machines (SVM) and Naïve Bayes (NB).

### 2.3.1 Random Forests (RF) algorithm

The complexity and heterogeneity of wetland regions calls for a classification method whereby multiple land-cover characteristics can be taken into account simultaneously. The RF classification algorithm has the capability to process high dimensional, multi-source datasets to form a combination of tree-predictors. It is frequently used for image classification due to its general robustness to over-fitting and quick training and prediction speed. The algorithm constructs several decision trees from a random subset of data points and ensembles the outcome of all trees in the forest to make a prediction for each pixel. Due the participation of a large number of decision trees, the RF is generally an accurate method for complex classifications. The prediction model constructed using RF is, however, often difficult to interpret. The RF algorithm is a more 'black-box' approach than other algorithms, as is constructs a number of trees for every point and thus troubleshooting is not straightforward. Figure 2.5 shows a visualisation of the sampling and classification process of the RF algorithm.

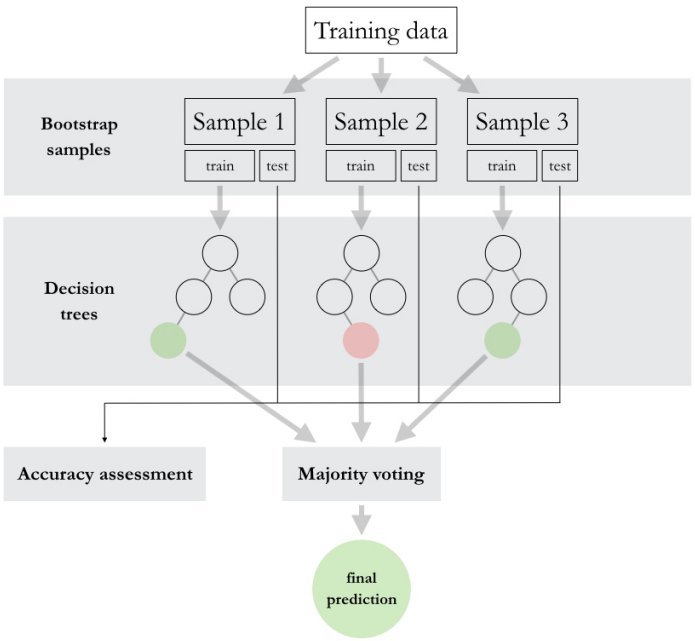



**Figure 2.5:** Example of Random Forests prediction model

## 2.4 MULTITEMPORAL REMOTE SENSING APPROACHES FOR WETLAND MAPPING

There is a growing awareness of the inter-dependency of society and the spatial and temporal extent of water bodies, which has lead to an increase of research in this area. The local services and regional climate regulation capacity of wetlands in particular highlights the global importance of monitoring the status of wetlands.

The use of remote sensing technologies for wetland research has been increasing rapidly over the past 50 years. Applications vary from change detection to monitoring carbon release to analyses of hydrological process (Guo et al., 2017). The main challenges in global wetland delineation is their non-unified nature. Wetlands vary greatly in physical characteristics, the only common factor being the presence of water. Differences in vegetation, permanence and shape, makes consistent detection a challenging endeavor (Gallant, 2015). The analogy of wetlands as a "moving target" that portray the hydro-biological cycle rather than a single land-cover (Gallant, 2015), appropriately exemplifies the difficulty of remotely monitoring wetlands. Former approaches in wetland monitoring via remote sensing include the use of aerial photography, multispectral imagery, Radio Detection and Ranging (RADAR) and Light Detection and Ranging (LiDAR). Several trade-offs must be made in choosing a suitable satellite data-set; such as spatial and temporal resolution, the ability for sensor to detect specific wetland features and its sensitivity to cloud-coverage.

### 2.4.1 Optical remote sensing of wetlands

The main limitations of using optical imagery for wetland monitoring are the influence of clouds, and the inability for the sensors to detect water bodies underlying vegetation. However, the visible, near/mid/shortwave-infrared bands of optical sensors are sensitive to moisture, and can thus effectively separate wet and dry surfaces. The visible bands are also sensitive to chlorophyll content, and are thus useful in distinguishing between different vegetation types and their status. Band ratios (in the form of indices such as Normalized Difference Vegetation Index (NDVI) or Normalized Difference Water Index (NDWI)) are particularly useful for areas where water and vegetation coexist (Ma et al., 2019). Several researchers have used optical remote sensing data to characterize the hydrological dynamics of wetland systems. The surface water dynamics of depressional wetlands in Douglas County, Washington (Halabisky et al., 2016), temporary wetlands in Texas (Collins et al., 2014) and the vegetated Macquarie Marshes in Australia (Thomas et al., 2015) have been reconstructed using a time series of Landsat satellite imagery. Research purposes that can tolerate lower resolution (500 m versus < 50 m), often exploit images obtained by the Moderate Resolution Imaging Spectroradiometer (MODIS) sensory aboard the Terra satellite, due to shorter revisit time and greater number of bands. For example, the flood dynamics of the Inner Niger Delta over a time-span of 10 years has been reconstructed using MODIS imagery (Ogilvie et al., 2015). (Alonso et al., 2020) and (Zimba et al., 2018) couple hydrological field data with MODIS imagery to reproduce inundation extents of wetlands in the Palo Verde national park in Costa Rica and Zambesi River Basin, respectively.

### 2.4.2 Microwave remote sensing of wetlands

Unlike optical imagery, Synthetic Aperture RADAR (SAR) is able to penetrate clouds and is thus often exploited for areas with frequent cloud-cover. Depending the operating wavelength, polarization and incidence angle of the SAR instrument, the resultant imagery shows varying degrees of penetration through vegetation and sensitivity to moisture and structure of land-covers. The limitation of using SAR imagery for wetland monitoring is the difficulty in analysing and interpreting images (Muro, 2019; Tsyganskaya et al., 2018; Gallant, 2015). The interactions between ground environmental conditions and the incoming signal are highly complex, and must often be supported by additional data sets to interpret change detection. An understanding of the relationship between sensor characteristics; such as wavelength, polarisation, geometric resolution, angle of incidence, and ground conditions, such as vegetation type and phenology, soil moisture and water depth, is required to accurately describe the satellite imagery (Tsyganskaya et al., 2018). The backscattering from wetland is a sum of many independent scattering features, which means straightforward methods such as Persistent Scattering (PS) monitoring hydrological behavior of wetlands (Mohammadimanesh et al., 2018). Monitoring seasonal change in wetland extent using multitemporal SAR imagery has been done in a number of recent studies. The main methods by which wetland extent change has been monitored are; visual interpretation, Interferometry SAR coherence methods and grey-level threshold classification (White et al., 2015; Muro, 2019; Dabboor and Brisco, 2019).

The repeat-pass SAR interferometry technique is not common in the context of monitoring wetlands because of the low coherence between images due to vegetation growth and changes in water level

(Mohammadimanesh et al., 2018). Monthly flooded area classification over a period of 5 years using low resolution SAR has been applied in the Sudd Wetland in South Sudan (Wilusz et al., 2017). Threshold classification was used to distinguish between dry land, flooded vegetation and open water. The results show a time-series of the monthly change in flood extent as well as a spatial representation of flood frequencies. A weather independent water-cover detection method has been developed and applied to the Alkaline Lakes region in Hungary (Gulácsi & Kovács, 2020).

### 2.4.3 Combined approaches

Synergistic use of optical and radar satellite imagery is common in the field of wetland mapping, as this method frequently outperforms single data-set analyses (Muro, 2019). Combined approaches are particularly useful for cases where optical imagery does not reach the desired temporal frequency due to frequent cloud cover. Combining Sentinel-1 (SAR) and Sentinel-2 (optical) has shown to lead to higher wetland classification accuracy than using only one of these (Slagter et al., 2020; Mahdianpari et al., 2019). There are, however, classification examples whereby the use of Sentinel-2 bands and (Sentinel-2-derived) indices reaches nearly the same accuracy as combination of Sentinel-1 and Sentinel-2 does (Muro, 2019). For a study whereby ALOS L-band SAR is used in combination with Landsat imagery to characterise the dynamics of water and vegetation in northern Australia, the (Landsat-derived) Normalized Difference Infrared Index (NDII) index was particularly effective at distinguishing flooded and non-flooded vegetation (Ward et al., 2014). In the context of wetland delineation, optical input variables have again shown to be of greater importance in the classification process with respect the SAR input variables (Hird et al., 2017). The relative importance of optical versus SAR input variables suggests the productivity of vegetation and occurrence of water detected by optical sensors (by employing vegetation and water indices) may be more indicative of wetland occurrence than moisture content of soil or surface structure are.

# 3

# STUDY AREA

This chapter provides an introduction to the study area. Section 3.1 consists a general introduction of the region and Section 3.2 provides more information on the local climate. Section 3.3 describes the past and projected environmental changes of the wetland area. Section 3.4 and Section 3.5 respectively provide details on the ecosystem vegetation and basin hydrology. Lastly, a brief description of the local communities and their activities is provided in Section 3.6.

## 3.1 GENERAL INFORMATION

The Mara (or Masirori) Wetland (1°31′45"S, 34°7′34"E) is located at the lower reaches of the The Mara River Basin (MRB); a transboundary watershed (roughly 13,750 km²) shared between Kenya and Tanzania (Mwangi et al., 2017). The Mara River (approximately 130 km) originates from the Mau forest complex in the Kenyan rift valley at about 3000 m a.s.l., and flows southwest through the Maasai-Mara and Serengeti National Reserves, after which it forms large wetland area before discharging into Lake Victoria (LV) at Musoma Bay, Tanzania at 1120 m a.s.l. (Bregoli et al., 2019). Discharge is measured at Mara Mines (MM), a gauging station located approximately 25 km upstream of the wetland. The wetland area is sloping only very gently from MM to the outlet of LV. The shape and general flood pattern of the wetland can be largely explained by the topographic features in the area. The narrow outlet to LV is bounded by steep hills on either side. Continuing upstream, the northern edge of the wetland is sharply defined by the Utimbaru fault (Kabete et al., 2012), whereas the southern edge is only gently sloping, which provides room for seasonal flooding. The upstream, eastern parts of the wetland are only seasonally inundated and consist of temporarily waterlogged marshes and herbaceous vegetation (Bregoli et al., 2019). The middle and lower parts of the wetland consists of a vast area of perennially saturated papyrus swamps, of which the canopies can reach up to 5 m in height (Bogers, 2007). Besides Cyperus papyrus, Typha domingensis (cattail or Typha) and Phragmites australis (common reed) can also be found in within the wetland area (Muraza et al., 2013). The papyrus is particularly dominant, covering virtually all of the water surface in the wetland. The papyrus plant is clearly recognizable in satellite imagery due to its unique structure and dense coverage.

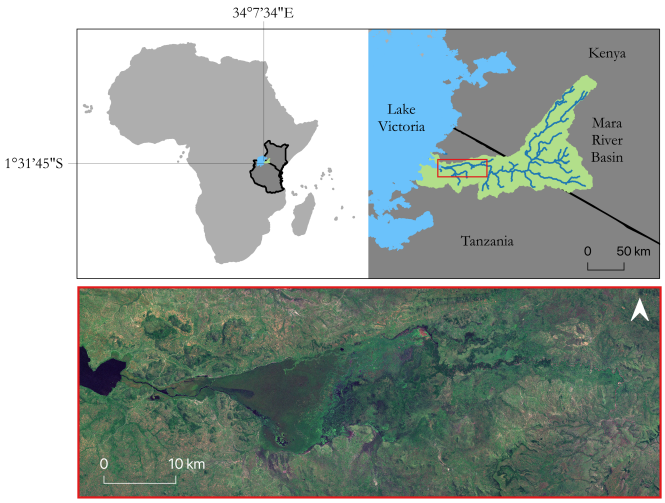



**Figure 3.1:** Map of the location of Lake Victoria (blue) and Mara River Basin (green) and the outline of the wetland region (red box)

## 3.2 CLIMATE

The rainfall regime in the area is bimodal, with two distinct rainy seasons between March - June and October - December. The 'short' rainy season (alluding to the duration of the precipitation events rather than the duration of the season), between October and December marks the start of the hydrological year and is driven by the southward movement of the Intertropical Convergence Zone (ITCZ), the 'long' rainy season is the result of the ITCZ returning north. Inter-annual variability in the timing and length of the wet and dry seasons is shown to be rather high due to the dependence of surface water temperature of the Indian Ocean (Black, 2005).

The rainfall amount decreases with altitude from east to west, the upper reaches of the river on the Kenyan side receive an average yearly rainfall of 1000 - 1750 mm whereas the lower reaches, including the Mara wetland, receives only 300 - 850 mm (USAID, 2018). Local average temperature ranges between 13 and 30 °C. June/July are usually the coolest months and September/October the warmest.

## 3.3 DRIVERS OF CHANGE AND FUTURE THREATS

The Mara Wetlands in Tanzania have undergone rapid changes in the past years, attributed to land use change upstream as well as a number of other possible anthropogenic stressors (Bregoli et al., 2019). The ecological conditions of the MRB have been shown to have been relatively stable over time scales of 1000 to 2000 years before present (Subalusky et al., 2019). The high level and reliable nature of the rainfall patterns, and the fertile soils has attracted communities and encouraged farming practices. The rapid hydrological and ecological changes in the basin, especially since the 1960's, are attributed to anthropogenic factors (Subalusky et al., 2019; Houlahan and Findlay, 2004; Bregoli et al., 2019; Mwangi et al., 2017; of Tanzania, 2017). There has been a reduction in vegetation cover in the upper catchments due to clearing of forests and conversion of Savannah grasslands into agricultural areas. Forest loss in the MRB has been investigated by Mwangi et al., which reports a loss rate of 36 $km^2$/year between 2003 and 2014. The loss of rooted trees has impacted the water and sediment transport downstream. In addition, high population growth rates in the aforementioned districts (estimated 28% between 2002 and 2012) has lead to large-scale land conversion (natural vegetation to cropland and pasture) and has put pressure on natural resources from the wetland (USAID, 2019). Projected changes in the regional climate include an increase in temperature, changes in variability and frequency of rainfall events and longer duration of drought events. These changes, coupled with high levels of poverty and a heavy reliance on rainfed agriculture, make the communities around the wetlands particularly vulnerable to climate change.

## 3.4 VEGETATION AND ECOSYSTEM FUNCTIONING

Precipitation that falls within the MRB runs off into streams, which combine to form Mara River. Pollutants from agricultural, mining and other industrial practices in the basin dissolve in or are adsorbed by the water and sediment moving downstream to the wetland area. The primary pollutants in the Mara River include organic nitrogen and nitrates, of which almost 30 % is removed by the wetland after a retention time of only about 36 hours (Mayo et al., 2013). Aside from entrapment of pollutants by the sediment in the wetland, much of the incoming nitrogen is removed by plant uptake and can be traced back in the biomass of several species. Both the sediment and vegetation thereby play an important role in protecting LV from the pollutant loads from economic activities. In addition to pollutant uptake, the wetland has been estimated to trap 90% of the in-flowing suspended sediment (Bregoli et al., 2018).

## 3.5 HYDROLOGY OF THE MARA WETLAND

General trends on the decadal hydrological changes are relatively well understood. In these studies the river-wetland-lake system is generally understood as follows. The Mara River flows from the Mau forest through the Serengeti, then west-ward through croplands before entering the Mara Wetland, after which the flow turns Southward and finally continues West before discharging into LV. The extreme expansion of the wetland in recent years has motivated research on land-cover changes and resulting hydrological and sedimentological changes occurring in the MRB (Bregoli et al., 2019; Mutie et al., 2006; Mwangi et al., 2017, Hulsman et al., 2018; Dutton et al., 2018; Mayo et al., 2013; Defersha et al., 2012, Kiragu et al., 2011). Table A.1 in Appendix A provides an overview of areas estimates as calculated in previous studies. Past research concurs on the influence the deforestation upstream has had on the runoff and sediment transport of the river. Visible effects from the loss of forest include an increase in suspended sediment levels in the

river due to the lack of tree roots to anchor the soil, accelerated bank erosion (Kiragu et al., 2011) , and clogging of river channel causing avulsions and leading to wider distribution of water-flow (Bregoli et al., 2019).

The complete water balance of the wetland is to date not well understood, especially the seasonal dynamics of the system and the relative influence of the LV, the Mara River and local precipitation on the expansion process. Precipitation patterns govern the seasonal stage level fluctuations of LV, and usually show an annual peak between May and July and are lowest between August and December. The wetland is connected to the eastern edge of LV and thus wetland inundation patterns could potentially be driven (in part) by lake level fluctuations. The degree to which backwater from the lake influences the inundation patterns of the wetland seasonally is not yet understood. In addition, lateral and/or groundwater inflow to the wetland region, and the evaporative fluxes from the wetland have not been quantified to date. Without seasonal analyses, the relative influence of Mara River discharge, local rainfall and stage levels of LV cannot be evaluated.

## 3.6 POPULATION AND CONTEMPORARY LIVELIHOOD

There are a number of (largely rural) communities living along the lower reaches of the Mara River, and in particular around the wetland area. The wetland area covers part of four districts; the Tarime and Rorya districts in the the North-East and North-West, respectively, and the Butiama and Musoma Urban district in the South-East and South-West, respectively. Within these districts there are several villages situated along the wetland edge because of the ecosystem services the wetland provides. The main economic and life-sustaining activities are livestock keeping, smallholder farming (mainly foodcrops such as maize and sorghum) (Mati et al., 2008), and mat making from papyrus plants (Kema, 2010). Every household from those interviewed in the study by Kema in 2010, were involved with crop cultivation activities within the wetland area. This is likely due to the favorable, wet conditions of these soils. The wetland floodplain is sometimes made accessible for crop cultivation and livestock grazing through burning activities (Bogers, 2007; USAID, 2018). Livestock is similarly dependent on the fertile wetland areas for the natural green vegetation for grazing.

# 4

# DATA ACQUISITION AND PRE-PROCESSING

In this chapter the processing tools, datasets and data pre-processing steps are described. Section 4.2 provides information on the remote sensing datasets used in the research. This includes an account of the pre-processing steps performed on the Sentinel-2 (S2) data. Information on the hydro-meteorological datasets (river discharge, lake water levels, precipitation and evapotranspiration) is provided in Section 4.3.

## 4.1 GOOGLE EARTH ENGINE (GEE) PROCESSING PLATFORM

Google Earth Engine (GEE) is freely-accessible, interactive, cloud-based processing platform for performing geo-spatial analyses. Using GEE as a processing platform avoids the need to store images locally, and allows for greater computing power (Gorelick et al., 2017). The GEE web-based code editor uses a JavaScript API for accessing and processing data. The GEE data catalog includes a large variety of publicly available geo-spatial datasets.

## 4.2 REMOTE SENSING DATA

The S2 satellite imagery is freely available and acquired via the GEE data catalog and processed in the GEE code editor.

### 4.2.1 Sentinel-2 (S2)

The S2 data products are acquired by the MultiSpectral Instrument (MSI) aboard the Sentinel-2A and Sentinel-2B satellites, launched in June 2015 and March 2017, respectively. The MSI measures at 13 separate spectral bands, ranging from from visible (Blue, Green, Red) to Short Wave InfraRed (SWIR) (wavelengths range from 493 nm to 1374 nm). The spatial resolution depends on the spectral band, an overview of the spectral resolutions is provided in 4.2.1. The S2 data product used for this study is the Level-1C Top of Atmosphere (TOA) reflectance, available in the GEE catalog. The atmospherically corrected Level-2A Bottom of Atmosphere (BOA) reflectance dataset in the GEE catalog is only available from the beginning o 2019 to present for scenes covering the study area. Since the research does not rely on highly detailed, plot-scale reflectance values, additional atmospheric correction is deemed unnecessary. The satellite images are used rather to detect clearly contrasting land-covers, which show contrast regardless of the blueish, foggy appearance due to atmospheric scattering. Unlike SAR data, optical satellite images are affected by cloud cover and thus additional masking steps are required to obtain a suitable set of images. The additional pre-processing steps performed in GEE are outlined below.

Chlorophyll-containing vegetation absorbs in the red and Near InfraRed (NIR) bands and reflects in the green band. Pure water has high absorbance in the visible, NIR and SWIR regions of the spectrum (Ma et al., 2019). The SWIR bands are sensitive to moisture content in soils and vegetation and are therefore useful in distinguishing wet and dry soils, and healthy and water-stressed vegetation. The additional penetration capacity of SWIR bands through organic matter makes these bands particularly useful in identifying flooded areas with vegetation cover (Lefebvre et al., 2019). B1, B2, B9 and B10 are sensitive to atmospheric conditions and may thus add noise to surface signals.

#### *Pre-processing Sentinel-2 Data*

The S2 images are filtered to location and time-period, the Level-1C image collection is filtered to contain only images covering the wetland and within the date range 01-01-2017 to 01-07-2020. The resulting image collection includes two overlapping MGRS (Military Grid Reference System) tiles; '36MXD' and '36MWD'. An additional filter is applied to include only the '36MXD' tile because this tile fully contains the site

**Table 4.1:** Sentinel-2 MSI satellite sensor spectral bands

| Spectral band | Band nr. | Wavelength (μm) | Resolution (m) |
|---|---|---|---|
| Coastal aerosol | B1 | 0.433 - 0.453 | 60 |
| Blue (B) | B2 | 0.458 - 0.523 | 10 |
| Green (G) | B3 | 0.543 - 0.578 | 10 |
| Red (R) | B4 | 0.650 - 0.680 | 10 |
| Red Edge 1 (RE1) | B5 | 0.698 - 0.713 | 20 |
| Red Edge 2 (RE2) | B6 | 0.733 - 0.748 | 20 |
| Red Edge 3 (RE3) | B7 | 0.773 - 0.793 | 20 |
| Near Infrared (NIR) | B8 | 0.785 - 0.900 | 10 |
| Narrow Near Infrared (NIR) | B8A | 0.855 - 0.875 | 20 |
| Water Vapor | B9 | 0.935 - 0.955 | 60 |
| Shortwave Infrared Cirrus (SWIR Cirrus) | B10 | 1.365 - 1.385 | 60 |
| Shortwave Infrared (SWIR1) | B11 | 1.565 - 1.655 | 20 |
| Shortwave Infrared (SWIR2) | B12 | 2.100 - 2.280 | 20 |

boundaries, the portion of the tile '36MWD' covering the wetland is already contained in the '36MXD' tile. See Figure 4.1 for an illustration. The image collection is filtered by applying a maximum cloud coverage of 20 percent, which leaves a total of 112 scenes. The final maximum cloud cover percentage is chosen iteratively based on the number of remaining images to achieve sufficient temporal resolution (roughly 2/month). A cloud mask function is applied to remaining dense (opaque) and cirrus clouds. The Level-1C cloud detection algorithm is based on reflectance thresholds applied to the the visible (B1 and B2) and SWIR (B10, B11 and B12) bands (Coluzzi et al., 2018). The algorithm is applied to create the 'QA60' bit band, which is used to apply a cloud masking function. An additional filter is applied based on the percentage of masked pixels within the bounded region around the wetland. Scenes in which more than 5 percent of pixels are masked are removed, leaving a total of 87 scenes. An example of two cloudy images is shown in Figure 4.2, of which image (a) is removed and image (b) is not.

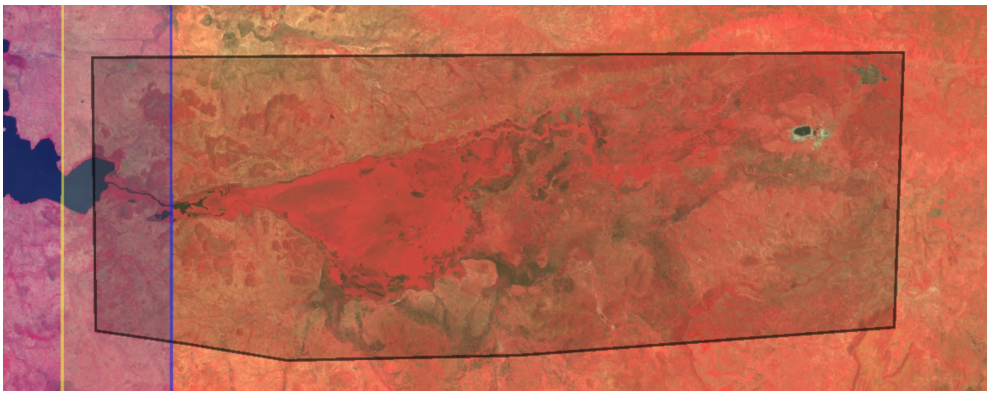
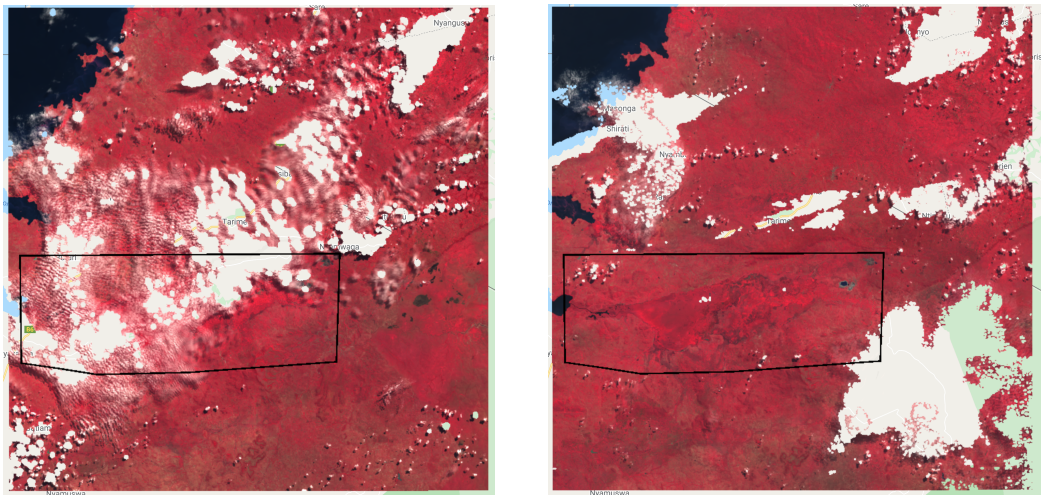


**Figure 4.1:** Illustration of a false color composites and the outline of the two the Sentinel-2 satellite image tiles 36MWD (blue) and 36MXD (yellow) and wetland delineation as set by (Bregoli et al., 2019) (black)

### *Vegetation and water indices*

The indices chosen are based on the assumption that the wetland will be distinguishable from upland areas by either the wetness of the ground, or by the state/type of vegetation present. A profuse amount of indices have been used for wetland mapping, the majority of which are used with the aim to detect open water and are sensitive to wetness only. Considering a large portion of the Mara Wetland is



**(a)** Example of a scene (01-06-2017) with a cloudy pixel percentage of 11.54 (for the entire scene) and a masked pixel percentage of 21.44 (for the set bounds, shown by black line

**(b)** Example of a scene (16-06-2019) with a cloudy pixel percentage of 15.49 (for the entire scene) and a masked pixel percentage of 1.68 (for the set bounds, shown by black line)

**Figure 4.2:** Cloud mask coverage

covered in dense vegetation, additional vegetation indices are included that are sensitive to greenness in addition to wetness. The water indices used are NDWI (Normalized Difference Water Index) and MNDWI (Modified Normalized Difference Water Index). The MNDWI is calculated in addition to the original NDWI because the use of NDWI alone tends to confuse wet areas with urban (Ludwig et al. 2019, Singh et al. 2015). The MNDWI has also been successful in detecting water features mixed with vegetation, deeper water, as well as areas with shallow surface ponding (Ogilvie et al., 2015). The vegetation indices used are NDVI (Normalized Difference Vegetation Index), Normalized Difference Moisture Index (NDMI) and NDII (Normalized Difference Infrared Index). The equation and corresponding S2 band combinations considered for this study are presented in Table 4.2. The indices are limited to the normalized difference type, because indices of different scales are given unequal weight unless these are normalized (Ludwig et al., 2019). Section B.1 in Appendix B provide composite maps showing the index response of the wetland in the wet and dry season.

**Table 4.2:** Spectral indices used as input variables for the Random Forests classifier

| Index | General Formula | S2 Band Formula | Source |
|---|---|---|---|
| Normalized Difference Water Index (NDWI) | $\frac{(G-NIR)}{(G+NIR)}$ | $\frac{(B3-B8)}{(B3+B8)}$ | (McFeeters, 1996) |
| Modified Normalized Difference Water Index (MNDWI) | $\frac{(G-SWIR1)}{(G+SWIR1)}$ | $\frac{(B3-B11)}{(B3+B11)}$ | (Xu, 2006) |
| Normalized Difference Vegetation Index (NDVI) | $\frac{(NIR-R)}{(NIR+R)}$ | $\frac{(B8-B4)}{(B8+B4)}$ | (Rouse et al., 1974) |
| Normalized Difference Moisture Index (NDMI) | $\frac{(NIR-SWIR1)}{(NIR+SWIR1)}$ | $\frac{(B8-B11)}{(B8+B11)}$ | (Gao, 1996) (Wilson & Sader, 2002) |
| Normalized Difference Infrared Index (NDII) | $\frac{(VNIR-SWIR1)}{(VNIR+SWIR1)}$ | $\frac{(B8A-B11)}{(B8A+B11)}$ | (Hardisky et al., 1983) |

### 4.2.2 Topographical data

The occurrence of wetlands depends strongly on the topographical characteristics of the landscape (Li et al., 2014). Topography is a measurable property and can thus be used as additional input variable to mask upland areas and improve classification accuracy. A Digital Elevation Model (DEM) of the wetland area is given in Figure 4.3.

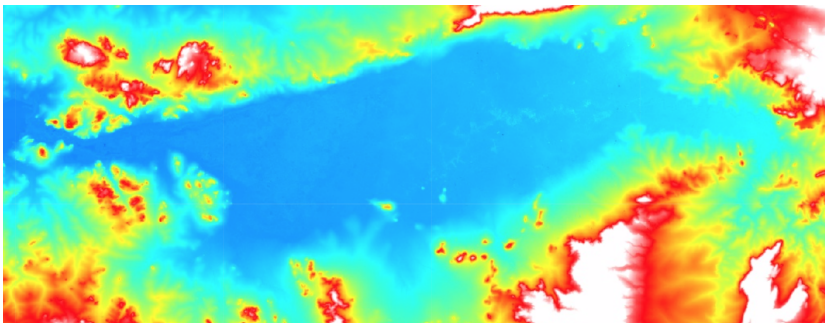


**Figure 4.3:** Digital Elevation Model (DEM) of wetland area

### 4.2.3 Orthophotography

The wetland was visited for field work in October - November 2017 by F. Bregoli Bregoli et al., 2019. Most part of the wetland is inaccessible due to the presence of dense aquatic vegetation, therefore the ground-truth data is in the form of high-resolution photographs acquired by an Unmanned Aerial Vehicle (UAV). The drone footage is used to recognize and validate land-cover classes in the training data sampling process. The photographs were acquired by F. Bregoli (Bregoli et al., 2019). The images were in the form of ortho-photographs that were projected to the Coordinate Reference System (CRS) EPSG:32736 - WGS 84. Figure 4.4 shows an example of a drone image along the edge of the wetland overlaying a Google Satellite image at two different times. The brown patch on the right image is the result of burning activities.

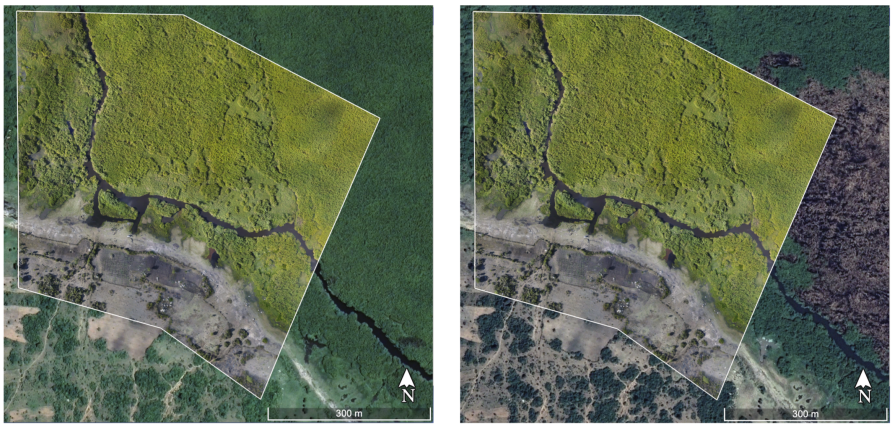



**Figure 4.4:** Example of an ortho-photograph acquired on 02-11-2017 overlaying two satellite image scenes from 28-12-2018 (left) and 02-04-2019 (right)

## 4.3 HYDRO-METEOROLOGICAL DATA

In this section the available hydro-meteorological datasets used in this study are introduced. Note that potential groundwater sources are not treated in this study. Although the wetland is believed to received some lateral inflow via groundwater (Bregoli et al., 2019), the lack of data hereof has limited the ability to quantify these flows. Nonetheless, these flows are taken into account in the final discussion.

### 4.3.1 Discharge

The discharge data is obtained from the Lake Victoria Basin Water Board (LVBWB) and is obtained from the gauge nearby the Mara Mine (MM) (see Figure 4.6) There are a few data-gaps (of max. 14 days), and no data from January 2020 due to the apparatus being washed away by an large flood event. Figure 4.5 shows the time-series of discharge data available from the start of 2017 up to the end of 2019.

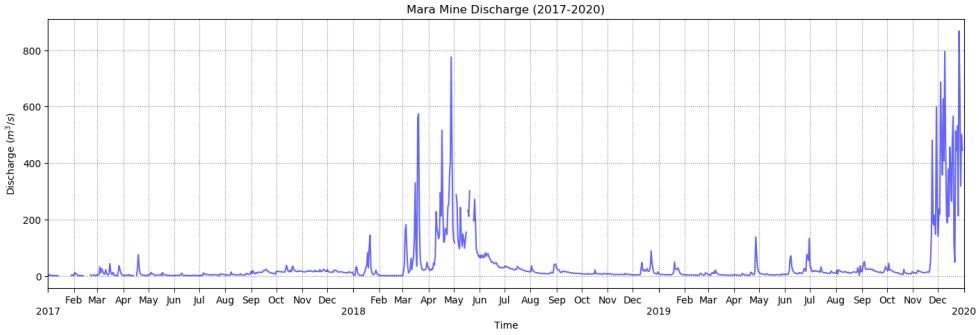



**Figure 4.5:** Discharge at Mara Mine for 2017-2020


**Figure 4.6:** Location of Mara Mine gauging station

### 4.3.2 Lake Victoria (LV) stage levels

The LV stage level data is obtained from the LVBWB. Figure 4.7 shows the measured stage levels for LV for the years 2017 to 2020. The lake level is measured in Mwanza (on the Southern coast of LV). There is a stage level instrument at Kirumi bridge (the outlet from the wetland to LV), but the data from here is incomplete. However, Figure 4.8 demonstrates that the lake level data used in this study corresponds to the levels measured at Kirumi Bridge.

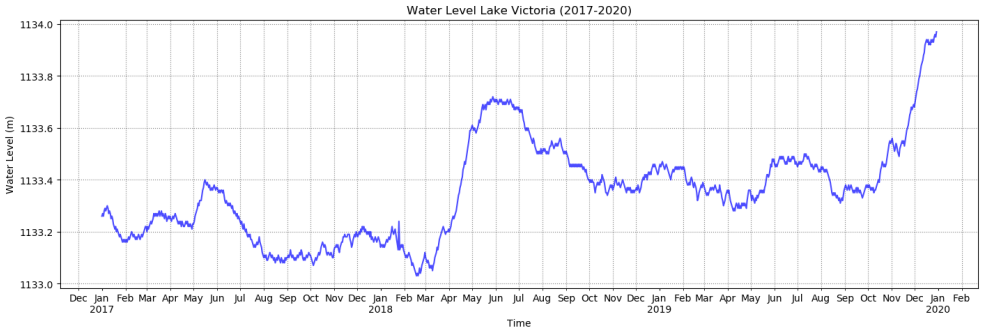



**Figure 4.7:** LV stage levels for the years 2017, 2018, 2019

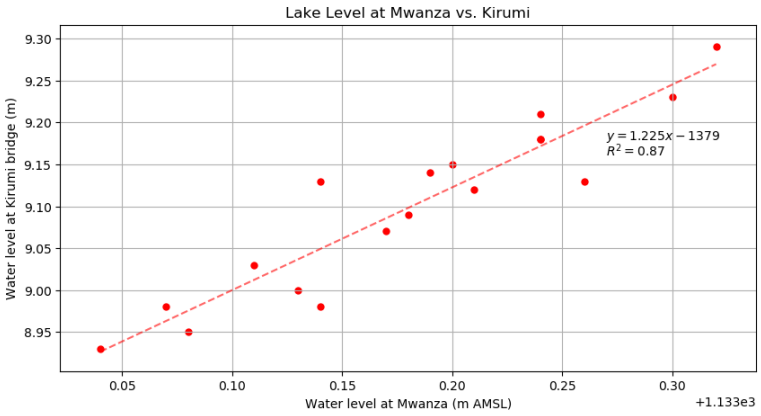



**Figure 4.8:** Correlation chart showing the water levels measured at Kirumi Bridge against those measured in Mwanza. Data from the water level at Kirumi Bridge is provided by F. Bregoli (Bregoli et al., 2019)

### 4.3.3 Precipitation

A timeseries chart of monthly accumulated precipitation derived from the TRMM (Tropical Rainfall Measuring Mission), PERSIANN-CDR (Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks-Climate Data Record), (CHIRPS) (Climate Hazards Group InfraRed Precipitation with Station Data), GPM-IMERG (Global Precipitation Measurement - Integrated Multi-satellitE), GSMaP (Global Satellite Mapping of Precipitation) and ERA5 (ECMWF Reanalysis) datasets for the years 2017, 2018, 2019, 2020 is given in Figure 4.9. Due to the lack of consistent information on the quality of the datasets for Eastern Africa, the median of all datasets is used for the analyses and the range is used as the uncertainty. Table A.2 in Appendix A provides an overview of the specifications of the precipitation datasets.

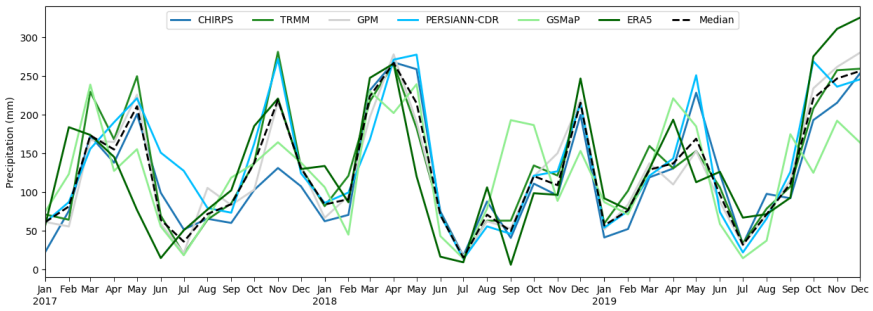



**Figure 4.9:** Regional mean monthly accumulated precipitation estimate from TRMM, PERSIANN-CDR, CHIRPS, GPM, GSMaP and ERA5 for the years 2017, 2018, 2019

### 4.3.4 Evapotranspiration (ET)

The MODIS sensor aboard the Terra satellite is a passive instrument, recording information of earth's surface over 36 spectral bands with a resolution between 0.25 - 1 km. The global 8-day MODIS (MOD16A2) Evapotranspiration (ET), 0.5 km resolution product is used for this study and was accessed via the GEE data catalog. The ET is a term describing the combined process of evaporation (from the ground surface) and transpiration (from vegetation), to account for the upward vertical water fluxes from vegetated and non-vegetated land-surfaces. The ET values are provided in 0.1 kg/m²/8-days. The average daily ET is derived from the total ET over 8-days and accumulated over each month.

# 5

# METHODOLOGY

This chapter provides and overview of the complete research structure, explains how the data and processing tools were used for to classify the images, assess the accuracy of the classifier, and carry out the water-balance analyses. Figure 5.1 below shows a schematisation of the entire research process, the break-down of Phase-I and Phase-II is given in Section 5.2 and Section 5.3, respectively. An explanation of how the the land-cover classes are analysed, sampled and classified and how the accuracy of the classification is assessed is given in Section 5.2. Section 5.3 outlines how the results from Phase-I are combined with the hydro-meteorological datasets for the final discussion. Details on the input data and pre-processing steps can be found in Chapter 4.

## 5.1 OVERALL RESEARCH STRUCTURE

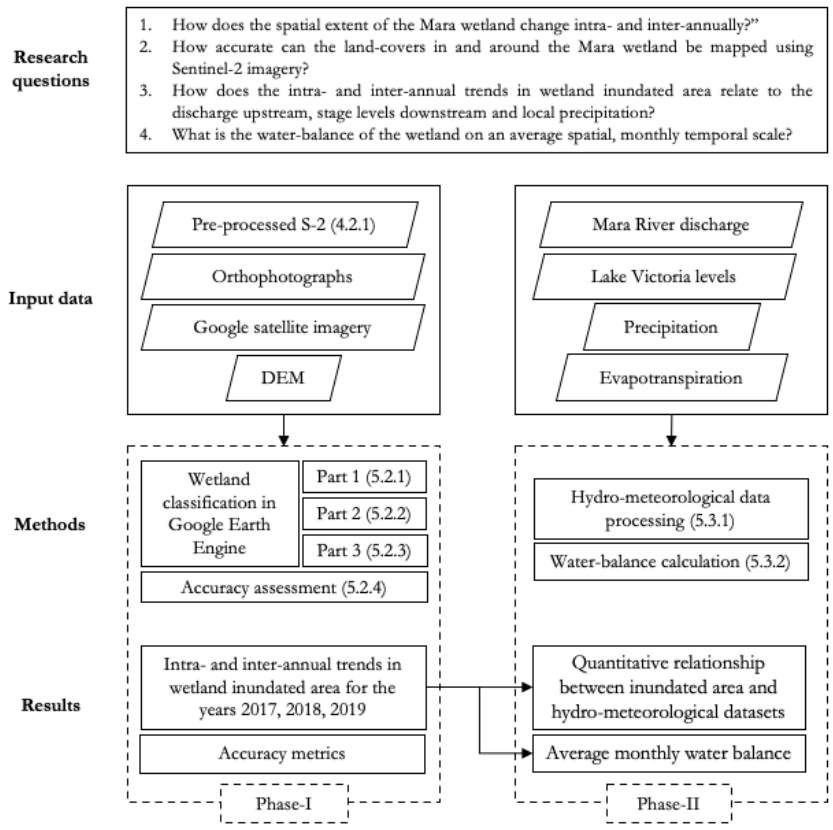



**Figure 5.1:** Flow chart of the overall research structure

## 5.2 PHASE-I: MULTITEMPORAL IMAGE CLASSIFICATION

This section covers the steps carried out in GEE to map the spatial and temporal dynamics of the land-cover classes in and around the Mara Wetland using remote sensing data. This phase is subdivided in three

parts. Part-1, outlined in Subsection 5.2.1, consists of the preliminary land-cover analyses and explanation of how the classification scheme is set up. Part-2 is described in Subsection 5.2.2 and provides details on how the training data is collected, how the topographic threshold is applied and how the classification is carried out to classify all of the images in the image collection. Part-3 consists of post-classification steps and the accuracy assessment, which are described in Subsection 5.2.3 and Subsection 5.2.4.

### 5.2.1 Land-cover analysis

This section describes Part-I of the classification procedure. This includes an explanation of how the land-cover classes are identified, sampled and analysed. How this translates to the design of the final classification scheme is explained at the end of the section.

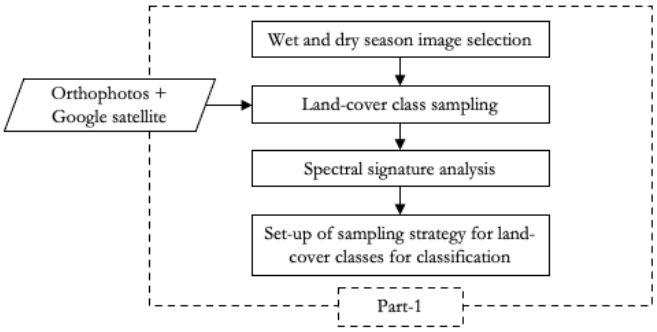



**Figure 5.2:** Flow chart of the steps carried out for the land-cover class analysis

*Land-cover classes identified in literature*

The selection of a suitable classification procedure and classifiers to distinguish between the land-cover classes of the region requires an understanding of which land-covers are present. Regrettably, this information cannot be obtained through field visits. Therefore, preliminary literature research is conducted and expert knowledge is obtained to gather information on the dominant species identified in previous studies. Land-cover classes around the wetland include agriculture, Savannah, forests, few urban areas, and bare ground. Table 5.2.1 gives an overview of classes recognized in literature (Mutie et al., 2006; USAID, 2018; Dessu and Melesse, 2012; Mati et al., 2008).

**Table 5.1:** Land-cover classes (from literature)

| **Land cover** | **Characteristics** |
|---|---|
| Wetland | |
| Open water | Open water without plant cover |
| Cyperus papyrus | Dense, fast-growing, rooted/free-floating aquatic species |
| Phragmites australis | Invasive, perennial, aquatic grass species |
| Typha domingensis | Emergent aquatic macrophyte |
| Upland | |
| Agriculture | Dryland cropland and pasture |
| Bare-land | Dry grass or no vegetation |
| Urban | Settlements, houses |
| Forest | Broadleaf forest, terrestrial vegetation |
| Savannah | Mixed wood-land grass-land ecosystem |

*Bi-seasonal image date selection*

Natural land-cover classes can change significantly over the year due to phenological events. Wetlands are highly dynamic and complex ecosystems due to the variability of source water, which influences the growth, decay and distribution of plants in and around the wetland (Dutton et al., 2018). Therefore, the wetland classes are analysed bi-seasonally (samples are taken for both the wet and dry season), in order to capture possible changes in greenness or wetness and thus changes in spectral reflectance within the same class. The wet and dry state images are taken from the year 2017 because of the availability of training data for that year. The reference image dates used are given in Table 5.2. The 'dry season' and a 'wet-season' images are chosen on the basis of capture date and cloud-coverage. The image must be acquired in cloud-free conditions and the 'dry season' image must have been captured around the same date as the UAV dataset (early November 2017). The resulting 'dry-season' image is from 3rd of November, acquired at the end of the dry period, just before the onset of rain (which marks the start of the climatic year). The 'wet season' images are from the 11th of June 2017, near the end of the rain-season.

**Table 5.2:** Image dates used for land-cover analysis

| Season | Date |
|---|---|
| Wet | 11-06-2017 |
| Dry | 03-11-2017 |

*Land-cover class sampling*

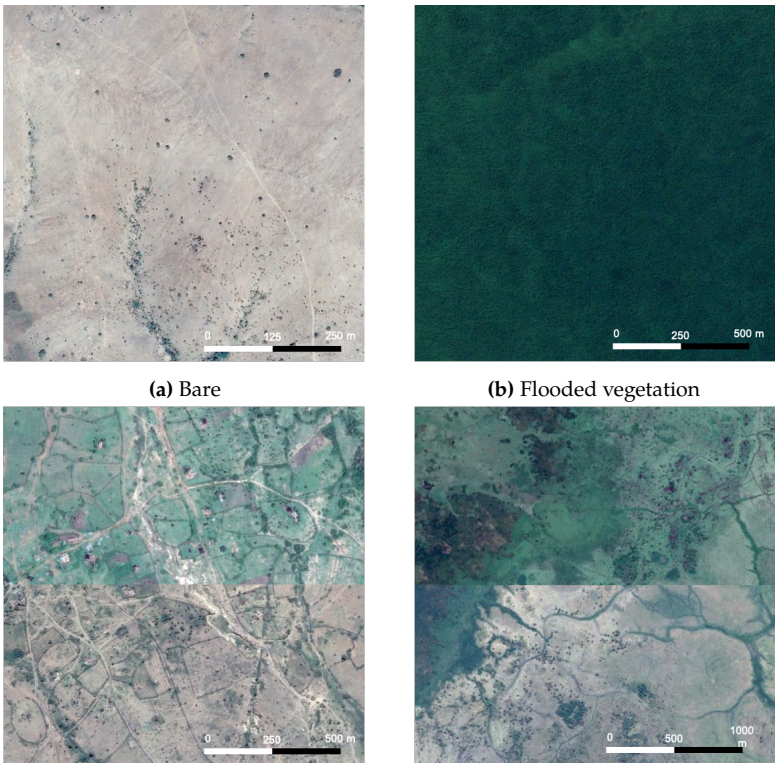
Drone (UAV) footage (see Figure 4.4), a historical time-series of Google satellite images and expert knowledge of the region are used as a reference for the sampling process. The land-cover classes sampled for the classification procedure are not precisely as listed those in . The 'urban' class is disregarded because this class is rare, and in the form of scattered homes and small roads, and thus spectrally indistinguishable. Furthermore, it was not possible to sample the flooded vegetation classes according to the species type, and is thus generally sampled as 'flooded vegetation'. However, the class is subdivided into two individual classes (FV1 and FV2) according to the different spectral reflectance revealed by the NDVI map. The 'forest' areas are not included in the classification because the upland areas where this class occurs are masked (see Figure 5.7).

Moreover, several additional classes are added in light of the classification goal; to distinguish between the 'wet' wetland and the 'dry' upland areas. The presence of water is variable in both space and time, thus several of the wetland and upland classes exist in both 'dry' and 'wet' states. The presence of moisture has an influence on the spectral characteristics of the land-cover class. Therefore, the land-covers are considered as different classes depending on the state of inundation. The resulting land-cover classes are listed in Table 5.3. Figure 5.3 shows satellite images of four of the land-cover classes in the dry and wet season. The samples are manually placed and labelled by drawing polygons in GEE and labeling these according to the land-class they represent. Lastly, dark brown patches, which appear in various shapes and sizes in the wetland, are the result of uncontrolled burning activities to create fertile pasture or to open-up land for cattle grazing (USAID, 2018). This class is sampled for the spectral signature plots (see Figure 5.4 and Figure 5.5), but not for the classification procedure because of the sporadic occurrence. Section B.3 in Appendix B contains examples of the spectral response of land-cover classes in time.

*Spectral signatures*

The outcome of the delineation process relies on a difference in spectral response for the different band-s/indices from S2 satellite image between the land cover classes. The band and index reflectance/sensitivity is analysed for various land-cover features for the wet and dry season, to identify suitable indices and combinations to best delineate the inundation of the wetland. Figure 5.4 shows spectral signature of the land-cover classes for the month June and November. Figure 5.5 show the index reflectance of land-cover classes for the month November and June. The median values of the band and index reflectance of the land-cover class samples can be found in Appendix A (Table C.2 and Table C.1). A detailed description of the spectral characteristics of the land-cover classes is provided in Table 5.3.

The eastern edge of the wetland is where the fluctuations in flooded area are most clearly visible when viewing a historical time-series of satellite images. The edge not only shifts position, but the land-cover of




**(a)** Bare **(b)** Flooded vegetation

**(c)** Agriculture (wet/dry season) **(d)** Floodplain (wet/dry season)

**Figure 5.3:** Google satellite images of land-classes used as a reference for collecting training data. The top and bottom half of figures (c) and (d) are the labeled class in wet and dry season, respectively

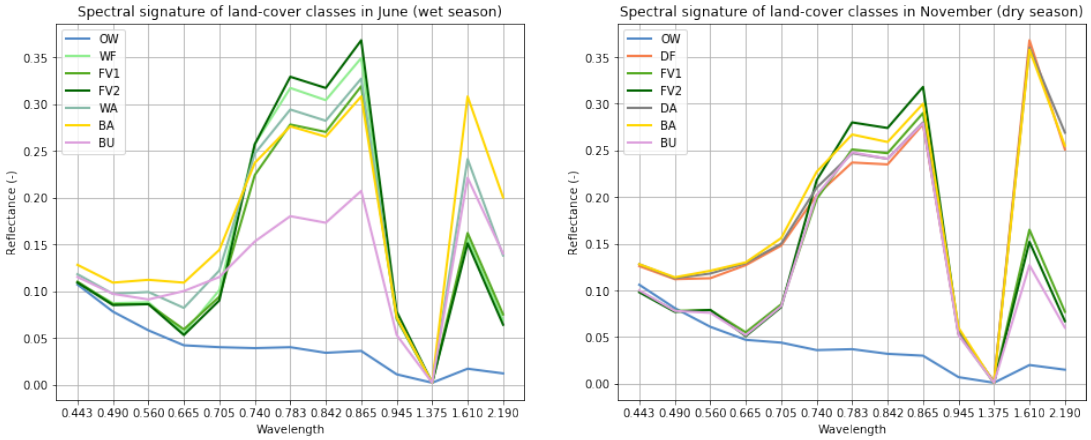



**Figure 5.4:** Spectral signature of land-cover classes sampled in June (left) and November (right)

the wetland at this moving edge is also continually undergoing change. The variation in water level due to seasonal patterns, and the simultaneous variation in water availability to soil and vegetation in this area induces the growth and decay of species.

The MNDWI, NDWI and NDVI show much higher sensitivity to open water features in both June and November than the NDII and NDMI, which show the same response for the open water and papyrus classes. The difference in response between papyrus and upland vegetation is quite clear for the NDII, NDVI and NDMI in November. These same two classes are, however, not easily distinguishable in June. Appendix A shows visualizations of the wetland area based on each of the indices. The MNDWI, NDWI and NDVI shows contrast between wetland vegetation and open water. The NDMI and NDII show similar response for the entire wetland area (open water and vegetated), but show more contrast between papyrus swamp and upland areas.

**Table 5.3:** Land cover classes based on Google Satellite and UAV Imagery

| **Land cover** | **Description** | **Spectral characteristics** |
|---|---|---|
| Wetland | | |
| Open water (OW) | Water with little to no vegetation coverage and varying in depth. Includes wide river reaches, deep water at the outlet to Lake Victoria and shallow seasonally inundated floodplain | Very low reflectance over the visible ($\leq$ 0.7 $\mu$m, B1-B3), NIR (B4-B8A), and SWIR (B11-B12) regions. High reflectance for NDWI, MNDWI and NDMI indices. Negative values for NDVI. |
| Flooded vegetation (FV) | Pixels containing inundated areas mostly covered in emergent vegetation, including dense papyrus swamp, water-logged agricultural land surrounding the main wetland and water-logged riparian vegetation. | Low reflectance in the visible range and for $SWIR_1$/$SWIR_2$ (due to penetration through organic matter and absorption by water) (Lefebvre et al., 2019). High NDII and NDMI due to sensitivity to the moisture levels in vegetation. High values for NDVI (above 0.6, indicates the highest possible density of green leaves). |
| Wet season floodplain (WF) | Sparsely vegetated, water-logged floodplain areas. | Low reflectance for $SWIR_1$/$SWIR_2$ and slightly lower NDVI than flooded vegetation due to lower reflectance in the NIR region caused by lower density of green leaves. High NDII and NDMI due to sensitivity to the moisture levels in vegetation, but slightly lower than flooded vegetation class. High NDVI, due to presence of green vegetation, but again slightly lower than flooded vegetation class. |
| Upland | | |
| Dry season floodplain (DF) | Dry floodplain areas with little to no vegetation coverage | High reflectance over the visible and SWIR regions due to absence of moisture and reflection by ground surfaces. Lower reflectance in the NIR region relative to vegetation classes due to the lack of chlorophyll. Lowest reflectance for the NDII, NDMI, NDVI and MNDWI indices due to lack of moisture and lack of green vegetation. |
| Dry season agriculture (DA) | Fallow land with little or no green vegetation | Very high reflectance over the visible and SWIR regions due to absence of water. Lower reflectance in the NIR region relative to green vegetation classes. Very low reflectance for the NDII, NDMI, NDVI and MNDWI indices. |
| Wet season agriculture (WA) | Cultivated land (crops or forestry) or pasture covered in green vegetation, where vegetation type is variable. | Slightly higher reflectance in $SWIR_1$/$SWIR_2$ than inundated vegetation classes due to absence of water. Lower reflectance (than inundated vegetation classes) for NDII, NDMI, NDVI and MNDWI due to lower moisture levels. |
| Bare (BA) | Dry upland areas with no vegetation, patches of terrestrial vegetation or few shrubs | Similar to dry season agriculture; very high reflectance over the visible and SWIR regions due to absence of water. Lower reflectance in the NIR region relative to dry season agriculture due to greater absence of vegetation. Very low reflectance for the NDII, NDMI, NDVI and MNDWI indices. |

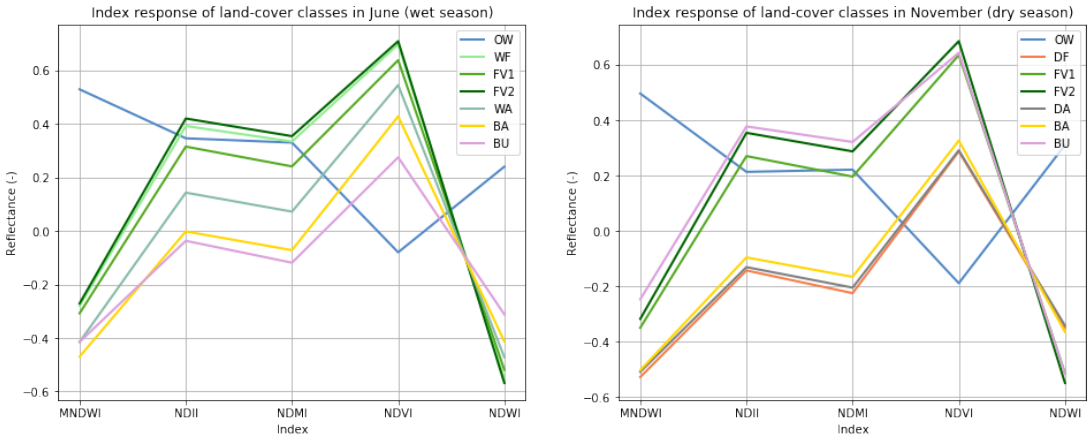



**Figure 5.5:** Spectral (index) response for land-cover classes sampled in June (left) and November (right)

### 5.2.2 Random Forests classification

This section describes Part-2 of the classification procedure. This includes an explanation of how the training data is collected, the masking of upland areas from using a DEM and a description of the Random Forests classification model. Figure 5.6 shows the flowchart of Part-2, including hierarchy of land-cover classes in the region.

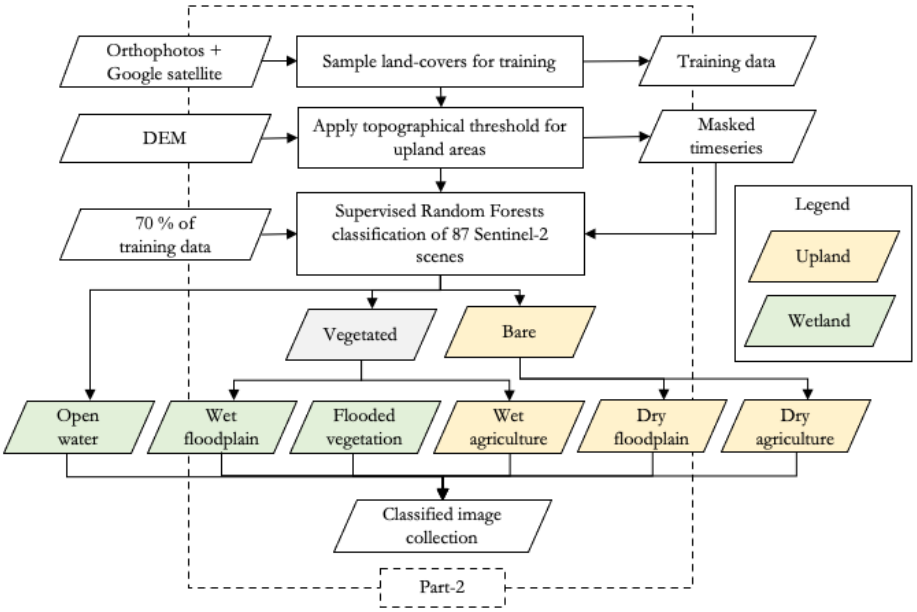



**Figure 5.6:** Flow chart of the steps carried out to collect training data and classify image collection

*Training data sampling strategy*

The orthophotographs available cover only a minor portion of the wetland area. Therefore, a historical time-series of Google Satellite images available using Google Earth Pro is used as an additional reference dataset to label the sampled land-cover classes. In order to produce an unbiased training dataset, the sampling strategy must ensure the proportions of the sample size for each of the land-cover classes are roughly representative of the actual land-cover proportions within the classification area (Millard & Richardson, 2015). Table 5.4 shows the estimated land-cover classes within the classification region and the corresponding final proportions of the samples taken for each of the land-cover classes. Note the map

proportions are rough estimates obtained through visual assessment, and will vary for each of the images. The times at which each of the training data samples are take are given in Table 5.5.

**Table 5.4:** Map and training data proportions

| **Land-cover class** | **Map proportion estimate** (%) | **Training data proportion** (%) |
|---|---|---|
| Open water | 4 | 3 |
| Flooded vegetation | 62 | 61 |
| Agriculture | 14 | 24 |
| Bare | 11 | 7 |
| Floodplain | 9 | 8 |

**Table 5.5:** Land-cover class sampling date

| **Land-cover class** | **Wet season image (06/2017)** | **Dry season image (11/2017)** |
|---|---|---|
| Open water (OW) | ✓ | ✓ |
| Flooded vegetation 1 (FV1) | ✓ | ✓ |
| Flooded vegetation 2 (FV2) | ✓ | ✓ |
| Bare (BA) | ✓ | ✓ |
| Floodplain wet (FW) | ✓ | |
| Floodplain dry (FD) | | ✓ |
| Agriculture wet (AW) | ✓ | |
| Agriculture dry (AD) | | ✓ |

The training data is split into two parts: 70% is used to train the algorithm and 30% is used as an independent validation dataset that is used in the accuracy assessment explained in Subsection 5.2.4. The input variables used for training the algorithm are; B4, B5, B6, B7, B8, B8A, B11, B12, NDVI, MNDWI, NDWI, NDII, NDMI and a digital elevation model. Details on the chosen input variables can be found in Table 4.2.1 and Table 4.2, for the bands and indices, respectively.

*Application of topographic threshold*

Topography is often a major control of the distribution of water in a landscape (Grabs et al., 2009). A Digital Elevation Model (DEM) is therefore used in addition to the spectral differences between land-cover features to distinguish wetland and non-wetland areas. The DEM is added as a band to each of the images in the image collection and is used to set a threshold that masks the upland areas (beyond the maximum extent of the wetland). By visual inspection, the upland areas are masked by setting a maximum threshold of 1185 m. Figure 5.7 shows the boundary of the unmasked area (black), which is used in the classification procedure.

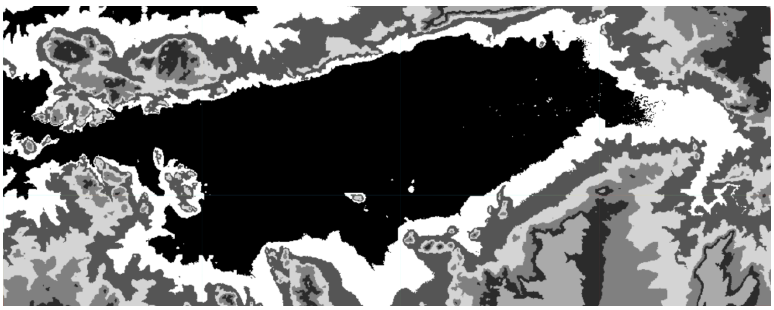


**Figure 5.7:** Contour map of the wetland area. The black region contains the wetland and the remaining part of the image is masked as upland areas.

### 5.2.3 Post-processing of classified imagery

This section describes Part-3 of the classification procedure (see Figure 5.9 for the steps). These steps are carried out after performing the first classification of the 87 S2 scenes that remained after filtering to date, tile, and maximum cloud cover percentage.

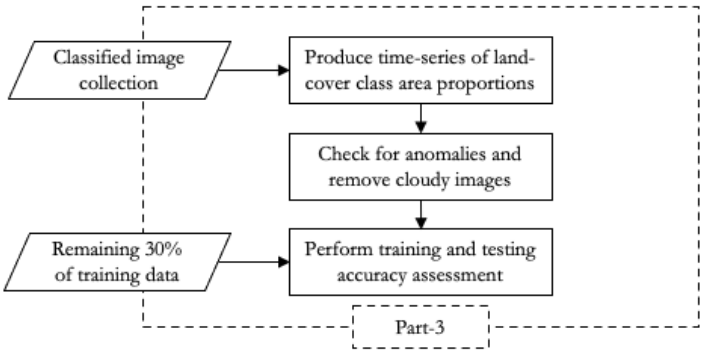



**Figure 5.8:** Post-classification steps and accuracy assessment

*Anomaly removal*

The total area classified for each of the land cover classes is calculated for each image and combined in a stacked area chart to visualise the seasonal land-cover area changes. A number of scenes show a sudden peak in a certain land-cover class. These images were revisited to check for potential cirrus cloud coverage that was not removed by the mask at during the pre-processing steps. The images that did reveal the presence of undetected clouds were manually removed from the image collection.

*Regional segmentation of flooded vegetation class*

The area is segmented study the occurrence of the 'flooded vegetation' according to the region of occurrence. The central part of the Mara Wetland forms an extensive, densely vegetated swamp. However, flooded vegetation also occurs beyond the main swamp and in the upstream reaches of the Mara River. A polygon that delineates the main swamp is added in order to distinguish between the class occurring within or beyond the main swamp. This is done in favor of a more detailed analysis of potential source waters.

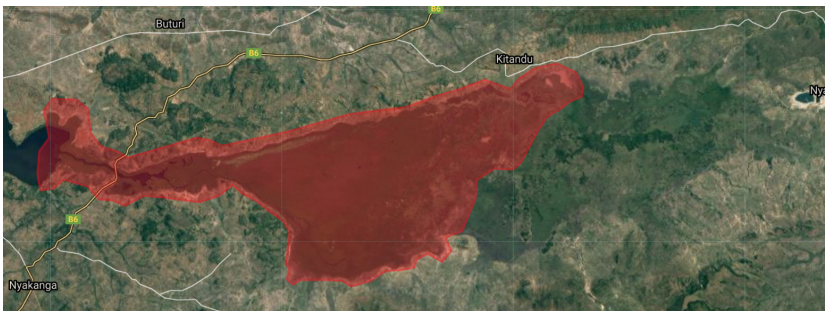



**Figure 5.9:** Main papyrus swamp delineation

### 5.2.4 Evaluation of area estimates

The two images from which the training data is sampled are evaluated by performing a conventional accuracy assessment (based on confusion matrices formed with the split training dataset to ensure an unbiased model evaluation). Ideally, the remaining area estimates (besides the estimates derived for the two training images) would be evaluated in the same manner. The lack of field observations or high resolution imagery acquired at the same time rules out the possibility to make a validation dataset. Therefore, the sequence of remaining images are evaluated according to area estimates from previous literature and the magnitude and timing of antecedent water fluxes.

*Accuracy assessment*

The sampled data is split into two; 70% is used for training and the remaining 30% of the sampled pixels that are not used to train the RF classifier are set aside for testing the performance of the classifier. A separate accuracy assessment is carried out for the training (re-substitution) and testing (validation) data. The accuracy is assessment is in performed in the GEE environment by computing a 2D confusion matrix for each of the datasets. The calculated metrics are the Overall Accuracy (OA), User's Accuracy (UA), Producer's Accuracy (PA) and Kappa coefficient. Figure 5.10 provides an illustration of the formulas used to calculate the accuracy metrics, where A, B, C, D resemble the land-cover classes. For example, $N_{AB}$ is the number of pixels classified as class B but are actually class A. $\sum_{An}$ and $\sum_{Ap}$ are the sum of the pixels *actually* in class A and the sum of pixels *classified* as class A, respectively.

**Predicted**

| **Actual** | **A** | **B** | **C** | **D** | **PA** |
|---|---|---|---|---|---|
| **A** | $N_{AA}$ | $N_{AB}$ | $N_{AC}$ | $N_{AD}$ | $\frac{N_{AA}}{\sum_{Ap}}$ |
| **B** | $N_{BA}$ | $N_{BB}$ | $N_{BC}$ | $N_{BD}$ | $\frac{N_{AA}}{\sum_{Bp}}$ |
| **C** | $N_{CA}$ | $N_{CB}$ | $N_{CC}$ | $N_{CD}$ | $\frac{N_{AA}}{\sum_{Cp}}$ |
| **D** | $N_{DA}$ | $N_{DB}$ | $N_{DC}$ | $N_{DD}$ | $\frac{N_{AA}}{\sum_{Dp}}$ |
| **UA** | $\frac{N_{AA}}{\sum_{An}}$ | $\frac{N_{BB}}{\sum_{Bn}}$ | $\frac{N_{CC}}{\sum_{Cn}}$ | $\frac{N_{DD}}{\sum_{Dn}}$ | - |

N = total number of points
n = actual number of points of class
p = number of points classified as class

**Figure 5.10:** Confusion matrix

$$\text{OA} = \frac{N_{AA} + N_{BB} + N_{CC} + N_{DD}}{N}$$

$$\text{K1} = \frac{\sum_{Ap}\sum_{An} + \sum_{Bp}\sum_{Bn} + \sum_{Cp}\sum_{Cn} + \sum_{Dp}\sum_{Dn}}{N^2}$$

$$\text{Kappa} = \frac{OA - K1}{1 - K1}$$

*Antecedent water fluxes*

The lack of extensive ground-truth data limits the ability to validate the classification results over the whole study period. An alternative evaluation method to nonetheless evaluate to the area estimates resulting from the classification procedure is discussed below.

The seasonal inundation patterns are assumed to be (at least in part) resulting from the fluctuations of (a combination) of local precipitation, upstream discharge and water levels of Lake Victoria (LV). The magnitude of peak area estimates are thus expected to show a positive correlation with the antecedent magnitudes of the aforementioned components. This hypothesis is tested by comparing the peak inundation areas with the accumulation of precipitation and discharge, and the change in the water level of LV over the three months preceding the peak date. The result of one or more of these comparisons are expected to show a positive correlation, and if so, will add confidence to the area estimates obtained from the classification procedure.

The discharge is acquired from the Mara Mine (MM) gauging station (see Figure 4.6). The discharge volume is accumulated over the 3 months preceding the peak inundation date, the precise months are not fixed due to the inter-annual variability of the peak inundation. The (satellite-based) precipitation datasets used are PERSIANN-CDR and CHIRPS, both products provide daily rainfall estimates in mm. The average of the two datasets is compared with the peak inundation area and the range between these used to estimate the uncertainty. The spatial average precipitation is taken over the wetland area (with a buffer of approximately 2 km to account for infiltration and runoff around the main wetland. Only the local rainfall average is calculated with the assumption that the precipitation over LV is represented by the water level and the precipitation over the upstream areas of Mara Basin is similarly included in the discharge at MM. The increase in lake level is calculated as the difference between the lake level at the moment of peak inundation and that of 3 months prior.

## 5.3 PHASE-II: HYDROMETEOROLOGICAL ANALYSES

In Phase-II of the research, the time-series of the inundated area computed in Phase-I is correlated with the magnitudes of surrounding water balance components (discharge upstream and lake water level downstream). Furthermore, Subsection 5.3.2 describes the method employed to compute the monthly water balance.

### 5.3.1 Stage- and discharge-surface area correlations

The acquisition dates of the classified scenes are extracted from the classification time-series. The hydrometeorological measurements from these dates are plotted against the total inundated area (open water, flooded vegetation, wet floodplain). A travel time of 2 days (from the discharge gauge to the wetland) is taken into account for the discharge correlations (Bregoli et al., 2019).

### 5.3.2 Waterbalance calculation

Assuming the following conceptual framework for the wetland water-balance:

$$\Delta S/\Delta t = S_i + AP_{in} + Q_{in} + L_{in} - AE_{out} - Q_{out} - L_{out}$$

Where:

$S_i$ = initial wetland storage
$A$ = wetland area
$P_{in}$ = net precipitation
$Q_{in}$ = surface water inflow (Mara River)
$Q_{out}$ = surface water outflow
$E_{out}$ = net evaporation
$L_{in}$ = lateral inflow (groundwater)
$L_{out}$ = lateral outflow (groundwater)

The water balance is computed in the form of a time-series due to the absence of bathymetry data required to compute the initial storage ($S_i$). Furthermore, water balance is computed on a monthly basis, considering the average frequency of area estimates is 1-2 per month. The water balance of the wetland estimated by accumulating the monthly (known) inflows (discharge and precipitation) and outflows (evapotranspiration). An average inundation area is assumed for each given month, and linearly interpolated for the months with no area estimates. The lateral inflow and outflow is not quantified in this study because of the lack of groundwater data. Based on speculative assumptions from previous studies, the groundwater may function as water storage and/or baseflow for the wetland (Bregoli et al., 2019).

The regional average evapotranspiration (ET) is derived over the years 2017, 2018 and 2019 (see Subsection 4.3.4 for a description of the data used). The regions used are shown in Figure 5.11. A time-series of the average ET over the main swamp can be derived because of its permanence (a once-defined region can be used to extract data over the study-period). Due to the dynamic nature of the remaining classes, the average ET could not be derived in the same way. Thus, the total average ET is assumed as the average between the ET from the main swamp (dark green) and the total wetland area (light green). The range of the two represents is used as the uncertainty. The ET is converted to volumetric units using the inundated area estimates and are accumulated over each month. The discharge volume is derived from the discharge rate and accumulated over each month. The components are combined in a graphical form to illustrate the net monthly water availability (additional to the unknown initial storage). The lateral (groundwater) inflow and outflow are neglected in the computation, but considered as potential drivers in the discussion. The total uncertainty range is computed as the summation of the uncertainty from the precipitation and the evaporation.

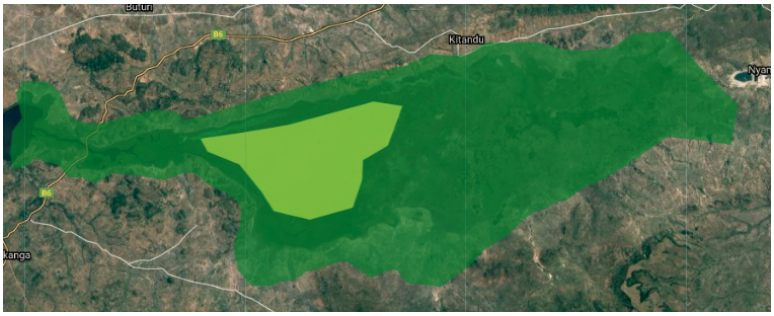



**Figure 5.11:** The delineation of the total wetland area (dark green) and permanent swamp (light green)

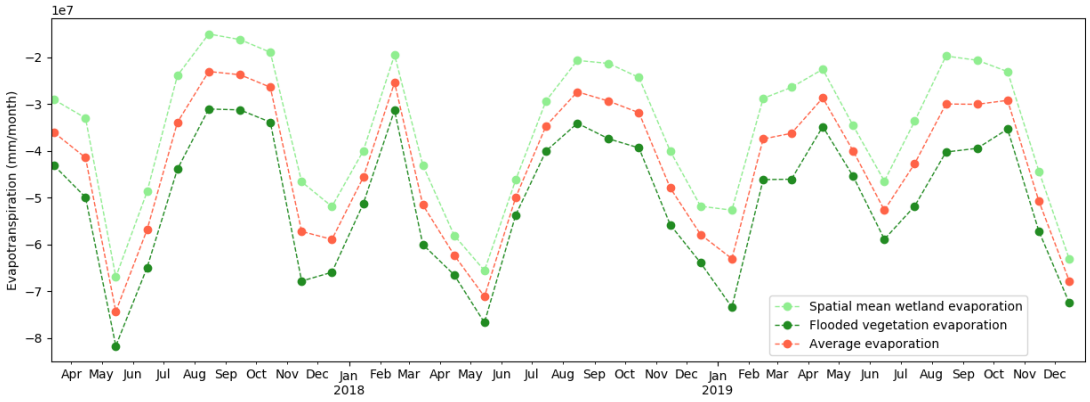



**Figure 5.12:** Time-series showing monthly accumulated evapotranspiration for the whole wetland area (dark green), the flooded vegetation swamp (light green) and the average of these (red)

# 6

# RESULTS

This chapter presents the main results related to the research questions. The results from the Phase-I classification procedure are provided in Section 6.1. The results from the evaluation of the area estimate are given in Subsection 6.1.4 and Subsection 6.1.5. The Phase-II hydro-meteorological analysis results are given in Section 6.2. The results are discussed in the following chapter (Chapter 7).

## 6.1 PHASE-I: MULTITEMPORAL WETLAND CLASSIFICATION RESULTS

Herein the results are provided for the Phase-I classification procedure outlined in Section 5.2.

### 6.1.1 Time-series of land-cover class area estimates

Figure 6.1 shows the areas of the land-cover classes within the study region. Linear change is assumed in between observations solely to support visual interpretation. Only scene-derived area estimates are used in the quantitative analyses. The small dips in total area for a few images is due to the application of a cloud-mask. The bottom four classes in the chart (Flooded vegetation (swamp), Open water, Wet floodplain, Flooded vegetation) are considered inundated wetland area, the remaining classes are considered dryland. The maximum total inundated wetland area (sum of flooded vegetation, open water and wet floodplain) is 707 km² and minimum of 294 km². The statistics of the wetland land-cover classes (open water, flooded vegetation, wet floodplain) are provided in Table 6.1. See Table C.3 in Appendix B for a complete list of area estimates.

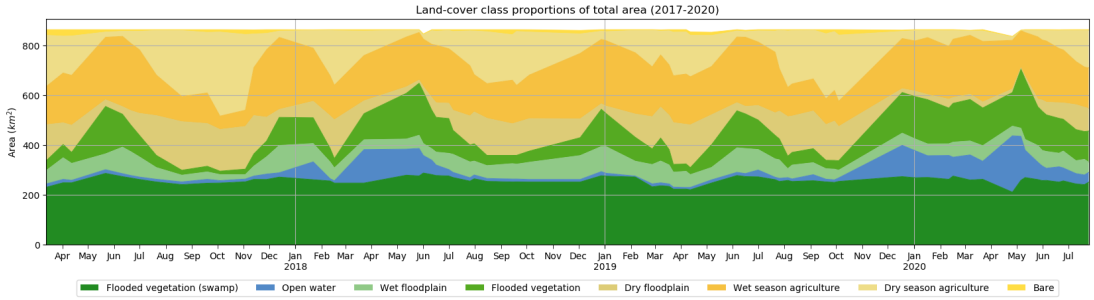



**Figure 6.1:** Time-series stacked area plot of the land-cover classes resulting from the classification procedure

**Figure 6.2:** Link to a video showing the spatiotemporal dynamics of the land-covers of the Mara Wetland for all 73 classified scenes. Access by clicking or scanning the QR code.

### 6.1.2 Classification of sampled images

Figure 6.3 shows the classification result of the sample image for June (11/06/2017). Figure 6.4 shows the classification result of the sample image for November (03/11/2017).

**Table 6.1:** Wetland class area statistics

| Land-cover class | Minimum | Maximum | Mean + Sd |
|---|---|---|---|
| Open water (OW) | 3.71 | 226.34 | 38.67 ± 44.34 |
| Flooded vegetation (FV) | 248.06 | 512.38 | 357.71 ± 68.48 |
| Wet floodplain (WF) | 18.04 | 109.73 | 60.70 ± 20.74 |

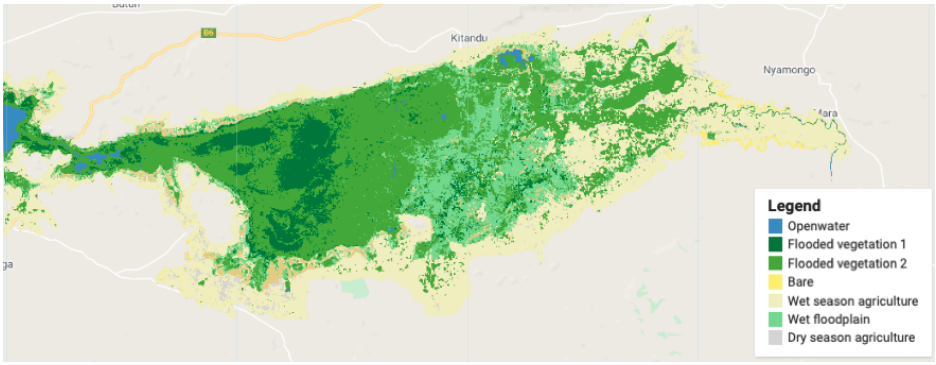



**Figure 6.3:** Classified map for June training image (11-06-2020)

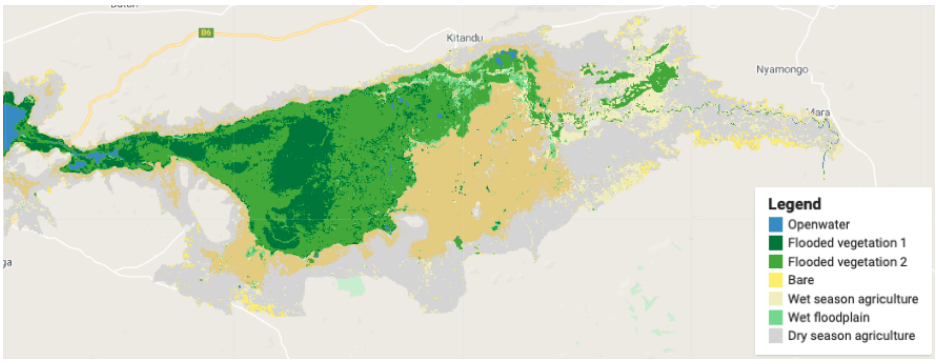



**Figure 6.4:** Classified map for November training image (03-11-2020)

### 6.1.3 Land-cover class occurrence frequency

Figure 6.5 shows the frequency of pixels classified as the open water over the study period. The scale runs from 0 to 100 % of the time, the map shows pixels with a low frequency of inundation as transparent red and those with 100 % frequency of inundation as opaque blue.

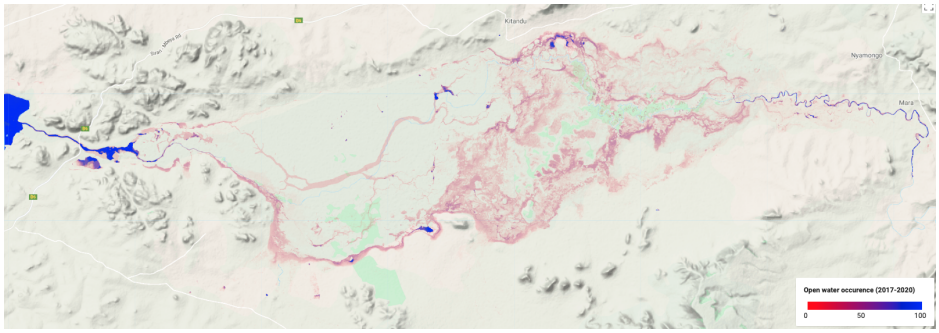



**Figure 6.5:** Open water covered area frequency map for 2017-2020

Figure 6.6 and Figure 6.7 show the frequency of pixels classified as flooded vegetation and wet floodplain, respectively. The opacity of green indicates the percentage occurrence, transparent for zero occurrence and green for pixels permanently classified as the respective class.

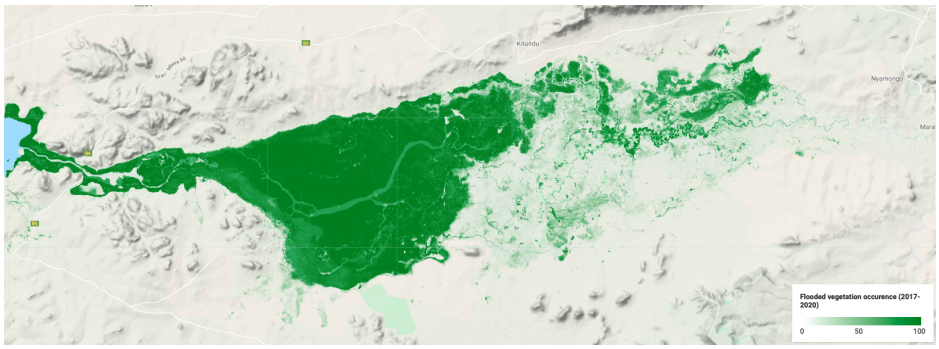



**Figure 6.6:** Flooded vegetation covered area frequency map for 2017-2020

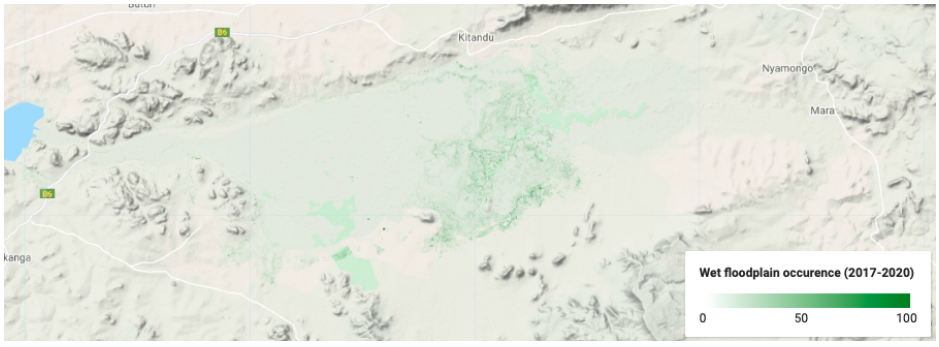



**Figure 6.7:** Wet floodplain frequency map for 2017-2020

### 6.1.4 Accuracy assessment

The overall training (re-substitution) accuracy is 99.9% for the classification of all land-covers (see Figure 6.8 for the confusion matrix and Table 6.2 for accuracy statistics). The overall accuracy derived from an independent validation set is 98.6% (see Figure 6.9 for the error matrix and Table 6.3 for accuracy statistics). Note the Flooded vegetation 1 and Flooded vegetation 2 classes are combined, as these belong to a single class. The 'bare' and 'agriculture dry' classes are similarly combined.

*Training accuracy*

| Actual \ Predicted | OW | FV1 | FV2 | BA | AW | WF | AD | DF | PA |
|---|---|---|---|---|---|---|---|---|---|
| **OW** | 12037 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 100.0 |
| **FV1** | 0 | 71134 | 933 | 0 | 0 | 0 | 0 | 0 | 100.0 |
| **FV2** | 0 | 667 | 122125 | 0 | 1 | 10 | 0 | 0 | 99.99 |
| **BA** | 0 | 0 | 0 | 35966 | 52 | 0 | 153 | 0 | 99.86 |
| **AW** | 0 | 0 | 0 | 88 | 20377 | 1 | 13 | 7 | 99.47 |
| **WF** | 0 | 2 | 51 | 0 | 0 | 8536 | 0 | 1 | 99.37 |
| **AD** | 0 | 0 | 0 | 1481 | 18 | 0 | 21796 | 5 | 99.90 |
| **FD** | 0 | 0 | 0 | 0 | 3 | 0 | 12 | 15481 | 99.90 |
| **UA** | 100.0 | 100.0 | 99.96 | 99.77 | 99.64 | 99.87 | 99.89 | 99.92 | - |

**Figure 6.8:** Confusion matrix representing re-substitution accuracy (training accuracy). OW = Open water, FV1 = Flooded vegetation 1, FV2 = Flooded vegetation 2, BA = Bare, AW = Agriculture wet, WF = Wet floodplain , AD = Agriculture dry, DF = Dry floodplain, PA = Producer's accuracy, UA = User's accuracy

**Table 6.2:** Overall accuracy and Kappa coefficient for training data

| | **Combined classes*** | **Individually sampled classes** |
|---|---|---|
| Total classified | 310950 | 310950 |
| Total correctly classified | 310686 | 307452 |
| Overall accuracy (%) | 99.9 | 98.9 |
| Kappa | 1.00 | 0.99 |

* FV1 = FV2 and BA = AD

*Testing accuracy*

| Actual \ Predicted | OW | FV1 | FV2 | BA | AW | WF | AD | DF | PA |
|---|---|---|---|---|---|---|---|---|---|
| **OW** | 4945 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 99.96 |
| **FV1** | 0 | 25692 | 5069 | 0 | 0 | 8 | 0 | 0 | 99.97 |
| **FV2** | 0 | 3498 | 48982 | 0 | 16 | 155 | 0 | 0 | 99.68 |
| **BA** | 0 | 0 | 0 | 14168 | 340 | 0 | 1037 | 0 | 97.81 |
| **AW** | 0 | 0 | 21 | 450 | 8179 | 6 | 85 | 7 | 93.50 |
| **WF** | 0 | 28 | 467 | 0 | 3 | 3147 | 0 | 5 | 86.22 |
| **AD** | 0 | 0 | 0 | 1961 | 90 | 0 | 7867 | 61 | 98.49 |
| **DF** | 0 | 3 | 1 | 0 | 12 | 8 | 46 | 6609 | 98.95 |
| **UA** | 100.00 | 99.89 | 99.10 | 97.29 | 94.66 | 94.68 | 98.55 | 98.91 | - |

**Figure 6.9:** Confusion matrix for classification of independent validation data (testing accuracy). OW = Open water, FV1 = Flooded vegetation 1, FV2 = Flooded vegetation 2, B = Bare, AW = Agriculture wet, WF = Wet floodplain, AD = Agriculture dry, DF = Dry floodplain, PA = Producer's accuracy, UA = User's accuracy

**Table 6.3:** Overall accuracy and Kappa coefficient for validation data

| | **Combined classes*** | **Individually sampled classes** |
|---|---|---|
| Total classified | 132968 | 132968 |
| Total correctly classified | 131154 | 119589 |
| Overall accuracy (%) | 98.6 | 89.9 |
| Kappa | 0.98 | 0.87 |

* FV1 = FV2 and BA = AD

### 6.1.5 Prior area estimates

As shown in Table A.1, the wetland area estimates for the years 2017 and 2018 range from 388 to 517 km². Due to the significant changes the wetland has undergone for the past 100 years (especially in recent decades) (Bregoli et al., 2019), the other estimates (before 2017) are not considered representative of the wetland as it is for the study period (2017-2020). The peak area in this study is 707 km², a large part of this area is only temporarily wet due to the flood event occurring at acquisition time. Since the floodplain is

only covered in open water for a short period, a more representative figure for the maximum wetland area are the peak areas in the occasion there is no flooding beyond the main papyrus swamp. The peak area in those conditions are 555, 511, 543 and 537 km². The minimum wetland area is in the order of 300 km², with a minimum of 295 km².

### 6.1.6 Antecedent water fluxes

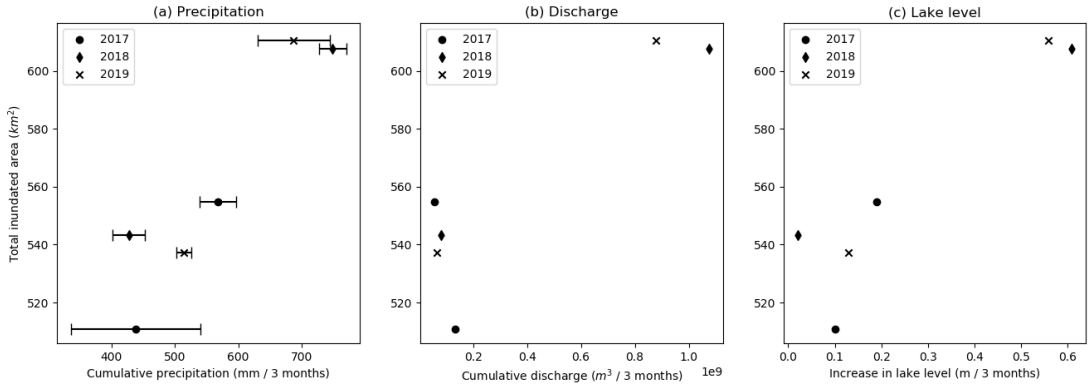



**Figure 6.10:** Relationships between the peak inundated area (2/year for the years 2017, 2018, 2019) and change in local water fluxes over the 3 months preceding the peak date. (a) cumulative precipitation (average between CHIRPS and PERSIANN-CDR and error indicates range between the data-sets), (b) cumulative discharge amount and (c) difference in Lake Victoria water level between the peak-flooding date and that of 3 months prior

## 6.2 PHASE-II: HYDRO-METEOROLOGICAL ANALYSES RESULTS

This section contains the correlation charts between the aerial extent of inundated area obtained in Phase-I and the available, local hydrological field-data (water levels of Lake Victoria (LV) as the downstream component and the discharge upstream for the same date as the area estimate). The average monthly inundated area is also calculated for each month available, and these are correlated with the average monthly precipitation over the entire the Mara Basin. Total inundated area is the sum of the open water, flooded vegetation 1 and 2 and wet floodplain classes.

Figure 6.11 shows the inundated area values against the water level in LV. Additional correlation charts for stage-area are provided in Appendix D. Figure 6.12 shows the total inundated area values against the discharge in $m^3/s$ upstream the wetland at the Mara Mine gauging station (see Figure 4.6). In Figure 6.13 the discharge against the open water area estimates. The monthly water balance of the wetland is given in Figure 6.14.

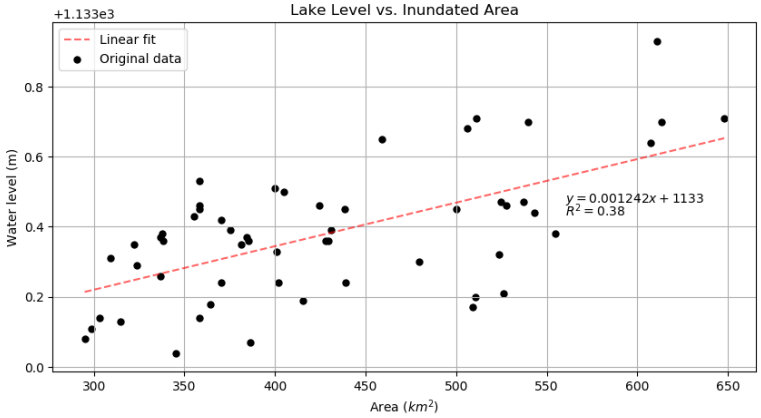



**Figure 6.11:** Correlation chart showing Lake Victoria water levels against inundated area

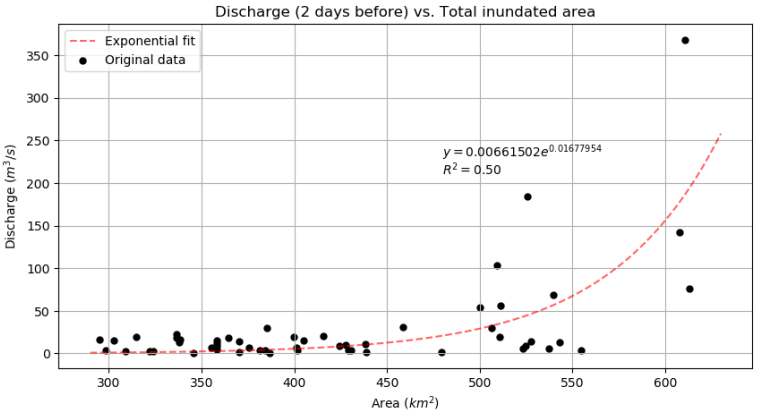



**Figure 6.12:** Discharge (2 days before date of area estimate) and estimated inundated area

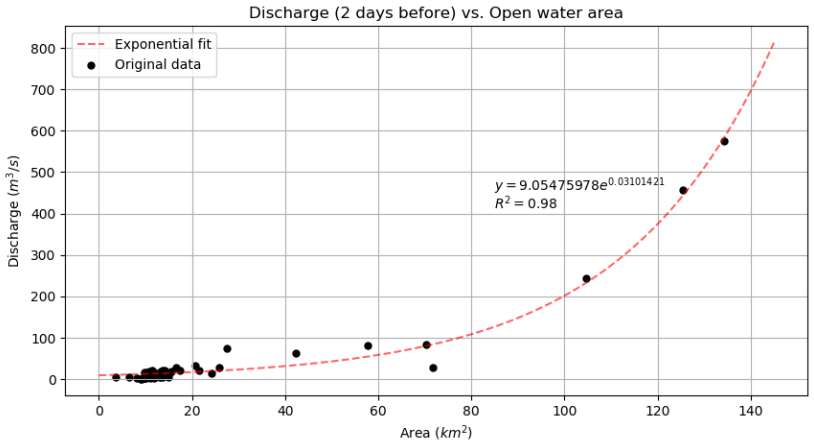



**Figure 6.13:** Discharge (2 days before date of area estimate) and estimated open water area

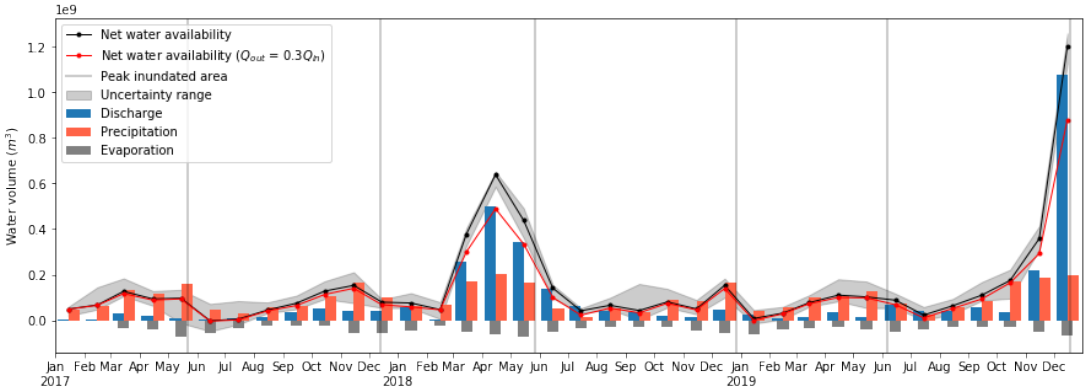



**Figure 6.14:** Monthly water balance of the wetland based on the net precipitation, discharge, and evaporative fluxes accumulated over one month. The uncertainty range is computed as the sum of the uncertainty from the precipitation and evaporation estimates. The vertical lines correspond to the dates of peak inundated area, estimated from the classification procedure. The red line is the water availability based on the assumption the outflow from the wetland into LV is 30% of the inflow from the Mara River (based on the estimation made by Bregoli et al., 2019).

# 7

# DISCUSSION

In this chapter the results presented in the previous chapter are examined, and the methodological approach to obtain the results is re-evaluated by considering the underlying assumptions and limitations. Section 7.1 discusses the classification procedure and the accuracy of the area estimates. In Section 7.2 the results from the hydro-meteorological analyses are examined. The last section (Section 7.3) explores the potential additional sources of error.

## 7.1 CLASSIFICATION (PHASE-I)

In this section the methods and results of the classification procedure (Phase-I) are discussed. In Subsection 7.1.1 the classification approach is reviewed, and the limitations of the chosen satellite, the classifier, the training process are raised. Subsection 7.1.2 contains a discussion of the accuracy obtained in the classification procedure.

### 7.1.1 Classification procedure

The optical sensors aboard the Sentinel-2 (S2) satellite are unable to penetrate cloud-cover nor detect water underneath the thick layer of aquatic vegetation, unlike L-band SAR, which can penetrate through both clouds and dense vegetation (White et al., 2015). The densely vegetated nature of the wetland contributed to the challenges in deriving an appropriate method to, regardless of the dense aquatic vegetation, distinguish between wetland and dryland areas. In this study the spectral characteristics of flooded vegetation are used to distinguish between wetland and dryland areas. The classification of flooded vegetation functions as a proxy for the presence of water. Using a threshold method to define flooded vegetation was not possible due to differing intra-class phenological phases of vegetation. The complexity of the land-covers calls for a method whereby the multiple characteristics can be taken into account simultaneously. The RF classification algorithm has been chosen because of its capability to process high dimensional, multi-source datasets to form a combination of tree-predictors. The RF algorithm is trained using 14 input variables; 5 indices, 8 bands and a DEM. The classifier is trained bi-seasonally, thereby accounting for the different phenological phase of vegetation, to effectively distinguish between vegetation classes independent of season. The classification procedure is semi-automated (a single training dataset is used to classify all scenes). The classification accuracy achieved (see Subsection 6.1.4) confirms the suitability and capability of using these methods to map vegetated wetland regions.

The intricacy of the interactions between water availability and vegetation in wetlands continues to be a significant challenge in meaningfully characterising and mapping wetland areas (Thomas et al., 2015). The forces controlling the water levels in wetlands, and thereby also the presence and diversity of species are both autogenic (internally driven, such as plant transpiration and plant water uptake) and allogenic (externally-driven, such as precipitation and temperature) (Koning, 2005). The moisture-vegetation feedbacks result in a heterogeneous, fast-changing landscape with indistinct classes and complex gradients (Adam et al., 2014). The classification process by definition involves generalization, and information loss when classifying graded landscapes is therefore inevitable. Pixels may contain multiple, spectrally distinct features which can lead to ambiguous classes and high confusion rate.

For this study the classes are chosen with the aim to distinguish wetland and dryland areas. The 'wet floodplain' class is chosen in recognition of a transitional zone between the main swamp and dryland. This class is likely to exhibit a significant amount of variability due to the fine-scale, heterogeneous/mixed complexion of water and vegetation in the landscape. Thus, although this land-cover has shown to be sufficiently spectrally unique for it to be successfully distinguished from other classes, there may still be significant differences in physical characteristics amongst classified pixels. The intra-class differences on the precise degree of waterlogging or precise vegetation type and vigor is not conveyed in the results, which limits the ability for a more detailed interpretation. The supervised classification approach executed

in this study requires manual selection of training data and thus contains a certain level of subjectivity. The necessary user input means this method is not founded entirely on the inherit differences within the landscape. There may be certain features in the landscape that have very distinct characteristics but belong to the same land-cover class in the results of this study. Image segmentation, fuzzy and object-based classification are examples of alternative, less bias-prone approaches to determining land-cover classes and selecting training data (Dronova et al., 2011). In addition, the results should be interpreted keeping in mind the aforementioned limitations of categorising graded landscapes, and that there will always be intra-class variability.

### 7.1.2 Classification accuracy

The results from the the accuracy assessment (see Subsection 6.1.4) demonstrate the consistency and effectiveness of the classifier in mapping the land-cover classes. The assessment carried out is, however, limited in its ability to validate the classification of all 73 scenes. The temporal proximity of the study period and inaccessibility of the wetland both contribute to the lack of coincident ground reference data necessary for a more rigorous accuracy metric. In addition, high-resolution, historical satellite imagery was found to be scarce and inconsistent with the acquisition dates of the S2 scenes. In the ideal case, misclassifications would be quantified with the use of field surveys conducted at several moments during the study period, coincident with the acquisition date of the S2 scenes. Without such validation datasets, the area estimates are further evaluated with an alternative method using the data available and reasoned judgement.

The seasonal expansion and contraction of the wetland clearly follows a bi-modal regime (see Figure 7.1). The timing of the peaks and troughs in aerial extent show exact correspondence to the lake level fluctuations and follow from plausible magnitudes of antecedent precipitation. These factors provide additional confidence in the classification in its ability to capture the extreme wet and dry states of the wetland, and thus the timing of the peaks in the area time-series. The accuracy of the *magnitude* of the aerial inundation extent is, however, more challenging to assess. The aerial extent of the inundation is not necessarily proportional to its volumetric equivalent, due to the irregular (and largely unknown) bathymetry of the wetland. Thus, an increase in area may not correspond to a proportional increase in water volume. Regardless of the bathymetry, the hypothesis remains that a larger peak in area is associated with a greater amount of water availability (detectable by a higher lake level, greater accumulated amount of antecedent discharge and/or local precipitation). Figure 6.10 shows a positive trend in at least one or more of the aforementioned waterbalance components, which again affirms the reliability of the classification results.

## 7.2 HYDROLOGICAL ANALYSES (PHASE-II)

In this section the methods and results of the hydrological analyses (Phase-II) are discussed. In Subsection 7.2.1 the overall approach of Phase-II is revisited, this includes a discussion of the underlying assumptions and their potential impact on the results. The overall limitations of the research methods are also explored. In Subsection 7.2.2 hydrological significance of the results from Phase-II of the research are discussed. The results from the the monthly water balance are explored in Subsection 7.2.3. Additional sources of error are discussed at the end of the chapter, in Section 7.3.

### 7.2.1 Approach, assumptions and limitations

The approach taken for the hydrological analysis of the system is an alternative to constructing a regular hydrological model. In the hypothetical case the Mara River would discharge directly into Lake Victoria (LV), a regular hydrological model (that accounts for varying lower boundary conditions) would suffice. Due to the complex geometry and irregular bathymetry of the wetland, a hydrological model of the Mara River must assume the beginning of the wetland as the downstream boundary. Without knowledge on the backwater curve from LV, the boundary conditions of the model cannot be estimated, and must be assumed to be uninfluenced by fluctuating water levels or must be (Stoop, 2017).

Instead of a model, the hydrology of the wetland is investigated by analysing the relationship between the spatiotemporal behavior of inundation and the local waterbalance components. Figure 7.1 shows the waterbalance components (lake level, discharge, precipitation, evapotranspiration) in parallel with the time-series of area estimates. In the approach taken for this study, the aerial extent of the inundated area is used as a proxy for the amount of water present in the wetland. This method inherently assumes a proportional relationship to area and water volume, which is not always the case due to the bathymetry of the wetland. The components that are quantified in this study are the upstream and downstream

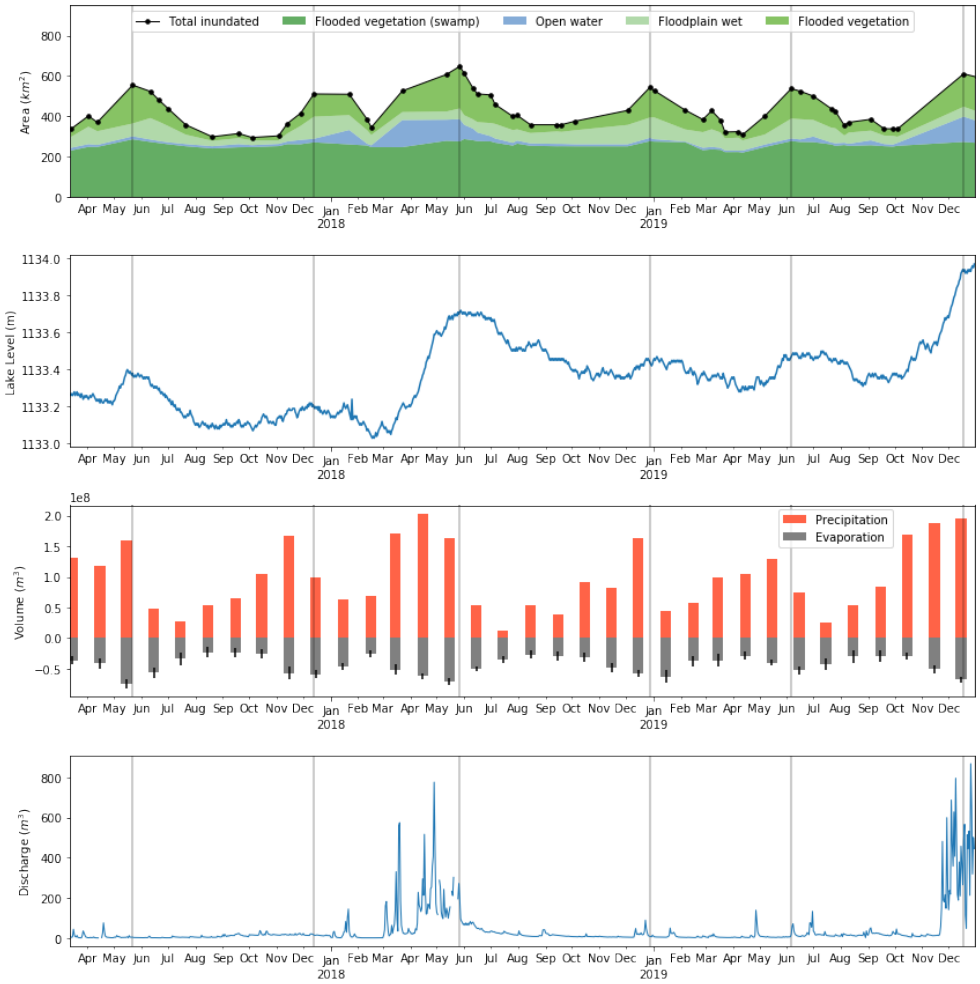



**Figure 7.1:** Time-series chart of wetland area and waterbalance components. The grey lines correspond to the inundated area peaks.

boundary conditions (Mara River discharge and LV water level) and the additional inflows and outflows (local precipitation and evaporation). Lateral groundwater inflow may well have an influence on the hydrology of the system, but could not be included due to the lack of data hereof. Similarly, the initial storage of the wetland is not quantified due to the absence of a bathymetry model. Lack of data on wetland water storage volume has also limited the waterbalance analyses to a monthly flux, rather than a depiction of the total water availability which includes the initial storage.

The temporal resolution of the area estimates are ascribed to the revisit frequency of the S2 satellite and the degree of cloud-coverage at the acquisition time. Rather than a continuous, daily time-series, which is usually the resolution of a hydrological model, the results provide a bi-monthly snapshot of the wetland. The classified pixels do not reveal by which process the land-covers appear. For example, flooded vegetation may appear due to the growth of vegetation from inundated areas, or it may appear due to the inundation of the already vegetated areas. These processes occur at very different time-scales. If the scenes had been captured at a higher frequency, the distinction between 'fast' and 'slow' processes could reveal more about the hydrological dynamics. A 'fast' process is, for example, the inundation of the wetland from an incoming flood. A 'slow' process is, for example, the growth of vegetation on the wet floodplain.

The influence of the discharge is analysed by assuming a travel time of 2 days from the gauge to the wetland. This assumption can only reveal how discharge contributes to the 'fast' processes. However, a persistent, but subtle increase in discharge may have long-term effects on the growth of wetland vegetation and this process would not be captured in these analyses. The waterbalance is calculated on a monthly basis which may, at least partly, attenuate the effects of processes occurring at different time-scales.

### 7.2.2 Wetland dynamics

*Intra- and inter-annual inundation pattern*

The aerial extent of wetland classes show high seasonal variability (see Figure 6.1). The seasonal flood pattern strongly follows the expected bi-modal (peaks occur twice yearly) regime. The timing of the peaks are, however, variable between years. The inter-annual variability is not of consequence for this study, since the hydrology is analysed by considering the waterbalance components (over the same period) in parallel. The area of the main flooded vegetation swamp remains relatively constant with a mean of 256 $km^2$ and standard deviation of 15.7 $km^2$. When excluding the peak flood events, the area of open water (no vegetation coverage) is on average about 13 $km^2$ with a standard deviation of 4.7 $km^2$. The spatial distribution of the fluctuations in wetland area are visualized in the animation provided in Figure 6.2. In the animation, as well as in the wetland occurrence frequency map (see Figure 6.6), one can see the wetland is comprised in part of a distinctive permanent swamp. The seasonal increase in wetland area beyond the permanent swamp occurs mostly upstream, eastern edge of the wetland, due to the topographical constraints near the outlet to LV. The seasonal emergence of wetland classes occurs partly as an extension of the main swamp. However, there are instances where the inundation springs up in areas that are seemingly isolated from the rest of the wetland. Rainfall is likely to play a role in inundating these specific areas, however, this could also be an indication of groundwater recharge fluxes.

*Lake Victoria stage - inundated area correlations*

The peak areas are indicated with vertical lines in each chart. The time-series of the water level of LV (Figure 4.7) show analogous fluctuations to the area estimates. The timing of the peak and troughs in the inundation extent and those of the water level of LV downstream are markedly concurrent, conveying the impression the level of the lake controls the seasonal fluctuations of the wetland. The *magnitude* of the extremes in area and lake level are, however, not clearly corresponding. For example, the first two peaks in area (05/2017 and 12/2017) and the 4th and 5th (12/2018 and 06/2019) are approximately the same (~580 km²). However, the lake level during the 4th and 5th peak is clearly higher than the first two. The antecedent rainfall and discharge for the 4th and 5th peak are not clearly lower than the first two. Thus, in the case the lake has influence beyond the main swamp, this is expected to show with higher 4th and 5th peaks. When plotting the magnitudes of inundation with lake level (see Figure 6.12 for the correlation chart), the relationship shows to be positive and linear. A possible explanation for the correlation is the influence of precipitation. The water of LV is estimated to source between 76% (Vanderkelen et al., 2018) and 85% (Obiero et al., 2012) from direct precipitation over the lake. The precipitation over the lake is subject to the same bi-modal rainfall regime as the that over the wetland (Vanderkelen et al., 2018). The bi-modal pattern of the stage levels of the lake are thus unsurprising, and are thereby not automatically linked to the inundation of the wetland. The flow of the Mara River is subcritical (Stoop, 2017), thus the backwater of the lake is likely to have at least some influence on the water availability of the wetland. Additional examination of the spatial distribution of the inundation is required to determine whether the congruence is indirect (due to both the wetland and the lake being influenced by the same rainfall regime) or direct (due to the backwater effect). The spatial pattern of the wetland (for a time-lapse video see Figure 6.2) reveals the seasonal expansion of vegetated wetland areas occurs mostly in the upstream reaches rather than as an extension of the main swamp on the LV side. This challenges the hypothesis that the coherence between inundation area and LV water levels are causal, since that would require the backwater from LV to reach over 40 km upstream.

*Mara River discharge - inundated area correlations*

The noteworthy correspondence between the area and the lake levels is not exhibited by the discharge hydrograph. The discharge rate is relatively consistent for most of the study period (with an average of approximately 11 $m^3/s$, excluding the peak flood events in March - June 2018 and November - December 2019). When converted to monthly accumulated discharge (see Figure 6.14), the seasonal pattern is more evident, but nonetheless not as distinct as that of the lake and of the precipitation. The correlation chart between inundated area and discharge upstream (Figure 6.12) shows an exponential relationship. However, the discharge for the area estimates below approximately 450 km² remains relatively constant. Figure 6.13 shows the correlation between the area estimate for open water and discharge upstream. For the majority of the data points (61/73) the open water area is below 20 km². The remainder of the points, however, follow a nearly perfect exponential trend. The wetland itself forms a depression in the landscape with a permanent swamp in the center. The landscape beyond the main swamp is approximately 0.035%

moving upstream (Bregoli et al., 2019). The flooding of a sloped landscape could justify the exponential increase in inundated area with greater discharge amounts. With a greater amount of discharge, the inundation will extend to the shallower reaches and thus spread over a larger area.

*Local precipitation - inundated area trends*

The local monthly accumulated precipitation time-series (Figure 6.14) follows the expected bi-modal regime, with peak months occurring twice yearly, in April/May and November/December. The relationship with precipitation and inundated wetland area is analysed on a monthly scale. Although the precipitation is available at a higher temporal resolution, the area estimates are only available, on average, twice monthly. The precipitation occurring precisely on the acquisition date of the area estimates could be misleading, as it does not account for the accumulated antecedent precipitation. The monthly water balance shows a consistent trend of high precipitation accumulation for three months preceding the peak wetland area. The influence of precipitation is further discussed below in Subsection 7.2.3.

### 7.2.3 Waterbalance

The monthly waterbalance (see Figure 6.14) visualises the quantifiable, monthly accumulated in and outflows of the wetland. Note the lateral groundwater inflow, overland flow, and backwater from LV are not included in the net water availability. The net outflow from the wetland into LV could not be determined because the gauge situated at the outlet of the wetland (Kirumi Bridge) is under the influence of the backwater effect. The outflow is estimated based on the discharge measurements from a field-study conducted in 2018 (Bregoli et al., 2019). The precipitation, discharge and evaporation are assumed to play a significant role in governing the waterbalance of the wetland. The timeseries of net water availability shows a clearly seasonal cycle, with a net positive water availability during the wet season and a net zero, or negative water availability in the dry season. This trend is mirrored in the occurrence of peak dates, which follow directly after three months of high net water availability. For only 7/36 months of the study period the accumulated discharge is greater than the accumulated precipitation, indicating local precipitation plays a significant part in the seasonal inundation pattern of the wetland. The Nyando Wetland, located approximately 160 km North-East of the Mara Wetland, similarly receives water from runoff, river discharge, direct precipitation, lake backwater flow and groundwater recharge. Isotope analyses have shown the major water source is direct precipitation (Obiero et al., 2012), suggesting the results from the calculated waterbalance in this study are plausible.

The seasonal contraction of the wetland demonstrated in the timeseries of area estimates (see Figure 6.1) suggests there should be seasonal net water shortage. The net monthly water availability as calculated in this study does not consistently reach *below* zero for the dry seasons. The positive water availability in the dry season can be justified when accounting for the constant outflow of wetland into LV. An estimate 30% of the discharge from the Mara River is conveyed through the main channel and reaches the outlet, the remaining 70% spreads over the floodplain (Bregoli et al., 2019). Assuming a constant outflow of 30% of the discharge at Mara Mine, the monthly waterbalance reaches below zero consistently, twice yearly during the dry season. The consistently high water availability year-round could also be the result of errors in the evapotranspiration and precipitation estimates, which is discussed in Section 7.3.

The magnitude of the total inundated area and the magnitude of the preceding precipitation amount do not show a clear relation (for example, the first peak is higher than the second, but does not show significantly higher water availability during the preceding months). There are several possible explanations for this. Firstly, this be the result of uncertainties in the evapotranspiration and precipitation datasets (as discussed below). Moreover, the magnitude in the area is not proportional to volume of water present due to bathymetric effects, and thus a high volumetric precipitation input should not necessarily be followed by a high inundated area. Lastly, the total 'inundated area' includes both open water and flooded vegetation classes, the complexity of the dynamics of the growth and decay of vegetation means the water availability and wetland area will unlikely exhibit a simple, linear relationship.

## 7.3 ADDITIONAL SOURCES OF ERROR

This study builds on the past research on the Mara Wetland, and the results show a broad consistency with previous area estimates. Furthermore, the patterns exhibited by the aerial fluctuations in wetland area are plausible when examined in correspondence with the surrounding water balance components. The prevailing uncertainties in the area estimates are discussed in Subsection 7.1.2. Additional uncertainty is introduced in Phase-II of the research from the evapotranspiration (ET) and precipitation datasets.

The lack of field-measurements has lead to an overall scarcity of ground-truth data, and has limited the possibilities to validate/calibrate the remotely sensed data. The potential uncertainties are thus important to address, especially in the water balance estimates which involves large-scale averaging. A discussion on the uncertainty of the ET and precipitation are provided in Subsection 7.3.1 and Subsection 7.3.2 below.

### 7.3.1 Uncertainty of evapotranspiration data

In light of a lack of multi-temporal field estimates of evaporation/transpiration in the Mara Wetland region, this study uses remotely sensed ET data. The ET dataset has a spatial resolution of 500 m and is an estimate of the sum of the evaporation and transpiration. The wetland is a complex, heterogeneous landscape of different vegetation species with varying phenology and density, in addition, the spatial distribution of water levels adds even more heterogeneity to the region. All of the aforementioned physical characteristics of the land-covers influence the magnitude of the evaporation/transpiration. Despite this, average ET between that of the main swamp and the wetland as a whole is used in this study. The average is taken because the changing spatial distribution of the wetland classes (wet floodplain, open water) made it impossible to obtain a multi-temporal ET dataset in a similar manner as was done for the main swamp (which is partly permanent). The averaging process inherently introduces error to the ET estimates. The ET results can be evaluated based on the magnitude of the flux that is expected for the particular climatic conditions. The Mara River originates in Kenya and the upstream reaches of the river go through an equatorial climate. The wetland itself, located further to the west, has a "Tropical Savannah" climate (based on the Köppen climate classification (Chen & Chen, 2013)). High evaporation rates are expected for this climate, and for some months during the dry season (usually around January - March and July - October) the evaporation could even be greater than the precipitation. In the results the evaporation is only greater than precipitation for a few exceptions, which may suggest the ET is underestimated. The permanent swamp and open water features have a greater evaporation than the remainder of the wetland area, but this does not come through due to the spatial averaging process. This is likely the reason for the low ET rates during the dry season.

### 7.3.2 Uncertainty of precipitation data

As shown in Figure 4.9, there are significant differences between the precipitation datasets. Without reliable field estimates of precipitation, it was not possible to select which of the datasets is most accurate. Therefore, the median of all datasets was used for in this study. This approach inherently contains error and has likely reduced the intensity of the peaks, and may thereby have misrepresented the magnitude of the seasonal rainfall variability. In addition to the uncertainty in the *per pixel* accuracy of the precipitation, the spatial averaging of the precipitation also contributed to the overall uncertainty of the waterbalance estimates. The spatial average of precipitation is used and multiplied by a constant area (bounded by a margin around the maximum wetland extent), regardless of the inundation area at the time. The precipitation has a heterogeneous spatial distribution that is not represented due to averaging process, and this may again have limited the ability of the final precipitation timeseries to reveal peaks. In addition, as mentioned in the previous section (Subsection 7.3.1), the monthly precipitation is nearly always greater than the evapotranspiration, which is likely due to an underestimate of the evapotranspiration but could also be caused by an overestimate of the precipitation. The precipitation is calculated as the monthly accumulated amount over a constant area (margin around the maximum inundated area), which is greater than the area used to calculated the monthly accumulated evapotranspiration (which is the average inundated area for that particular month). The precipitation amount is not calculated using the average inundated area to account for the rainfall that falls just outside of the wetland, but nonetheless contributes to the water available to the wetland via runoff or subsurface flow. The larger area chosen may have lead to an over estimate of the precipitation, at least in relation to the ET.

# 8

# CONCLUSION

## 8.1 GENERAL CONCLUSIONS

In this study the seasonal inundation pattern of the Mara Wetland is reproduced for the years 2017, 2018, 2019 using Sentinel-2 (S2) multispectral satellite imagery. The classification is carried out using S2 bands, derived water/vegetation indices and a DEM as input variables for the Random Forests (RF) classification algorithm. A total of 73 scenes are classified into 7 individual land-cover classes; 3 wetland classes (open water, flooded vegetation, wet floodplain) and 4 dryland classes (dry floodplain, wet agriculture, dry agriculture, bare land). The overall classification accuracy achieved (based on an independent validation set, which is not used to train the classifier itself) is 98.6 %. The classified scenes show the seasonal variations (timing, extent, spatial pattern) of the different land-covers in and around the wetland, and together form a multi-temporal, landscape scale reconstruction of the local flood regime. The aerial inundation extents are used alongside hydrometeorological datasets to investigate the waterbalance and identify which processes govern the seasonal increase in water availability.

The results reveal that the Mara Wetland exhibits a strong bi-modal inundation pattern, analogous to the local precipitation regime. The marked concurrence between peak inundation and peak Lake Victoria (LV) water level indicates the area estimates are coherent and able to correctly capture the timing of seasonal land-cover changes. The inundation area estimates are correlated with the local water balance components (discharge upstream and lake level downstream) in the aim of gaining an understanding of the hydrological drivers behind the seasonal variation in water availability. The wetland flooded area and the lake level show a positive linear relationship, whereas the area and discharge correlate exponentially. The spatial pattern of flooding suggests the correlation between LV and inundation is due to indirect effects from precipitation over the lake, rather than direct backwater effects.

Peak discharge correlates particularly well with the inundated area in the event of a flood extending beyond the main swamp. The discharge timeseries, however, shows no clear link to the seasonal increase in wetland area in the event that water does not extend beyond the main swamp. The results are further exploited to compute the monthly waterbalance of the wetland, by including the accumulated precipitation and evapotranspiration fluxes. The monthly water balance shows a consistent trend of high net water availability for three months preceding the peak wetland area. Considering the absence of a consistent, seasonal variability in discharge and the improbable case the LV backwater extends beyond the main swamp, it is concluded that the local precipitation is the main driver of the bi-modal inundation pattern of the wetland. Nonetheless, during the dry season the constant base-flow that enters the wetland via the Mara River together with the backwater from Lake Victoria are at equilibrium at the extent of the main permanent swamp. This highlights the importance of the river during the low-rainfall months. Additionally, the occasional extreme flood events caused by high discharge of Mara River are also believed to be of ecological necessity, but this is outside of the scope of this study and should be further investigated in future research.

This study has illustrated the effectiveness of using the Random Forests classification algorithm to (semi-automatically) reproduce the variability of the land-cover classes of the Mara Wetland over a period of 3 years. The results reveal the seasonal dynamics of the wetland which has been largely unresolved to date. This knowledge can be used to optimise sustainable water resource management within the Mara River Basin and protect the hydrological and ecological equilibrium of the wetland. Especially considering the planned dam construction and projected climate change, knowledge on the natural variability of the wetland is crucial to informing future dam operations and climate adaptation strategies.

## 8.2 IMPLICATIONS AND RECOMMENDATIONS FOR THE MARA WETLAND

This research suggests the local precipitation plays an important role in the seasonal expansion of the Mara Wetland. The wetland's dependence on precipitation and the projected local climatic changes stresses the importance of quantifying the implications under various scenarios. The base-flow supplied by the Mara River together with the backflow from Lake Victoria appear to be at equilibrium at the extent of the permanent swamp during the dry season. Thus, although the lake and the river are unlikely to *drive* the bi-modal seasonal expansion beyond the main swamp, the flow of the river during low rainfall months is expected to *keep* the main swamp inundated. The relative importance of the dry-season riverflow for maintaining the hydric status of the wetland is of concern, considering the high water demand from river sources at this time for irrigation, dam operations and other local livelihood activities. Further research should be conducted on the exact magnitude of the discharge during base flow conditions. This information can be used to inform the water allocation strategies for the region. Aside from the baseflow, the extreme flooding of the outer reaches of the wetland by the river is expected to play an important role in maintaining overall ecological equilibrium. Further research on the ecological importance of large flood events is necessary to determine whether, and to what degree these events must be remain in spite of altered flow due to the installation of a dam upstream. Lastly, although this research does not quantify the groundwater fluxes, the component of groundwater should not be disregarded in future research.

## 8.3 POTENTIAL APPLICATIONS OF THE METHODOLOGY

Generally, there is a greater amount of data available for streamflow of rivers than there is of wetland hydrology (Halabisky et al., 2016). The vastness, inaccessibility and complexity of these landscapes makes it challenging, if not impossible, to analyse the hydrology using the only field-based measurements. The rapid development in the availability and quality of remote sensing data has contributed considerably to improving the scientific understanding of wetland ecosystems (Guo et al., 2017). This research exemplifies the advantages of utilizing remote sensing data for hydrological analyses. The remote sensing methods employed in this study, and in particular the land-cover classification procedure, is broadly applicable for investigating the multi-temporal behavior of wetland systems. TheRF algorithm is often trained based on the same image to which it is applied. In this study, the RF algorithm is trained bi-seasonally (based on two images) and applied to classify multiple (73) images spanning a time-range of three years. The successful multi-temporal application of the classification tool advocates the use of this method for research that aims to analyse the dynamics of a wetland region where there is a scarcity of multi-temporal ground-truth data.

Furthermore, this study tackles the challenges introduced by the presence of aquatic vegetation. Optical remote sensing does not have the same penetrative capacity through vegetation as most microwave satellites do. This shortcoming has limited the efficacy of using optical remote sensing to analyse the hydrology of densely vegetated wetland regions. This research overcomes the challenge of vegetation by employing methods that *draws* on the characteristics of aquatic vegetation rather than attempting to *evade* its presence. This opens up opportunities for the use of optical remote sensing in other wetland areas similarly characterised by a dense coverage of aquatic species.

This study reproduces the wetland in space and time, which, aside from hydrological analysis, can be useful for a number of other applications. For example, the spatial distribution of the wetland classes, and the dynamics thereof, could support ecological investigations on the presence, mixing and response of vegetation communities. Furthermore, the spatial pattern of wetlands could also be useful input for the development of climate adaptation strategies to protect species or local communities that rely on its ecosystem services.

# REFERENCES


Adam, E., Mutanga, O., Odindi, J., & Abdel-Rahman, E. M. (2014). Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers. *International Journal of Remote Sensing*, *35*(10), 3440–3458. https://doi.org/10.1080/01431161.2014.903435

Alonso, A., Muñoz-Carpena, R., & Kaplan, D. (2020). Coupling high-resolution field monitoring and MODIS for reconstructing wetland historical hydroperiod at a high temporal frequency. *Remote Sensing of Environment*, *247*. https://doi.org//doi.org/10.1016/j.rse.2020.111807

Black, E. (2005). The Relationship between Indian Ocean Sea-Surface Temperature and East African rainfall. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, *363*(1826), 43–47. https://tudelft.on.worldcat.org/oclc/5551925375

Bogers, E. (2007). Mara Dryland Becomes Wetland. http://www.ethesis.net/dryland/MARAWETLAND.pdf

Bregoli, F., Crosato, A., Paron, P., & McClain, M. (2019). Humans reshape wetlands: Unveiling the last 100years of morphological changes of the Mara Wetland, Tanzania. *Science of the Total Environment*, *691*, 896–907. https://doi.org/10.1016/j.scitotenv.2019.07.189

Bregoli, F., Crosato, A., Paron, P., & McClain, M. (2018). Investigating morphological responses to sediment flux alterations and land use changes in the mara wetland, tanzania. AGU Fall Meeting 2018. Washington DC, US.

Campbell, D., Keddy, P. A., Broussard, M., & McFalls-Smith, T. B. (2016). Small Changes in Flooding Have Large Consequences: Experimental Data from Ten Wetland Plants. *Wetlands*, *36*(3), 457–466. https://doi.org/10.1007/s13157-016-0754-7

Chen, D., & Chen, H. W. (2013). Using the Köppen classification to quantify climate variation and change: An example for 1901–2010. *Environmental Development*, *6*, 69–79. https://doi.org/https://doi.org/10.1016/j.envdev.2013.03.007

Collins, S. D., Heintzman, L. J., Starr, S. M., Wright, C. K., Henebry, G. M., & McIntyre, N. E. (2014). Hydrological dynamics of temporary wetlands in the southern Great Plains as a function of surrounding land use. *Journal of Arid Environments*, *109*, 6–14. https://doi.org/10.1016/j.jaridenv.2014.05.006

Coluzzi, R., Imbrenda, V., Lanfredi, M., & Simoniello, T. (2018). A first assessment of the Sentinel-2 Level 1-C cloud mask product to support informed surface analyses. *Remote Sensing of Environment*, *217*, 426–443. https://doi.org/https://doi.org/10.1016/j.rse.2018.08.009

Costanza, R., D'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R. V., Paruelo, J., Raskin, R. G., Sutton, P., & van den Belt, M. (1997). The value of the world's ecosystem services and natural capital. *Nature*, *387*(6630), 253–260. https://doi.org/10.1038/387253a0

Council, N. R. (1995). Wetland Characterization: Water, Substrate, and Biota, In *Wetlands: Characteristics and boundaries*. Washington, DC, The National Academies Press. https://doi.org/https://doi.org/10.17226/4766

Dabboor, M., & Brisco, B. (2019). Wetland monitoring and mapping using synthetic aperture radar (D. Gökçe, Ed.). In D. Gökçe (Ed.), *Wetlands management*. Rijeka, IntechOpen. https://doi.org/10.5772/intechopen.80224

Defersha, M. B., Melesse, A. M., & McClain, M. (2012). Watershed scale application of WEPP and EROSION 3D models for assessment of potential sediment source areas and runoff flux in the Mara River Basin, Kenya. *Catena*, *95*, 63–72. https://doi.org/10.1016/j.catena.2012.03.004

Dessu, S. B., & Melesse, A. M. (2012). Modelling the rainfall-runoff process of the Mara River basin using the Soil and Water Assessment Tool. *Hydrological Processes*, *26*(26), 4038–4049. https://doi.org/10.1002/hyp.9205

Dronova, I., Gong, P., & Wang, L. (2011). Object-based analysis and change detection of major wetland cover types and their classification uncertainty during the low water period at Poyang Lake, China. *Remote Sensing of Environment*, *115*(12), 3220–3236. https://doi.org/10.1016/j.rse.2011.07.006

Dudgeon, D., Arthington, A. H., Gessner, M. O., Kawabata, Z. I., Knowler, D. J., Lévêque, C., Naiman, R. J., Prieur-Richard, A. H., Soto, D., Stiassny, M. L., & Sullivan, C. A. (2006). Freshwater biodiversity: Importance, threats, status and conservation challenges. *Biological Reviews of the Cambridge Philosophical Society*, *81*(2), 163–182. https://doi.org/10.1017/S1464793105006950

Dutton, C. L., Subalusky, A. L., Anisfeld, S. C., Njoroge, L., Rosi, E. J., & Post, D. M. (2018). The influence of a semi-Arid sub-catchment on suspended sediments in the Mara River, Kenya. *PLoS ONE*, *13*(2), 1–19. https://doi.org/10.1371/journal.pone.0192828

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., Husak, G., Rowland, J., Harrison, L., Hoell, A., & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations - A new environmental record for monitoring extremes. *Scientific Data*, *2*, 1–21. https://doi.org/10.1038/sdata.2015.66

Gallant, A. L. (2015). The challenges of remote monitoring of wetlands. *Remote Sensing*, *7*(8), 10938–10950. https://doi.org/10.3390/rs70810938

Gao, B. (1996). NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, *58*(3), 257–266. https://doi.org/https://doi.org/10.1016/S0034-4257(96)00067-3

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

Grabs, T., Seibert, J., Bishop, K., & Laudon, H. (2009). Modeling spatial patterns of saturated areas: A comparison of the topographic wetness index and a dynamic distributed model. *Journal of Hydrology*, *373*(1-2), 15–23. https://doi.org/10.1016/j.jhydrol.2009.03.031

Gulácsi, A., & Kovács, F. (2020). Sentinel-1-imagery-based high-resolutionwater cover detection on wetlands, aided by google earth engine. *Remote Sensing*, *12*(10), 1–20. https://doi.org/10.3390/rs12101614

Guo, M., Li, J., Sheng, C., Xu, J., & Wu, L. (2017). A review of wetland remote sensing. *Sensors (Switzerland)*, *17*(4), 1–36. https://doi.org/10.3390/s17040777

Halabisky, M., Moskal, L. M., Gillespie, A., & Hannam, M. (2016). Reconstructing semi-arid wetland surface water dynamics through spectral mixture analysis of a time series of Landsat satellite images (1984-2011). *Remote Sensing of Environment*, *177*, 171–183. https://doi.org/10.1016/j.rse.2016.02.040

Hardisky, M., Klemas, V., & Smart. (1983). The influence of soil salinity, growth form, and leaf moisture on the spectral radiance of Spartina Alterniflora canopies. *Photogrammetric Engineering and Remote Sensing*, *48*, 77–84.

Hayashi, M., van der Kamp, G., & Rosenberry, D. O. (2016). Hydrology of Prairie Wetlands: Understanding the Integrated Surface-Water and Groundwater Processes. *Wetlands*, *36*, 237–254. https://doi.org/10.1007/s13157-016-0797-9

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., … Thépaut, J.-N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/https://doi.org/10.1002/qj.3803

Hird, J. N., DeLancey, E. R., McDermid, G. J., & Kariyeva, J. (2017). Google earth engine, open-access satellite data, and machine learning in support of large-area probabilistic wetland mapping. *Remote Sensing*, *9*(12). https://doi.org/10.3390/rs9121315

Houlahan, J. E., & Findlay, C. S. (2004). Estimating the 'critical' distance at which adjacent land-use degrades wetland water and sediment quality. *Landscape Ecology*, *19*(6), 677–690. https://doi.org/10.1023/B:LAND.0000042912.87067.35

HSU. (2019). Spectral reflectance. http://gsp.humboldt.edu/OLM/Courses/GSP_216_Online/lesson2-1/reflectance.html

Hu, S., Niu, Z., Chen, Y., Li, L., & Zhang, H. (2017). Global wetlands: Potential distribution, wetland loss, and status. *Science of The Total Environment*, *586*, 319–327. https://doi.org/https://doi.org/10.1016/j.scitotenv.2017.02.001

Huffman, G. J., Bolvin, D. T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E. J., Sorooshian, S., Tan, J., & Xie, P. (2018). NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG), 35. https://pmm.nasa.gov/sites/default/files/document%7B%5C_%7Dfiles/IMERG%7B%5C_%7DATBD%7B%5C_%7DV5.2%7B%5C_%7D0.pdf

Huffman, G., Adler, R., Bolvin, D., Gu, G., Nelkin, E., Bowman, K., Hong, Y., Stocker, E., & Wolff, D. (2007). The trmm multisatellite precipitation analysis (tmpa): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, *8*. https://doi.org/10.1175/JHM560.1

Hulsman, P., Bogaard, T. A., & Savenije, H. H. (2018). Rainfall-runoff modelling using river-stage time series in the absence of reliable discharge information: A case study in the semi-arid Mara River basin. *Hydrology and Earth System Sciences*, *22*(10), 5081–5095. https://doi.org/10.5194/hess-22-5081-2018

Ilyas, S., Xu, X., Jia, G., & Zhang, A. (2019). Interannual Variability of Global Wetlands in Response to El Niño Southern Oscillations (ENSO) and Land-Use. *Frontiers in Earth Science*, *7*, 1–14. https://doi.org/10.3389/feart.2019.00289

Kabete, J. M., Groves, D. I., McNaughton, N. J., & Mruma, A. H. (2012). A new tectonic and temporal framework for the Tanzanian Shield: Implications for gold metallogeny and undiscovered endowment. *Ore Geology Reviews*, *48*(9), 88–124. https://doi.org/10.1016/j.oregeorev.2012.02.009

Kema, W. M. (2010). The significance of Ecosystem Services in sustaining people's livelihoods; a case study in Mara wetland, Musoma and Tarime District, Tanzania.

Kiragu, G., Home, P., MATI, B., & GATHENYA, J. (2011). Assessment of Suspended Sediment Loadings and their impact on the Environmental Flows of Upper Transboundary Mara River, Kenya. *Nile Basin Water Science & Engineering Journal*, *4*(2), 59–70.

Koning, C. O. (2005). Vegetation patterns resulting from spatial and temporal variability in hydrology, soils, and trampling in an isolated basin marsh, New Hampshire, USA. *Wetlands*, *25*(2), 239–251. https://doi.org/10.1672/1

Lefebvre, G., Davranche, A., Willm, L., Campagna, J., Redmond, L., Merle, C., Guelmami, A., & Poulin, B. (2019). Introducing WIW for detecting the presence of water in wetlands with landsat and sentinel satellites. *Remote Sensing*, *11*(19), 10–14. https://doi.org/10.3390/rs11192210

Li, Z., Xu, J., Shilpakar, R. L., & Ma, X. (2014). Mapping wetland cover in the greater Himalayan region: A hybrid method combining multispectral and ecological characteristics. *Environmental Earth Sciences*, *71*(3), 1083–1094. https://doi.org/10.1007/s12665-013-2512-y

Liew, S. C. (2001). Principles of remote sensing. http://www.crisp.nus.edu.sg/~research/tutorial/rsmain.htm

Ludwig, C., Walli, A., Schleicher, C., Weichselbaum, J., & Riffler, M. (2019). A highly automated algorithm for wetland detection using multi-temporal optical satellite data. *Remote Sensing of Environment*, *224*, 333–351. https://doi.org/10.1016/j.rse.2019.01.017

Ma, S., Zhou, Y., Gowda, P. H., Dong, J., Zhang, G., Kakani, V. G., Wagle, P., Chen, L., Flynn, K. C., & Jiang, W. (2019). Application of the water-related spectral reflectance indices: A review. *Ecological Indicators*, *98*, 68–79. https://doi.org/10.1016/j.ecolind.2018.10.049

Mahdianpari, M., Salehi, B., Mohammadimanesh, F., Homayouni, S., & Gill, E. (2019). The first wetland inventory map of newfoundland at a spatial resolution of 10 m using sentinel-1

and sentinel-2 data on the Google Earth Engine cloud computing platform. *Remote Sensing*, *11*(1). https://doi.org/10.3390/rs11010043

Mati, B. M., Mutie, S., Gadain, H., Home, P., & Mtalo, F. (2008). Impacts of land-use/cover changes on the hydrology of the transboundary Mara River, Kenya/Tanzania. *Lakes and Reservoirs: Research and Management*, *13*(2), 169–177. https://doi.org/10.1111/j.1440-1770.2008.00367.x

Mayo, A. W., Muraza, M., & Norbert, J. (2013). International Journal of Water Resources and Environmental Engineering The Role of Mara River Basin Wetland in Reduction of Nitrogen Load to Lake Victoria, *5*(12), 659–669. https://doi.org/10.5897/IJWREE2013.0444

McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, *17*(7), 1425–1432. https://doi.org/10.1080/01431169608948714

Merriam-Webster. (n.d.). Hydroperiod [https://www.merriam-webster.com/dictionary/hydroperiod, Last accessed on 2017-09-03].

Millard, K., & Richardson, M. (2015). On the importance of training data sample selection in Random Forest image classification: A case study in peatland ecosystem mapping. *Remote Sensing*, *7*(7), 8489–8515. https://doi.org/10.3390/rs70708489

Millennium Ecosystem Assessment. (2005). *Ecosystems and human well-being: Synthesis*. Island Press, Washington, DC.

Mohamed, Y. A., Bastiaanssen, W. G., & Savenije, H. H. (2004). Spatial variability of evaporation and moisture storage in the swamps of the upper Nile studied by remote sensing techniques. *Journal of Hydrology*, *289*(1-4), 145–164. https://doi.org/10.1016/j.jhydrol.2003.11.038

Mohammadimanesh, F., Salehi, B., Mahdianpari, M., Brisco, B., & Motagh, M. (2018). Wetland Water Level Monitoring Using Interferometric Synthetic Aperture Radar (InSAR): A Review. *Canadian Journal of Remote Sensing*, *44*(4), 247–262. https://doi.org/10.1080/07038992.2018.1477680

Muraza, M., Mayo, A. W., & Norbert, J. (2013). Wetland Plant Dominance, Density And Biomass In Mara River Basin Wetland Upstream Of Lake Victoria In Tanzania. *International Journal of Scientific & Technology Research*, *2*(12), 348–359.

Muro, J. (2019). Mapping intra- and inter-annual dynamics in wetlands with multispectral, thermal and SAR time series, 116.

Mutie, S. M., Mati, B., Home, P., Gadain, H., & Gathenya, J. (2006). Evaluating land use change effects on river flow using USGS geospatial stream flow model in Mara River Basin, Kenya. *Proceedings of the 2nd Workshop of the EARSeL SIG on Land Use and Land Cover*, 28–30.

Mwangi, H. M., Lariu, P., Julich, S., Patil, S. D., McDonald, M. A., & Feger, K. H. (2017). Characterizing the intensity and dynamics of land-use change in the Mara River Basin, East Africa. *Forests*, *9*(1), 1–17. https://doi.org/10.3390/f9010008

Obiero, K. O., Wa'Munga, P. O., Raburu, P. O., & Okeyo-Owuor, J. B. (2012). Community Based Approach to the Management of Nyando Wetland, Lake Victoria Basin, Kenya., In *The people of nyando wetland: Socioeconomics, gender and cultural issues*. Nairobi, 2012 KDC - VIRED - UNDP. http://www.undp.org/content/dam/kenya/docs/energy%7B%5C_%7Dand%7B%5C_%7Denvironment/Nyando%20Book%20-%20FINAL%E2%80%8BMOST-internet.pdf%7B%5C#%7Dpage=44

of Tanzania, U. R. (2017). Conservation Investment Plan for Mara Wetlands, United Republic Of Tanzania. https://www.climatelinks.org/resources/conservation-investment-plan-mara-wetlands

Ogilvie, A., Belaud, G., Delenne, C., Bailly, J. S., Bader, J. C., Oleksiak, A., Ferry, L., & Martin, D. (2015). Decadal monitoring of the Niger Inner Delta flood dynamics using MODIS optical data. *Journal of Hydrology*, *523*, 368–383. https://doi.org/10.1016/j.jhydrol.2015.01.036

Okamoto, K., Ushio, T., Iguchi, T., Takahashi, N., & Iwanami, K. (2005). The Global Satellite Mapping of Precipitation (GSMaP) project, In *International geoscience and remote sensing symposium (igarss)*. https://doi.org/10.1109/IGARSS.2005.1526575

Physics, M. (2016). The electromagnetic spectrum. https://www.miniphysics.com/electromagnetic-spectrum_25.html

Poff, N. L. R., Allan, J. D., Bain, M. B., Karr, J. R., Prestegaard, K. L., Richter, B. D., Sparks, R. E., & Stromberg, J. C. (1997). The natural flow regime: A paradigm for river conservation and restoration. *BioScience*, *47*(11), 769–784. https://doi.org/10.2307/1313099

Rains, M. C., Leibowitz, S. G., Cohen, M. J., Creed, I. F., Golden, H. E., Jawitz, J. W., Kalla, P., Lane, C. R., Lang, M. W., & Mclaughlin, D. L. (2016). Geographically isolated wetlands are part of the hydrological landscape. *Hydrological Processes*, *30*(1), 153–160. https://doi.org/10.1002/hyp.10610

*Ramsar convention* [Intergovernmental Treaty]. (1971). https://www.ramsar.org/sites/default/files/documents/library/scan_certified_e.pdf

Rouse, J., J. W., Haas, R. H., Schell, J. A., & Deering, D. W. (1974). Monitoring Vegetation Systems in the Great Plains with Erts. In *Nasa special publication* (p. 309).

Semeniuk, C. A., & Semeniuk, V. (1995). A geomorphic approach to global classification for inland wetlands. *Classification and Inventory of the World's Wetlands*, 103–124. https://doi.org/10.1007/978-94-011-0427-2_9

Semeniuk, C. A., & Semeniuk, V. (2004). Wetlands Research Association (Inc). https://www.ramsar.org/sites/default/files/forum/wi/ACF357.pdf

Singh, K. V., Setia, R., Sahoo, S., Prasad, A., & Pateriya, B. (2015). Evaluation of NDWI and MNDWI for assessment of waterlogging by integrating digital elevation model and groundwater level. *Geocarto International*, *30*(6), 650–661. https://doi.org/10.1080/10106049.2014.965757

Slagter, B., Tsendbazar, N.-E., Vollrath, A., & Reiche, J. (2020). Mapping wetland characteristics using temporally dense Sentinel-1 and Sentinel-2 data: A case study in the St. Lucia wetlands, South Africa. *International Journal of Applied Earth Observation and Geoinformation*, *86*, 102009. https://doi.org/https://doi.org/10.1016/j.jag.2019.102009

Sorooshian, S., Hsu, K. L., Gao, X., Gupta, H. V., Imam, B., & Braithwaite, D. (2000). Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bulletin of the American Meteorological Society*. https://doi.org/10.1175/1520-0477(2000)081<2035:EOPSSE>2.3.CO;2

Stoop, B. (2017). *Morphology of the Mara River* (Master). Delft University of Technology.

Subalusky, A. A., Dutton, C., Hill, T., Aleman, J., Kinyanjui, R., Staver, C., & Post, D. (2019). Reading the Historical Record of the Mara Using Sediment Cores from the Mara Wetland. Final Report.

Thomas, R. F., Kingsford, R. T., Lu, Y., Cox, S. J., Sims, N. C., & Hunter, S. J. (2015). Mapping inundation in the heterogeneous floodplain wetlands of the Macquarie Marshes, using Landsat Thematic Mapper. *Journal of Hydrology*, *524*, 194–213. https://doi.org/10.1016/j.jhydrol.2015.02.029

Tockner, K., & Stanford, J. A. (2002). Riverine flood plains: Present state and future trends. *Environmental Conservation*, *29*(3), 308–330. https://doi.org/10.1017/S037689290200022X

Tsyganskaya, V., Martinis, S., Marzahn, P., & Ludwig, R. (2018). SAR-based detection of flooded vegetation–a review of characteristics and approaches. *International Journal of Remote Sensing*, *39*(8), 2255–2293. https://doi.org/10.1080/01431161.2017.1420938

USAID. (2018). Fact Sheet: Valuing Ecosystem Services of Mara Wetlands, Tanzania.

USAID. (2019). Vulnerability and Adaptation in the Mara River Basin. https://www.climatelinks.org/projects/atlas

U.S.EPA. (2008). *Methods for evaluating wetland condition: Wetland hydrology* (tech. rep.) [EPA-822-R-08-024]. Office of Water, U.S. Environmental Protection Agency, Washington, DC. EPA-822-R-08-024.

Vanderkelen, I., Van Lipzig, N. P., & Thiery, W. (2018). Modelling the water balance of Lake Victoria (East Africa) - Part 1: Observational analysis. *Hydrology and Earth System Sciences*, *22*(10), 5509–5525. https://doi.org/10.5194/hess-22-5509-2018

Ward, D. P., Petty, A., Setterfield, S. A., Douglas, M. M., Ferdinands, K., Hamilton, S. K., & Phinn, S. (2014). Floodplain inundation and vegetation dynamics in the Alligator Rivers region (Kakadu) of northern Australia assessed using optical and radar remote sensing. *Remote Sensing of Environment*, *147*, 43–55. https://doi.org/10.1016/j.rse.2014.02.009

White, L., Brisco, B., Dabboor, M., Schmitt, A., & Pratt, A. (2015). *A collection of SAR methodologies for monitoring wetlands* (Vol. 7). https://doi.org/10.3390/rs70607615

Wickel, B. A., Colditz, R., Ressl, R., Kucharski, J., & Salinas-Rodríguez, S. (2020). *EGU2020-12991 presentation*. https://doi.org/10.5194/egusphere-egu2020-12991

Wilson, E., & Sader, S. (2002). Detection of forest harvest type using multiple dates of landsat tm imagery. *Remote Sensing of Environment*, *80*, 385–396. https://doi.org/10.1016/S0034-4257(01)00318-2

Wilusz, D. C., Zaitchik, B. F., Anderson, M. C., Hain, C. R., Yilmaz, M. T., & Mladenova, I. E. (2017). Monthly flooded area classification using low resolution SAR imagery in the Sudd wetland from 2007 to 2011. *Remote Sensing of Environment*, *194*, 205–218. https://doi.org/10.1016/j.rse.2017.03.005

Xu, H. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, *27*(14), 3025–3033. https://doi.org/10.1080/01431160600589179

Zedler, J. B., & Kercher, S. (2005). Wetland resources: Status, trends, ecosystem services, and restorability. *Annual Review of Environment and Resources*, *30*, 39–74. https://doi.org/10.1146/annurev.energy.30.050504.144248

Zimba, H., Kawawa, B., Chabala, A., Phiri, W., Selsam, P., Meinhardt, M., & Nyambe, I. (2018). Assessment of trends in inundation extent in the Barotse Floodplain, upper Zambezi River Basin: A remote sensing-based approach. *Journal of Hydrology: Regional Studies*, *15*, 149–170. https://doi.org/10.1016/j.ejrh.2018.01.002

# A

# APPENDIX A

## A.1 PREVIOUS WETLAND AREA ESTIMATES

**Table A.1:** Reported wetland area estimates from various studies

| Data used | Wetland area (km²) | Source |
|---|---|---|
| Landsat MSS | 1973: 286 | (Mati et al., 2008)<br>(Mutie et al., 2006) |
| Landsat TM/ETM | 1984: 604<br>2000: 1394 | (Mati et al., 2008)<br>(Mutie et al., 2006) |
| GLOWS | 2007: 204.46 | (Mayo et al., 2013) |
| Ministry of Natural Resources and Tourism | 2018: 390<br>2018: 517 (flooded) | (USAID, 2018) |
| German map (Reimer, 1911) | 1911: 113<br>1963: 270<br>1973: 288 | (Bregoli et al., 2019) |
| Landsat 4 | 1984: 213 | (Bregoli et al., 2019) |
| Landsat 7 | 2002: 344 | (Bregoli et al., 2019) |
| Google Earth | 2017: 388 | (Bregoli et al., 2019) |

## A.2 SPECIFICATIONS OF PRECIPITATION DATASETS

**Table A.2:** Precipitation datasets specifications

| Dataset | Resolution (°) | Precipitation units | Availability in GEE | Method | Source |
|---|---|---|---|---|---|
| TMPA: Tropical Rainfall Measuring Mission (TRMM) 3B42 - Multi-satellite Precipitation Analysis | 0.25 | mm/hr | 1998-2019 | Satellite data (IR/MW), gauge. Combines instrument precipitation estimate and the sampling of microwave data with fill-in using infrared estimates. Merges data from multiple satellites, including SSMI, SSMIS, MHS, AMSU-B and AMSR-E | (G. Huffman et al., 2007) |
| GPM-IMERG: Global Precipitation Measurement - Integrated Multi-satellitE Retrievals for GPM | 0.1 | mm/hr | 2000-present | Satellite (IR/MW), gauge. Quasi-global best estimate precipitation rate, successor of TRMM. IMERG is the algorithm that provides rainfall estimates combining data from all passive-microwave instruments in the GPM Constellation. Merges satellite microwave precipitation estimates and microwave-calibrated infrared (IR) satellite estimates and precipitation gauge analyses. | (G. J. Huffman et al., 2018) |
| CHIRPS: Climate Hazards Group InfraRed Precipitation with Station Data v2 | 0.05 | mm/day | 1981 - present | Satellite data (IR/MW), gauge. Uses satellite imagery with in-situ station data, combines data from CFS v2, TRMM 3B42, CHPClim. | (Funk et al., 2015) |
| PERSIANN-CDR: Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks-Climate Data Record | 0.25 | mm/day | 1983 - present | Satellite (IR), gauge. Uses Gridded Satellite (GridSat-B1) IR data that are derived from merging ISCCP B1 IR data, along with GPCP version 2.2. | (Sorooshian et al., 2000) |
| ERA5: ECMWF Reanalysis | 0.1 | mm/day | 1979 - 2020 | Reanalysis. Global weather model. Hourly estimates of atmospheric variables (50+). ERA5 replaces its predecessor, the ERA-Interim reanalysis. Uses 4DVar data assimilation in CY41R2, RTTOV11 radiative transfer. | (Hersbach et al., 2020) |
| GSMaP (Global Satellite Mapping of Precipitation) Operational and Reanalysis | 0.1 | mm/hr | 2014 - present | Satellite, reanalysis, gauge and non-gauge corrected. Values are estimated using multi-band passive microwave and infrared radiometers from the GPM Core Observatory satellite and with the assistance of a constellation of other satellites + NOAA/CPC gauge measurements | (Okamoto et al., 2005) |

# B

# APPENDIX B

## B.1 INDEX COMPOSITE MAPS

Figure B.1 and Figure B.2 show the calculated NDMI response for the wetland in June and November 2017. The NDMI shows sensitivity to moisture, but does not show clear distinction between open water and vegetated wetland.

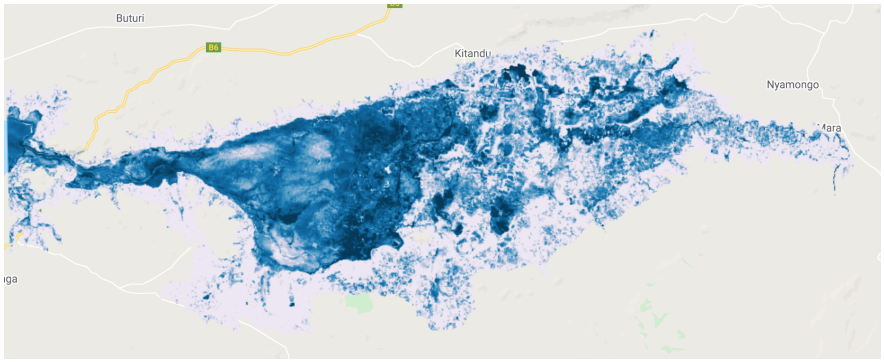



**Figure B.1:** NDMI visualisation for 11-06-2017, adjusted with a new minimum and maximum

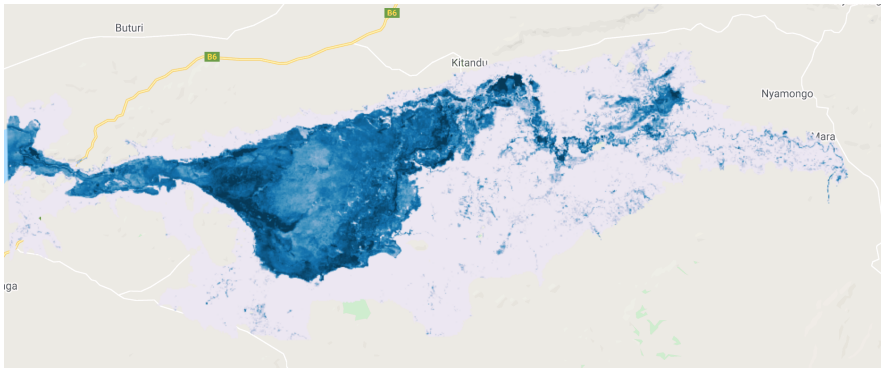



**Figure B.2:** NDMI visualisation for 03-11-2017, adjusted with a new minimum and maximum

Figure B.3 and Figure B.4 show the NDVI response for the wetland in June and November 2017. The NDVI map clearly shows that the wetland is more easily distinguishable in the dry season (November). Unlike the NDMI, the NDVI is sensitive to open water (which can be see at the outlet to Lake Victoria on the west). In June, the upland vegetation, around the wetland, does not show much difference from the wetland itself.

Figure B.5 and Figure B.6 show the calculated MNDWI response for the wetland in June and November 2017. The MNDWI shows sensitivity to moisture, and clearly shows contrast between open water and vegetated wetland. In both June and November the upland areas and areas around the wetland show low values of MNDWI.

Figure B.7 and Figure B.8 show the calculated NDII response for the wetland in June and November 2017. The NDII shows sensitivity to vegetation, the dry upland areas can clearly been seen in red, especially in November. The NDII shows no contrast between open water and wetland area.

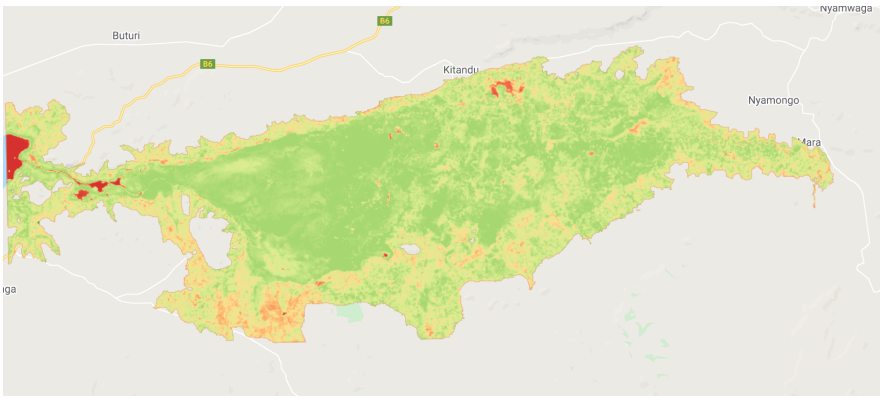



**Figure B.3:** NDVI composite for 11-06-2017 from S2 image collection

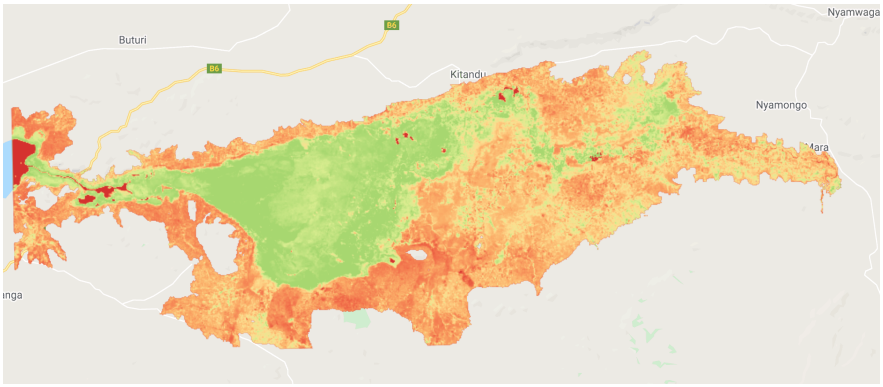



**Figure B.4:** NDVI composite for 03-11-2017 from S2 image collection

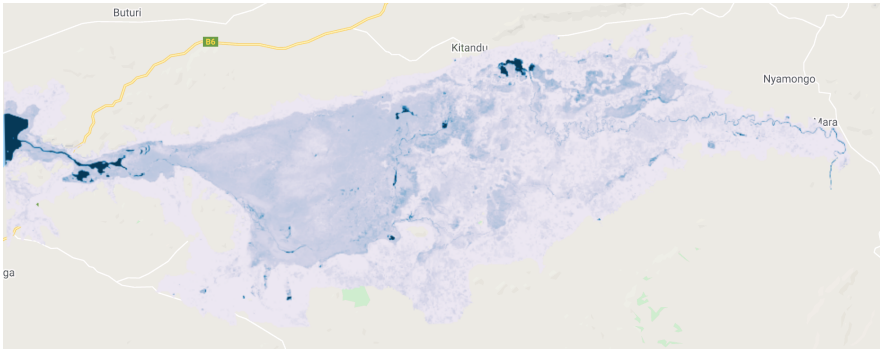



**Figure B.5:** MNDWI composite for 11-06-2017 from S2 image collection

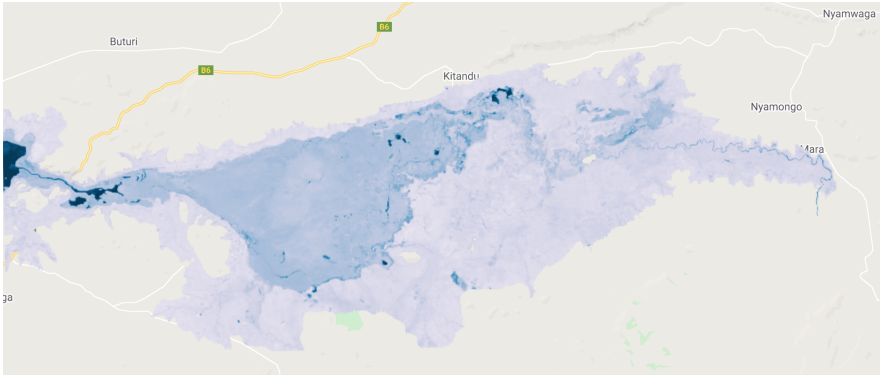



**Figure B.6:** MNDWI composite for 03-11-2017 from S2 image collection

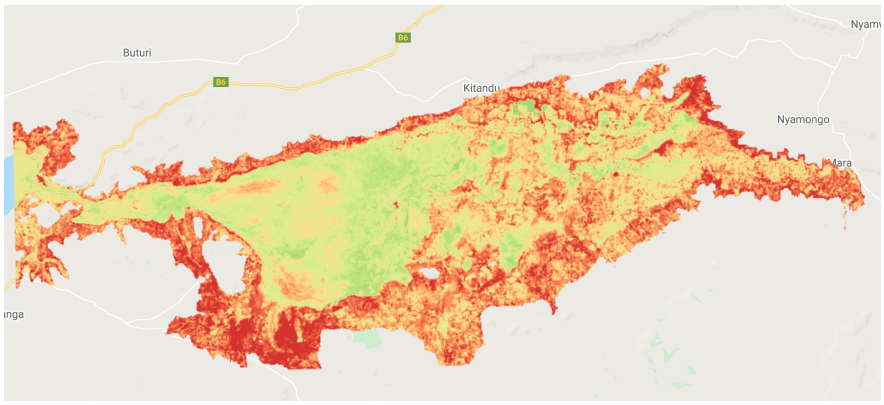



**Figure B.7:** NDII composite for 11-06-2017 from S2 image collection

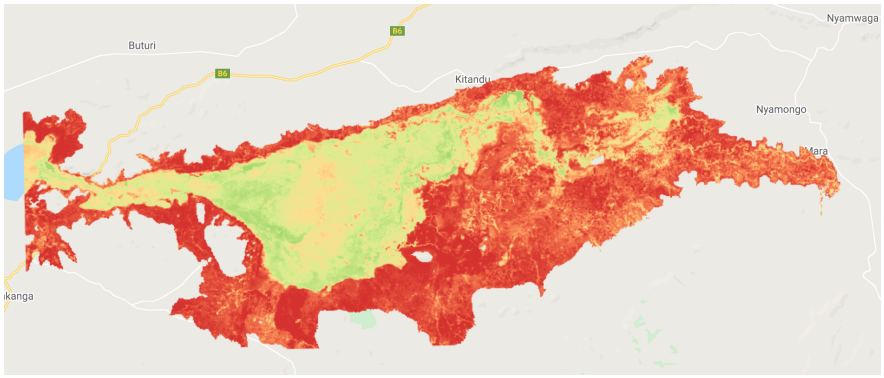



**Figure B.8:** NDII composite for 03-11-2017 from S2 image collection

## B.2 UNSUPERVISED CLASSIFICATION

An unsupervised K-means clustering classification is performed with 6 classes for both the wet and dry season images. The images are sampled from all bands and calculated indices except B1, B2, B9 an B10, because these bands are sensitive to atmospheric properties and might thus add noise to the data. The results for the classification for the June and November images are shown in Figure B.9 and Figure B.10, respectively. The same colors of the classes do not necessarily represent the same classes for both images. The results from the unsupervised classification clearly shows the wetland area is distinguishable from the upland areas, and that the wetland area decreases from June to November. The open water at the western edge of the wetland is a separate class in both images. The dense Papyrus swamp is classified as two separate classes in June (dark and light green) and as only one in November (dark green). The upland areas contains various classes which are not directly recognizable without further analysis.

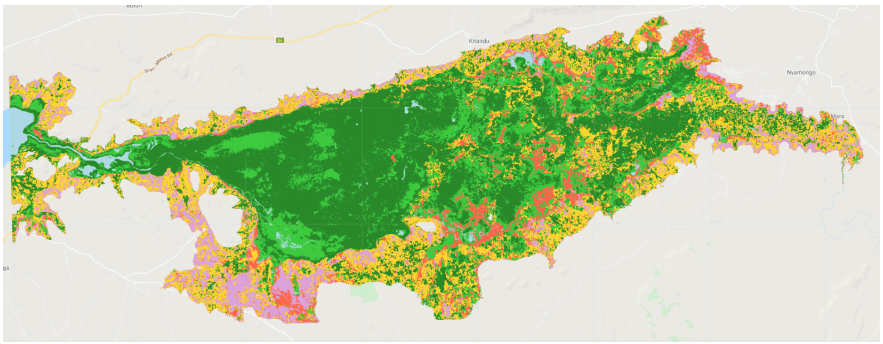


**Figure B.9:** Unsupervised K-means clustering classification for 06/2017

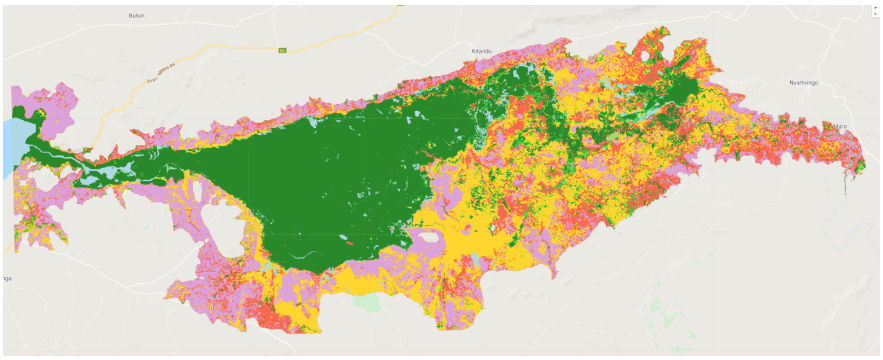


**Figure B.10:** Unsupervised K-means clustering classification for 11/2017

## B.3 MULTI-TEMPORAL SPECTRAL RESPONSE

Since the wetland must be delineated over the entire image collection, a time-series analysis of the spectral response of the land-classes is carried out to visualise the spectral changes in the sampled classes undergo. This is a useful step in determining what bands/indices the training data should be sampled, to be relevant for the entire series, rather than only the image from which the samples are taken.

The NDWI, MNDWI and NDVI indices exhibited the greatest difference in response between classes. An example of a time-series spectral response for the sampled classes is provided in Figure B.11 and Figure B.12.

The spectral response of the open water class shows the least overlap with other classes for the duration of the time-series. That implies the open water class can likely be distinguished from other classes with the most certainty.

The peaks for the 'Floodplain' class in April 2018 and December 2019 are likely due to the flooding of some pixels within the sample (samples are take in close to the fluctuating edge of the wetland), and thus shows reflectance values similar to open water. The 'Bare' class does not show as much variation and is thus likely not wetting or greening so much much from one image to the next. The 'Agriculture' class shows clear seasonal patterns due to farming practices, and show a very similar reflectance as 'Bare' in the dry season (August, September, October, November, depending on the year) due to the lack of vegetation and moisture at this time of the year. After the open water class, the next most separable classes are those belonging to either 'vegetated' or 'bare'. The bands that are particularly useful in separating bare from vegetated are the VNIR, NIR, SWIR 1 and SWIR 2 bands (B8A, B8, B11 and B12). The individual bands B11 and B12 show contrast between vegetation and bare ground, and the normalized difference between the B11 and B8/B8A (NDMI and NDII) show similar sensitivity to vegetation.

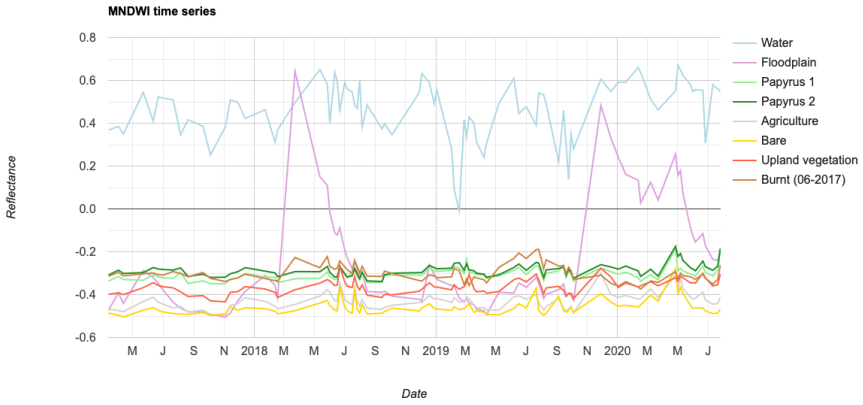



**Figure B.11:** MNDWI training data timeseries

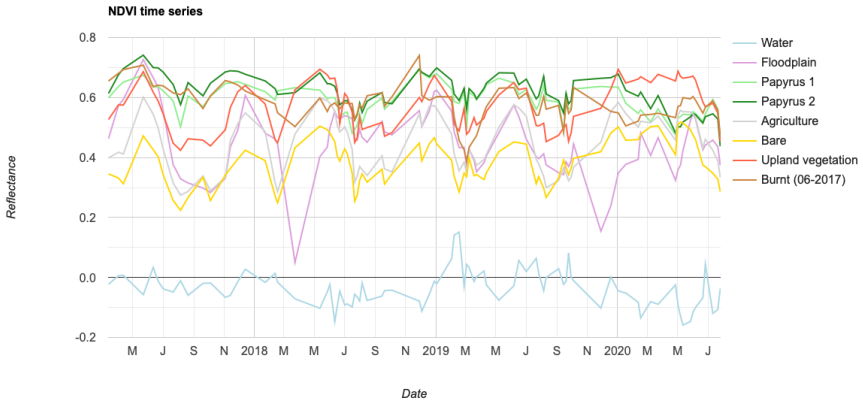



**Figure B.12:** NDVI training data timeseries

# C

# APPENDIX C

## C.1 SPECTRAL RESPONSE OF TRAINING DATASET

### C.1.1 Wet season training data

**Table C.1:** Median band and index response for training data samples obtained in the wet season

| Band/Index | OW | FV1 | FV2 | AW | BA | WF |
|---|---|---|---|---|---|---|
| AVE_DSM: | 1134.7 | 1145.0 | 1146.5 | 1193.0 | 1240.5 | 1148.0 |
| B1: | 0.107 | 0.110 | 0.109 | 0.118 | 0.128 | 0.111 |
| B10: | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 | 0.002 |
| B11: | 0.017 | 0.162 | 0.151 | 0.241 | 0.308 | 0.178 |
| B12: | 0.012 | 0.075 | 0.064 | 0.138 | 0.200 | 0.081 |
| B2: | 0.078 | 0.086 | 0.085 | 0.097 | 0.109 | 0.090 |
| B3: | 0.058 | 0.086 | 0.086 | 0.099 | 0.112 | 0.092 |
| B4: | 0.042 | 0.059 | 0.053 | 0.082 | 0.109 | 0.061 |
| B5: | 0.040 | 0.094 | 0.090 | 0.122 | 0.144 | 0.108 |
| B6: | 0.039 | 0.224 | 0.257 | 0.247 | 0.237 | 0.257 |
| B7: | 0.040 | 0.278 | 0.329 | 0.294 | 0.276 | 0.312 |
| B8: | 0.034 | 0.270 | 0.317 | 0.282 | 0.265 | 0.302 |
| B8A: | 0.036 | 0.319 | 0.368 | 0.327 | 0.308 | 0.345 |
| B9: | 0.011 | 0.070 | 0.078 | 0.072 | 0.072 | 0.073 |
| MNDWI: | 0.529 | −0.308 | −0.271 | −0.415 | −0.470 | −0.317 |
| NDII: | 0.346 | 0.315 | 0.420 | 0.143 | −0.002 | 0.326 |
| NDMI: | 0.330 | 0.241 | 0.354 | 0.072 | −0.072 | 0.268 |
| NDVI: | −0.080 | 0.638 | 0.709 | 0.545 | 0.428 | 0.662 |
| NDWI: | 0.240 | −0.519 | −0.568 | −0.472 | −0.413 | −0.526 |

### C.1.2 Dry season training data

**Table C.2:** Median band and index response for training data samples obtained in the dry season

| Band/Index | OW | FV1 | FV2 | AD | BA | FW |
|---|---|---|---|---|---|---|
| AVE_DSM: | 1134.7 | 1145.0 | 1146.5 | 1188.3 | 1240.1 | 1149.0 |
| B1: | 0.106 | 0.099 | 0.098 | 0.128 | 0.128 | 0.126 |
| B10: | 0.001 | 0.001 | 0.001 | 0.002 | 0.001 | 0.002 |
| B11: | 0.020 | 0.165 | 0.152 | 0.360 | 0.358 | 0.368 |
| B12: | 0.015 | 0.077 | 0.067 | 0.269 | 0.255 | 0.251 |
| B2: | 0.081 | 0.078 | 0.077 | 0.113 | 0.114 | 0.112 |
| B3: | 0.061 | 0.079 | 0.079 | 0.118 | 0.121 | 0.113 |
| B4: | 0.047 | 0.055 | 0.051 | 0.129 | 0.130 | 0.127 |
| B5: | 0.044 | 0.085 | 0.082 | 0.150 | 0.156 | 0.148 |
| B6: | 0.036 | 0.198 | 0.218 | 0.210 | 0.227 | 0.202 |
| B7: | 0.037 | 0.251 | 0.280 | 0.247 | 0.267 | 0.237 |
| B8: | 0.032 | 0.247 | 0.274 | 0.241 | 0.259 | 0.235 |
| B8A: | 0.030 | 0.290 | 0.318 | 0.280 | 0.300 | 0.278 |
| B9: | 0.007 | 0.053 | 0.057 | 0.055 | 0.060 | 0.054 |
| MNDWI: | 0.496 | −0.350 | −0.318 | −0.509 | −0.503 | −0.528 |
| NDII: | 0.213 | 0.270 | 0.354 | −0.131 | −0.096 | −0.143 |
| NDMI: | 0.221 | 0.196 | 0.287 | −0.205 | −0.166 | −0.225 |
| NDVI: | −0.189 | 0.635 | 0.685 | 0.291 | 0.326 | 0.287 |
| NDWI: | 0.310 | −0.514 | −0.549 | −0.342 | −0.365 | −0.350 |

## C.2 LAND-COVER CLASS AREA ESTIMATES

**Table C.3:** Land-cover classification area estimates (in km²) for all 73 Sentinel-2 scenes

| Date | OW | FV (main) | FV | WF | DF | WA | DA | BA |
|---|---|---|---|---|---|---|---|---|
| 3/13/17 | 13.359 | 230.161 | 39.182 | 53.881 | 144.657 | 153.296 | 205.851 | 22.544 |
| 4/2/17 | 13.628 | 248.400 | 53.016 | 86.769 | 87.056 | 200.033 | 147.931 | 26.097 |
| 4/12/17 | 9.925 | 248.218 | 44.077 | 68.048 | 108.910 | 200.237 | 158.156 | 25.359 |
| 5/22/17 | 14.851 | 285.974 | 190.360 | 63.600 | 26.723 | 250.799 | 21.143 | 9.481 |
| 6/11/17 | 11.871 | 272.982 | 132.025 | 106.587 | 30.275 | 282.393 | 20.233 | 6.564 |
| 6/21/17 | 9.987 | 267.347 | 106.565 | 95.511 | 58.182 | 269.696 | 49.011 | 6.632 |
| 7/1/17 | 9.751 | 261.490 | 86.623 | 81.029 | 88.054 | 255.506 | 73.765 | 6.713 |
| 7/21/17 | 10.861 | 251.507 | 48.809 | 46.958 | 159.668 | 162.913 | 179.008 | 3.207 |
| 8/20/17 | 11.038 | 240.755 | 19.774 | 26.977 | 194.930 | 100.747 | 264.901 | 3.808 |
| 9/19/17 | 15.621 | 246.402 | 22.608 | 29.942 | 171.495 | 122.803 | 243.256 | 10.805 |
| 10/4/17 | 9.740 | 247.185 | 13.723 | 24.289 | 167.277 | 53.624 | 337.520 | 9.573 |
| 11/3/17 | 11.093 | 252.114 | 21.708 | 18.042 | 171.481 | 65.195 | 304.162 | 18.582 |
| 11/13/17 | 13.332 | 261.864 | 49.068 | 40.024 | 153.694 | 192.661 | 134.935 | 16.478 |
| 11/28/17 | 21.584 | 261.519 | 62.979 | 69.500 | 95.816 | 270.825 | 65.805 | 14.903 |
| 12/13/17 | 17.475 | 270.940 | 112.714 | 109.729 | 31.762 | 289.299 | 25.140 | 5.873 |
| 1/22/18 | 71.666 | 260.534 | 103.535 | 73.636 | 65.969 | 213.420 | 70.554 | 2.945 |
| 2/11/18 | 8.999 | 255.441 | 54.757 | 67.376 | 130.222 | 160.947 | 183.851 | 1.337 |
| 2/16/18 | 9.127 | 247.016 | 37.499 | 51.847 | 155.591 | 138.790 | 218.758 | 4.302 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 3/23/18 | 134.394 | 247.098 | 104.896 | 39.535 | 51.223 | 181.983 | 98.250 | 5.551 |
| 5/12/18 | 104.572 | 278.687 | 183.595 | 40.724 | 25.577 | 200.394 | 26.969 | 2.413 |
| 5/27/18 | 110.562 | 275.580 | 208.853 | 52.956 | 12.583 | 190.979 | 10.394 | 1.024 |
| 6/1/18 | 70.381 | 287.343 | 207.693 | 47.855 | 20.376 | 191.454 | 18.046 | 2.684 |
| 6/11/18 | 57.730 | 281.230 | 154.542 | 45.983 | 50.563 | 213.702 | 54.910 | 4.270 |
| 6/16/18 | 42.339 | 277.224 | 139.615 | 51.888 | 58.331 | 230.971 | 56.427 | 6.135 |
| 7/1/18 | 25.778 | 275.524 | 139.628 | 65.243 | 61.394 | 218.879 | 73.131 | 3.354 |
| 7/6/18 | 20.715 | 269.311 | 96.211 | 72.609 | 80.273 | 234.164 | 82.570 | 7.079 |
| 7/26/18 | 13.632 | 254.820 | 67.874 | 63.455 | 116.778 | 201.488 | 140.095 | 4.788 |
| 7/31/18 | 15.285 | 264.429 | 69.411 | 55.690 | 127.687 | 149.868 | 175.230 | 5.331 |
| 8/15/18 | 11.671 | 253.869 | 40.894 | 51.723 | 148.847 | 139.401 | 208.304 | 8.221 |
| 9/14/18 | 11.802 | 251.449 | 32.972 | 61.925 | 126.953 | 174.977 | 182.784 | 20.068 |
| 9/19/18 | 11.639 | 251.658 | 35.361 | 59.765 | 131.814 | 148.853 | 216.735 | 7.105 |
| 10/4/18 | 10.293 | 251.214 | 44.702 | 69.220 | 130.197 | 175.042 | 172.660 | 9.356 |
| 12/3/18 | 10.287 | 251.195 | 71.793 | 96.042 | 75.499 | 262.688 | 78.455 | 13.849 |
| 12/28/18 | 16.492 | 276.556 | 149.277 | 100.915 | 22.069 | 258.476 | 32.821 | 6.325 |
| 1/2/19 | 11.938 | 274.727 | 133.152 | 107.929 | 28.887 | 262.862 | 35.450 | 7.986 |
| 2/6/19 | 3.705 | 271.457 | 95.883 | 59.773 | 93.374 | 245.909 | 89.699 | 3.130 |
| 2/26/19 | 11.974 | 231.934 | 62.821 | 77.661 | 129.130 | 200.127 | 146.390 | 2.893 |
| 3/8/19 | 11.935 | 236.280 | 92.023 | 87.524 | 124.041 | 209.553 | 95.742 | 4.883 |
| 3/18/19 | 9.857 | 233.789 | 61.769 | 76.185 | 137.415 | 202.128 | 137.111 | 4.676 |
| 3/23/19 | 8.110 | 222.655 | 31.217 | 60.048 | 167.556 | 188.136 | 175.236 | 9.974 |
| 4/7/19 | 6.576 | 222.927 | 29.448 | 64.848 | 159.073 | 201.822 | 169.045 | 9.192 |
| 4/12/19 | 9.757 | 220.059 | 28.000 | 51.198 | 171.774 | 193.105 | 166.313 | 13.491 |
| 5/7/19 | 12.575 | 247.355 | 91.195 | 49.882 | 120.004 | 194.000 | 124.826 | 12.573 |
| 6/6/19 | 13.002 | 277.231 | 148.543 | 98.488 | 32.485 | 263.767 | 24.079 | 5.336 |
| 6/16/19 | 12.214 | 273.003 | 138.864 | 100.665 | 30.111 | 277.704 | 22.410 | 5.374 |
| 7/1/19 | 27.538 | 272.047 | 118.144 | 82.397 | 58.208 | 254.955 | 44.864 | 4.777 |
| 7/21/19 | 9.706 | 260.709 | 94.308 | 73.852 | 90.307 | 243.633 | 84.561 | 5.855 |
| 7/26/19 | 12.634 | 253.533 | 83.921 | 74.394 | 111.467 | 170.604 | 154.288 | 2.090 |
| 8/5/19 | 11.344 | 257.620 | 50.274 | 35.877 | 158.673 | 117.042 | 226.744 | 5.356 |
| 8/10/19 | 10.181 | 253.479 | 50.716 | 55.754 | 145.942 | 128.885 | 214.207 | 3.767 |
| 9/4/19 | 24.256 | 256.570 | 56.142 | 48.326 | 151.104 | 128.803 | 196.398 | 1.332 |
| 9/19/19 | 10.784 | 251.870 | 33.587 | 42.113 | 143.564 | 103.526 | 261.158 | 16.329 |
| 9/29/19 | 10.212 | 249.800 | 34.914 | 43.085 | 157.656 | 123.106 | 236.359 | 7.799 |
| 10/4/19 | 14.118 | 253.633 | 37.951 | 30.979 | 137.157 | 101.549 | 265.804 | 21.741 |
| 12/18/19 | 125.485 | 272.992 | 163.083 | 49.124 | 15.221 | 201.469 | 28.723 | 4.954 |
| 1/2/20 | 109.037 | 268.241 | 169.685 | 47.383 | 21.829 | 201.589 | 41.504 | 3.662 |
| 1/17/20 | 87.103 | 269.591 | 177.652 | 47.484 | 20.405 | 228.913 | 28.034 | 3.748 |
| 2/11/20 | 95.755 | 262.151 | 143.511 | 46.807 | 35.660 | 211.845 | 62.838 | 4.362 |
| 2/16/20 | 75.769 | 275.205 | 155.854 | 60.368 | 23.257 | 234.068 | 34.459 | 3.950 |
| 3/7/20 | 100.971 | 259.205 | 166.416 | 56.036 | 21.466 | 236.300 | 17.854 | 1.990 |
| 3/22/20 | 72.437 | 262.189 | 151.422 | 62.653 | 23.663 | 243.128 | 42.221 | 5.217 |
| 4/26/20 | 226.345 | 210.387 | 134.810 | 39.050 | 14.218 | 195.322 | 14.035 | 1.591 |
| 5/6/20 | 176.606 | 258.140 | 239.812 | 32.706 | 2.962 | 148.845 | 1.262 | 0.635 |
| 5/11/20 | 110.070 | 269.162 | 243.219 | 43.098 | 10.087 | 176.034 | 9.735 | 1.369 |
| 5/26/20 | 69.582 | 261.151 | 155.424 | 67.262 | 30.385 | 250.193 | 25.727 | 3.207 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 5/31/20 | 59.275 | 256.871 | 157.571 | 62.476 | 44.340 | 241.074 | 38.299 | 3.024 |
| 6/5/20 | 48.709 | 257.646 | 146.963 | 67.547 | 50.750 | 245.757 | 41.708 | 3.851 |
| 6/20/20 | 61.497 | 251.020 | 139.899 | 54.545 | 62.536 | 217.994 | 71.110 | 3.787 |
| 6/25/20 | 51.941 | 255.767 | 127.570 | 67.962 | 65.632 | 212.844 | 77.797 | 3.115 |
| 7/10/20 | 41.607 | 243.519 | 124.636 | 51.904 | 100.115 | 173.120 | 124.079 | 3.951 |
| 7/20/20 | 38.873 | 240.781 | 111.438 | 62.731 | 97.739 | 161.198 | 148.309 | 1.862 |
| 7/25/20 | 39.330 | 250.547 | 124.990 | 40.830 | 90.887 | 162.789 | 150.905 | 2.652 |

# D

## APPENDIX D

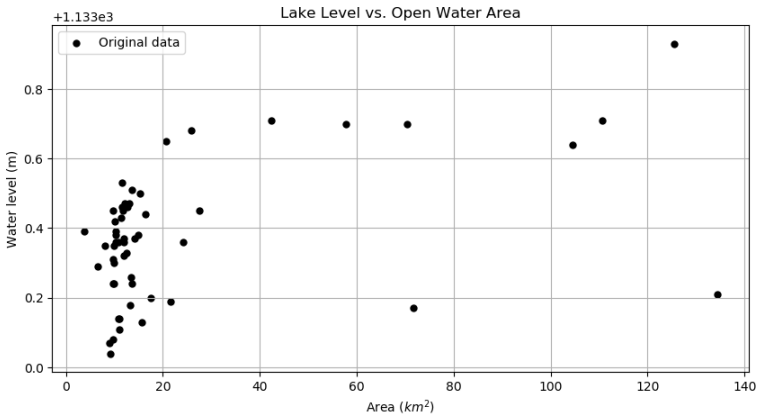



**Figure D.1:** Correlation chart showing Lake Victoria water levels against open water area

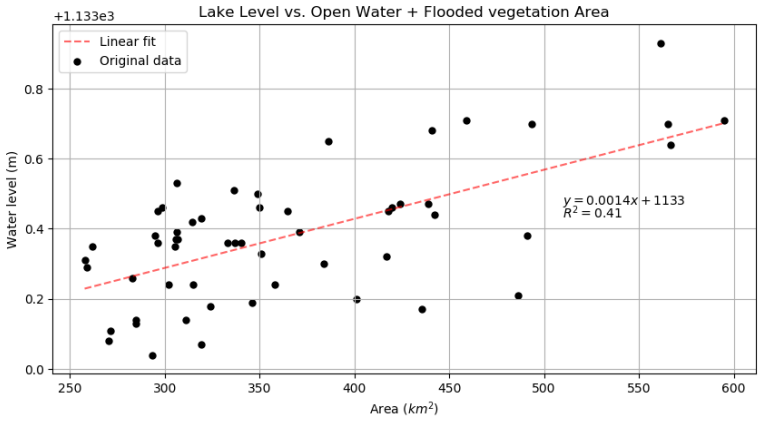



**Figure D.2:** Correlation chart showing Lake Victoria water levels against open water and flooded vegetation area

COLOPHON

This document was typeset using LaTeX. The document layout was generated using the `arsclassica` package by Lorenzo Pantieri, which is an adaption of the original `classicthesis` package from André Miede.