Towards a global high-resolution inundation map derived from remote sensing imagery: African continent application

Etienne Fluet-Chouinard

Degree of Master of Science

Department of Geography
McGill University
Montreal, Quebec, Canada

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Master of Science.

Acknowledgements

My master`s research and the exceptionally enriching experience that was acquired during this period could not have been possible without the help and support of many individuals. I first want to thank my supervisor, Prof. Bernhard Lehner for his trust and support during the development of my research project. Prof. Lehner's insight directed me toward a research project that made me feel like I would provide a valuable contribution to the field of global freshwater resource inventorying. The work presented in this thesis ended up being very close to my very first proposal on a wetland mapping project, hence demonstrating the level of agreement between Prof. Lehner's interests and mine. My project could not have taken the scale it has without the help of my committee member Prof. Margaret Kalacska who lent some of the computing horsepower to which she has access to run my scripts. In that regard, I also need to thank Eva Snirer for her technical support in using the computers. For sharing the Master`s experience in a new environment from start to end with me, and figuring it out as we went, I thank Camille Ouellet-Dallaire. I also want to thank the entire Global Land & Water Resource Modeling Lab and Prof. Ramankutty, who have made the lab a pleasant and enriching working environment. For her reviewing prowess and sharing my thesis related stress, I need to thank Avni Malhotra. I must also thank Dr. Lisa-Maria Rebelo for accepting to oversee my internship in Addis Ababa during my second year of master's, as well as the Global Environmental and Climate Change Centre (GEC3), Prof. Lehner and Mr. Warren for their financial support which made this internship possible. Also for their financial support, I would like to thank the organizing committee of the AGU Chapman Conference on Remote Sensing of the Terrestrial Water Cycle held in February 2012.

Biographical Sketch

Upon graduating from my Bachelor degree in Applied Environmental Geomatics from the University of Sherbrooke, and after completing internships on topics ranging from aquatic pollution to carbon-accounting forestry and agricultural soil erosion, I became convinced that my technical skills could also take me closer to working in the fields that interested me the most: hydrology and freshwater resources. Although I arrived late in the admission process, Prof. Lehner kindly accepted me as his master's student at McGill University. Working for Prof. Lehner allowed me to approach the topics of freshwater conservation and ecosystem services from my GIS and remote sensing background. My master's also introduced me to working at the global scale, which I now consider central to my research identity.

During my master`s, I had the opportunity to present my work at multiple conferences. I presented posters at the AGU Fall Meeting in San Francisco in December 2011, AGU`s Champan Conference on Remote Sensing of the Terrestrial Water Cycle in Hawaii in February 2012, and also at INTECOL's 9th International Wetlands Conference in Orlando in June 2012. Thanks to Profs. Lehner and Ramankutty`s contact network, I will be commencing doctoral studies in Limnology and Marine Science under the supervision of Prof. Peter McIntyre at the Center for Limnology of the University of Wisconsin – Madison. There, I hope to pursue global freshwater ecology research, building on the knowledge acquired at McGill, while maintaining my links with Prof. Lehner`s lab.

Table of Contents

List of Figures

List of Tables

List of Equations

Executive Summary

Wetlands are recognized as valuable landscapes for their contribution to biodiversity, ecosystem services and population livelihoods. However, current global wetland inventories do not spatially represent wetland extent at a spatial and temporal resolution appropriate for conservation and management purposes. Among the best existing global inventories, the *Global Lakes & Wetlands Database (GLWD*; Lehner & Döll, 2004) is a static database assembled from various existing data sources that unfortunately suffers from the inconsistency among its data sources. Another, the *Global Surface Water Extent Dataset (GSWED*; Prigent et al. 2007; Papa et al. 2010) produced from a multi-satellite method is capable of monthly measurements but possesses a coarse spatial resolution incapable of discriminating distinct surface water bodies. Faced with the limitations of current global inventories, a new methodological approach is required to provide the improved wetland inventory needed by the research and conservation communities.

This thesis investigates a methodology capable of producing a high-resolution (~ 500 m) surface water extent map by spatially downscaling the coarse resolution (~27 km) inundated area estimates of GSWED. The methodology inspired by Bwangoy et al. (2010) has a pragmatic and straight-forward design to ensure and ease its global application. The work of this thesis consists of an initial implementation and validation of the methodology across the African continent. The downscaling approach relies on the topographic and hydrographic information from the globally available HydroSHEDS data (Lehner et al., 2008) to distribute inundated area at the finer resolution to the most topographically inundation prone areas. Thirteen hydro-topographic variables were computed from HydroSHEDS and then consolidated into a single inundation probability map with the use of decision tree learners. The decision trees were trained on regional inundation maps and subsequently employed to generate a topographic probability of inundation map at high-resolution for the entire continent. The probability map is turned into an inundated/non-inundated map by

splitting the probability distribution into two (inundated/non-inundated) with a defined threshold value. A threshold value is chosen for each GSWED cell to produce an inundation map replicating the inundated area estimates of GSWED within the cell at the finer resolution. Distribution of inundated area for individual GSWED cells led to undesirable linear features predominant along the edge between cells of contrasting inundation levels. The presence of edge effects was reduced by a Moving Window Thresholding (MWT) process which reallocated inundated area among adjacent GSWED cells to smoothen the inundation distribution, but that also sacrifices authenticity to the original GSWED estimates. The output inundation maps were subsequently filtered with a majority filter to eliminate lone pixels and smoothened inundation outlines.

To represent the maximum wetland extent at different timescales, two sets of inundated areas estimates were downscaled as high-resolution inundation maps with this MWT downscaling procedure: 1) the mean annual maximum (MAMax) estimates were calculated for each cell from the monthly estimates of GSWED between 1993 and 2004; 2) the fusion maximum (MaxFusion) was generated from a fusion of the time-series maximum (TSMax) also calculated from GSWED, with the wetland area from GLWD. The MaxFusion estimates were produced to correct some data gaps of GSWED, as well as to offer a more complete and reliable maximum wetland extent map. The MAMax and MaxFusion estimates respectively totalled 1339 and 2779 thousand km^2 of wetland area across the continent; higher than most previous estimates for Africa.

Validation of the spatial distribution of inundation at the finer resolution exhibited high levels of agreement against reference regional maps (Overall Accuracy ~ 92%; KIA ~ 80%). Over selected wetland study sites, comparisons of the MaxFusion downscaled map with the global land cover GLC2000 (Mayaux et al. 2004) and wetland database GLWD indicated that the downscaled map possessed slightly lower but more consistent agreement with GLC2000 than GLWD did. Taken as a whole, some observable defects such as the remaining “edge effects

limit the reliability of the downscaled inundation maps generated from this initial implementation. These defects could be further reduced with optimization of the downscaling procedure and expansion of training/validation areas in future iterations of the methodology. Regardless, the level accuracy of the tested methodology is considered satisfactory to pursue production of a first version global inundation map. Possible follow-up applications making use of the downscaled inundation maps such as a global hydro-geomorphic wetland classification as well as prediction of the extent of future wetland could support global wetland and freshwater system assessments that are data limited today.

Résumé Exécutif

Bien que l'importance des milieux humides pour la biodiversité et les services écosystémiques soit reconnue, les bases de données actuellement disponibles ne sont pas en mesure de décrire globalement les charactéristiques biophysiques des milieux humides de façon utile à des fins de gestion et de conservation. Parmi les inventaires globaux de milieux humides disponibles, la *Global Lake & Wetland Database (GLWD)* (Lehner & Döll, 2004), , est une représentation statique constituée de plusieurs sources qui est cependant spatiallement inconsistente. Un autre inventaire, le *Global Surface Water Extent Dataset (GSWED)* (Prigent et al. 2007; Papa et al. 2010) produite à partir d'une méthode multi-satellitaire, possède une faible résolution spatiale incapable de différencier des plans d'eau distincts. Compte tenue des limites des actuels inventaires, une nouvelle approche est nécessaire pour générer le nouvel et amélioré inventaire que demande diverses communautés de chercheurs.

Cette thèse examine la possibilité de produire une représentation spatiale d'inondation et de milieux humides globale à haute résolution (~ 500 m) à partir d'une réduction d'échelle des estimés surfaces inondées de faible résolution (~27 km) du GSWED. La méthode inspirée par Bwangoy et al. (2010) a été développée de façon pragmatique et simple afin d'assurer son application globale sans heurt. Cette thèse est une application initiale de la méthodologie à des fins de validations sur le continent africain. La méthode de réduction de résolution repose sur de l'information topographiques et hydrographiques globales provenant des données HydroSHEDS (Lehner et al. 2008) pour distribuer la surface inondée de GSWED à la plus fine résolution aux endroits les plus prompt à l'inondation. Treize variables hydro-topographiques furent calculées à partir d'HydroSHEDS et ensuite consolidée dans une probabilité d'inondation calculée pour chaque pixel de haute résolution d'HydorSHED,S grâce à un arbre de décision. L'arbre de décision fut entraîné et validé avec des images satellitaires régionales d'inondation et fut utilisé par la suite pour générer des probabilités d'inondation sur l'ensemble du continent. La carte de probabilités d'inondation

résultante est ensuite transformée en carte d'inondation en utilisant une valeur seuil divisant la distribution de probabilités en deux. Une valeur seuil est choisie pour la surface de chaque cellule de GSWED pour répliquer la surface inondée de GSWED.

En tant qu'estimé de surface provenant de GSWED, les valeurs mensuelles entre 1993 et 2004 furent agrégée pour produire des estimés du maximum annuel moyen (MAMax) ainsi que le maximum historique (TSMax). Pour produire des estimées fiables représentant la surface maximale inondée et pour corriger pour certaines des lacunes des estimés de GSWED, les estimés du TSMax furent fusionnés à celles de GLWD, générant les estimées maximales de fusion (MaxFusion). La surface totale estimée pour l'ensemble du continent africain pour MAMax et MaxFusion est estimées respectivement à 1339 et 2779 milliers de km^2, plus élevée que la plupart des estimés précédents. La conversion des probabilités en carte d'inondation pour chaque cellule de GSWED a eu pour effet indésirable de produire des traits linéaires aux bordures entre les cellules de GSWED, particulièrement entre les cellules de niveau d'inondation contrastante. La présence de ces traits linéaires fut réduite grâce à l'usage de la valeur seuil en concert avec une fenêtre mobile permettant la réallocation de surface inondée entre les cellules voisines de façon à produire une représentation plus lisse, mais sacrifiant son authenticité par rapport aux estimés originaux de GSWED ou de la fusion.

La validation de la distribution spatiale de l'inondation à la plus fine résolution a démontré un bon accord (Précision globale ~ 92%; KIA ~ 80%) lorsque comparée à des cartes régionales d'inondations ou de milieux humides. De plus, la comparaison du résultat cartographique avec le GLC2000 et le GLWD sur quelques sites particuliers a indiqué une concordance avec GLC2000 consistante, malgré qu'inférieure à celle de GLWD. Dans l'ensemble, des problèmes visibles à l'œil, comme les traits linéaires restants limitent la fiabilité du produit résultant de cette première implémentation de la méthodologie. Ces défauts peuvent être

réduits dans le futur grâce à l'optimisation du procédure de réduction d'échelle ainsi que par l'ajout de données d'entraînement et de validation additionnels. Malgré les défauts mentionnés, le niveau de précision de la méthodologie testée peut être considérée suffisant pour poursuivre son développement et son application globale. D'autre projets de recherches découlant de celui-ci et faisant usage d'une carte d'inondation global à haute-résolution peuvent avoir faire à un classification hydro-géomorphique de milieux humides ainsi qu'une prédiction de la surface inondée dans le future sous des scénarios de changements climatiques.

List of Acronyms

AMSR-E: Advanced Microwave Scanning Radiometer for EOS

AVHRR: Advanced Very High Resolution Radiometer

DEM: Digital Elevation Model

EO: Earth Observation

ERS: European Remote Sensing satellite

ETM+: Enhanced Thematic Mapper

MAMax: Mean Annual Maximum inundated extent from GSWED estimates; can be used to describe the downscaled map or the estimates.

GLWD: Global Lakes and Wetlands Database – dataset of Lehner & Döll, 2004

GSWED: Global Surface Water Extent Dataset – dataset of Papa et al. 2010

MaxFusion: Historical Maximum Inundated Extent generated from the fusion of GLWD and GSWED; can be used to describe the downscaled map or the estimates.

JERS: Japanese Earth Resource Satellite

MWT: Moving Window Thresholding

SSM/I: Special Sensor Microwave/Imager

SAR: Synthetic Aperture Radar

SRG: Seeded Region Growing

SVT: Single Value Thresholding

UNEP: United Nations Environment Program

WCMC: World Conservation Monitoring Centre

1 Introduction

Wetlands can be loosely described as transitional ecosystems located at the interface between terrestrial and aquatic systems (Mitsch & Gosselink 2007). Wetlands possess distinguishing features such as standing water or a water table near the land surface as well as vegetation that is tolerant or adapted to saturated soils (Mitsch & Gosselink 2007). Wetland formation and evolution are constantly influenced by ecological, hydrological and geomorphic processes (Wood & Halsema 2008) and in turn, wetlands play an important role in the global hydrological and biogeochemical cycles (Finlayson et al. 2005). In the terrestrial water cycle, wetlands act as buffers that maintain base river flows during drier periods of the year, reduce peak-flows as well as recharge aquifers (Bullock & Acreman 2003). Globally, wetlands contribute to the carbon cycle, acting both as sinks and sources of carbon (Ringeval et al. 2010; Richey et al. 2002).

1.1 Biodiversity, Ecosystem Services & Livelihoods

Covering only 5 to 8% of Earth's ice free land surface, wetlands exhibit a disproportionately high number of species compared to marine and terrestrial ecosystems (Revenga et al. 2000, Mitsch & Gosselink 2007). Wetland and riparian species pool is drawn from both the terrestrial and aquatic communities (Robinson et al. 2002). Wetlands are biodiversity hotspots and the high level of endemism present in inland water systems result from the numerous physical barriers in catchments and river systems (Finlayson et al. 2005). In riparian systems, species diversity is affected by the main source of disturbance of inundation which defines the level of habitat heterogeneity and connectivity with the river channel (Ward et al. 1999).

Wetlands also perform a wide range of environmental functions that provide socio-economic benefits contributing to human well-being which together are referred to as ecosystem services (Finlayson et al. 2005). Examples of ecosystem services include *provisioning services* related to food and water; *regulating services* such as water filtration, regulation of floods, drought, land degradation,

and disease; *supporting services* such as soil formation and nutrient cycling; and *cultural services* such as recreational, spiritual, religious, and other nonmaterial benefits (Finlayson et al. 2005). Wetland-agricultural systems are often particularly important for impoverished and vulnerable groups that may rely on wetlands for coping strategies to food scarcity (Rebelo et al. 2009).

1.2 Wetland Degradation & Management

Wetlands are fragile ecosystems, easily degradable through natural processes or unsustainable exploitative human interventions. It is widely accepted that wetlands are being degraded at an unprecedented rate, vastly beyond that of natural loss, yet there are few reliable accounts of the current state of wetlands. The loss of wetlands worldwide has been estimated at 50% of those that existed since the start of the 20th century (Dugan 1993). Most of the losses come from the developed world, where wetlands have been drained and converted for industrial agricultural practices (Wood & Halsema 2008). Additionally, floodplain wetlands particularly are also threatened by the flow regulation of dam impoundment which alters inundation patterns and reduces lateral connectivity (Ward 1989; Tockner & Stanford 2002).

Overexploitation of wetland *provisioning* services is done at the expense of *regulating* and *supporting services* that mitigate the impacts of degradation and ensure ecosystem resilience (Finlayson et al. 2005; Gordon et al. 2009). The loss of regulating and supporting services can, over time, hinder the ability of wetlands to maintain provisioning services, leading to further reduction in human well-being (Finlayson et al. 2005). The subsequent impacts of wetland degradation and loss of resilience can be far-reaching and often irreversible, affecting not only the local communities that rely on wetlands for a range of benefits, but also those downstream whose livelihood strategies depend upon a particular supply of freshwater (Rebelo et al. 2009). The effect of climate change on the global water cycle is expected to exacerbate the degradation of many wetlands (Falkenmark et al. 2007). As a result, wetland dependent impoverished communities are more

vulnerable to the impacts of climate change (Finlayson et al. 2005). Wetland degradation is a threat to the unique flora and fauna present in these ecosystems. It is likely that freshwater dependent species are more at risk than their marine and terrestrial counterparts because of the acute pressure on inland water systems (Finlayson et al. 2005). The number of wetland-dependent species is considered an underestimate as many species-rich regions have not yet been inventoried (Finlayson et al. 2005), also making it difficult to assess wetland species loss.

Poor consideration of wetland importance in decision-making is one of the major factors leading to wetland degradation, and management decision influencing wetlands rarely consider the wider benefits and underestimate the value of wetlands because the links between wetlands and livelihoods have not been comprehensively demonstrated. Thus, lack of knowledge and information on how to exploit wetland services in a sustainable and equitable manner remains a hurdle towards an adequate use of the wetland resources (Rebelo et al. 2009). To achieve sustainable use of wetland resources in a climate change context, existing pressures on wetlands should be removed and effective management practices should be implemented to cope with the altered drought/flood cycles and mitigate its effects (Finlayson et al. 2005; McCartney & Smakhtin 2010).

1.3 Wetland Classification and Inventory

Inventorying activities consist of the primordial task of cataloguing the occurrence, extent and biophysical characteristics of wetlands. The Ramsar Convention on Wetlands, an intergovernmental treaty originally providing a framework for national action and international cooperation for the wise use of wetlands, defines wetland inventory as "the collection and/or collation of core information for wetland management, including the provision of an information base for specific assessment and monitoring activities" (Ramsar - Framework for Wetland Inventory - Resolution VIII.6). Inventory activities support subsequent, assessment and monitoring activities which consider the risks and changes affecting wetlands, are together considered critical for a sustainable wetland

management framework for the maintenance of ecosystem services (Davidson & Finlayson 2007; MacKay et al. 2009). Inventorying efforts however experience issues of appropriately defining what consists of a wetland. The variety of possible hydrological and ecological conditions of wetlands complicates the task of adopting a wetland definition easily transferable across regions and useable for diverse purposes, particularly for large scale inventories.

The most internationally accepted and globally relevant wetland definition is the one of the Ramsar Convention, which 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" (Matthews 1993). Ramsar's broad definition originates from a desire to capture all habitats of migratory birds, but retains hydrology and presence of water as its central criteria (Scott & Jones 1995). Ramsar's definition has not been adopted by all national inventories around the world and other definitions are in use, habitually reflecting the purpose and setting for which they were created (Wood & Halsema 2008).

1.4 Criticism of Global Inventories

Global wetland inventories are valuable to the Ramsar Convention for evaluating the effectiveness of its efforts to identify, protect and manage wetlands (MacKay et al. 2009). Current baseline inventories remain gravely incomplete and inconsistent, leading the Ramsar Convention to prioritize filling the gaps in baseline inventories as a means to monitoring the temporal and spatial dynamics of wetlands across the globe (Rosenqvist et al. 2007). With time, these broad recommendations from the Ramsar Convention for improving wetland inventories have proven to be overly ambitious for standard in situ inventorying by the members, and the convention is now further exploring usefulness of Earth Observation (EO) technologies (MacKay et al. 2009).

Ramsar's emphasis on filling wetland/inundation inventory gaps globally has been echoed by many in the literature in a variety of contexts: a) to assist in the assessments of ecosystem services, in a context of sustainable agriculture for poverty alleviation (IWMI 2007); b) to estimate surface water storage and volume change in hydrology and river discharge modelling (Alsdorf & Lettenmaier 2003); c) to validate river flooding modules of hydrologic and hydraulic models (Adam et al. 2010; Decharme et al. 2008); d) to improve accounting of greenhouse gas emissions (Melack & Hess 2004), for freshwater populations and habitat assessment (Dudgeon et al. 2006), for systematic freshwater conservation planning (Nel et al. 2009); and e) to improve understanding of the global water system (Naiman et al. 2011).

To address these calls for an improved wetland/inundation baseline inventory, this thesis attempts to provide a global wetland inventory that is superior to the currently existing ones (reviewed in section 2.2). The pragmatic approach taken (section 3) uses existing global datasets derived from remote sensing imagery to generate a global high-resolution inundation map. The approach specifically designed for global application has the potential to fill many of the existing gaps in current global inventories. The method used consists of downscaling global surface water extent data with topographic and hydrographic information (location of rivers and waterbodies). The downscaling methodology was inspired by the one Bwangoy et al. (2010) used for producing a wetland map of the Congo river basin.

2 Literature Review

2.1 Inventorying Methods

In response to the need for wetland inventorying, various methods have been designed to generate new baseline information on wetlands at different scales. Approaches reviewed in this section employing either earth observation (EO) technologies, topographic data, or a combination of the two are discussed in the context of providing a global baseline wetland data. The approaches incapable of generating a complete global coverage, such as ground surveys and interpreted aerial photography are excluded from the review.

2.1.1 Earth Observation

As recognized by Ramsar, the most promising approaches are methods employing EO (Davidson & Finlayson, 2007). Three main types of remote sensing methods are employed to map wetlands through either their distinct vegetation or the presence of surface water: optical sensor, Synthetic Aperture Radar and passive microwave. These methods are reviewed in this section.

Optical

Wetland and riparian areas can be mapped at moderate (~1 km) and high (< 500 m) resolutions using optical sensors through characterizing of their particular vegetation cover and of open water. Optical data, particularly assembled as time series, can delineate emergent and submerged wetland vegetation by detecting vegetation pigmentation and leaf water content (Townsend 2001; Baker et al. 2006; Vancutsem et al. 2009, Silva et al. 2008), but cannot penetrate through cloud cover nor extract information on sub-canopy inundation (Rosenqvist et al. 2007). These limitations prevent optical sensors alone from providing reliable baseline inundation extent data and capturing the dynamic aspects of wetland/inundation occurring beneath the canopy. Availability of optical data coverage globally has been used to produce global land cover maps (Loveland et al. 2000; Hodges et al. 2001; Friedl et al. 2007). In these land cover datasets, wetlands generally constitute a single class at a single time snapshot, capturing

mostly permanent wetland extents but not intermittent ones (Darras et al, 1998). The absence of intermittent wetlands from these land cover maps underestimates wetland cover and provides an example of the limitations of optical sensors for global scale wetland mapping.

Synthetic Aperture Radar

Synthetic Aperture Radar (SAR) sensors have shown good capabilities to measure and delineate surface water extent at high spatial resolutions (JERS~93m, ERS~100m, PALSAR~10-30m). Conclusive applications have been carried out over many regions: the Amazon (Hess 2003; Alsdorf et al. 2007), Congo (de Grandi et al. 2000, Rosenqvist & Brikett 2002, Alsdorf et al. 2007), Pantanal (Costa and Thelmer 2007), northern peatlands (Moghaddan et al. 2003) and others (Lowry et al. 2009). SAR is particularly appropriate over tropical regions because of its ability to penetrate cloud cover (Rosenqvist et al. 2007). The success of SAR is due to its non-specular backscattering, appropriate for capturing information about open water, submerged vegetation and sub-canopy inundation (Rosenqvist et al. 2007).

SAR operating in C and L bands are the most commonly used for wetland mapping. The short wavelength of C-band (5.6 cm) is appropriate to retrieve information on forest canopy structure because of the backscatter attenuation in forest canopy, and the longer wavelength of L-band (23 cm) can easily penetrate most vegetation canopies to provide information on sub-canopy inundation. Long-wavelength SAR (L-band) has been identified by the Ramsar Wetland Convention as particularly appropriate for reaching wetland inventory targets and filling the identified inventory gaps regarding inundation extent mapping (Rosenqvist et al. 2007). Newer sensors such as ALOS, Envisat and Radarsat-2 possess polarimetric capabilities, possibly yielding additional information on the canopy or soil (Rosenqvist et al. 2007).

Although SAR satellites can provide systematically collected data, large data volumes and unique vegetation backscatter have generally limited its use to a regional scale studies and prevented long-term assessment (Papa et al. 2010). Moreover, the varying levels of backscattering contamination by vegetation across regions make the delineation of wetlands site-specific (Prigent et al. 2007). An example of large scale SAR data from the Global Rain/Boreal Forest Mapping (GRFM/GBFM) projects demonstrated that temporally consistent SAR scenes could be combined as near-continental scale mosaics (Mayaux et al. 2000; Rosenqvist et al. 2004). The outcome of the projects established that two blanket acquisition dates are insufficient to capture the complex spatio-temporal variability of many wetland regions (Rosenqvist et al. 2007). The GRFM/GBFM mosaics were nonetheless utilized to categorize the wetland classes in the GLC2000 land cover map (Mayaux et al. 2004). Lastly, the use of object-oriented segmentation methods has proven adequate for processing SAR backscattering data and such methods have been used by many for wetland exercises (Costa 2002, Hess 2003, Hamilton 2007, Durieux 2007).

Passive Microwave

Passive microwave sensors such as SSM/I and AMSR-E are characterized by high temporal resolution (daily data, aggregated to monthly, globally) but low spatial resolution (0.25°), and have been used to measure surface water extent through changes inland surface emissivity. With passive microwave sensors, inundated areas can be detected because of their low microwave emissivity and high emissivity-polarization difference, even under dense canopy (Prigent et al. 2007). Because of the coarse resolution of passive microwave sensors, mixing models which distinguish the contribution from different end members (water, vegetation, bare soil, etc.) found in a given pixel can be used (Sippel et al. 1998; Mialon et al. 2005). The correction of the passive microwave land surface emissivity signal contribution from vegetation is usually performed using regionally developed empirical relationships (Prigent et al. 2007).

An example of this type of sensor for global inundation mapping was by Prigent et al. (2007) where passive microwave was used in combination with European Remote Sensing satellite (ERS) backscatter and AVHRR optical data. As part of this method, detection of inundation chiefly relies on the passive microwave measurements. However, in this case, passive microwave correction is accounted for through scatterometer response, or Normalized Vegetation Difference Index (NDVI) estimates from AVHRR in arid regions (Prigent et al. 2007). Prigent et al.'s (2007) multi-satellite method, updated and extended by Papa et al. (2010) with a new mixing model, shows difficulties discriminating between very saturated moist soil and standing open water, as noted by Adam et al. (2010). Some debate exists over whether passive microwave C-band can effectively detect sub-canopy inundation (pers. comm. R.Schröder – Jet Propulsion Laboratory - 2011). A global inundation product similar to that of Prigent et al. (2007) using AMSR-E and QuikSCAT sensors is in development and was tested over large Eurasian river basins (Schröder et al. 2010).

2.1.2 Topography

The distinct geomorphic features of wetlands conditioned by the presence of flowing or standing water can also be used to map wetlands. Not only the presence of water but its source, flow pathway, duration and depth of water in turn define the biophysical characteristics of wetlands. The geomorphology of the floodplain determines the degree of disturbance from flooding and dictates vegetation succession trajectories (Ward & Stanford 1995; Gurnell 1997). This relationship between topography and surface water has led to many efforts (discussed in this section) to delineate wetlands using topographic data alone. The wetland mapping approaches relying on topography can aim to delineate floodplains as distinct geomorphic features or to model hydrodynamic dynamics of inundations.

Delineating valley-bottoms as landforms is relevant in the context of wetland and riparian area mapping because they represent the extension of riverine inundation.

Different studies have used topographic attributes to delineate valley-bottoms for purposes of sediment deposition modeling (Gallant & Dowling 2003) or stream environment classification (Islam et al. 2008). In the simplest approaches, valley-bottoms were distinguished from hillslopes by a flatness index or a slope transition threshold while the more sophisticated method reviewed ensured valleys surrounded rivers by growing valley clusters around thalwegs (Straumann & Purves 2008). Similarly to valley landforms mapping, the possibility of basin morphology alone, quantified with geomorphic variables, being capable of delineating a known floodable area using geomorphic factors was investigated by Manfreda et al. (2008). They found that topographic index (Kirkby, 1975) strongly characterized areas prone to flood inundation. A threshold topographic index value was estimated to generate the flooded area matching the one reported by the Italian study basin's authorities.

Hydrodynamic models can also provide simulated surface water estimates where discharge and topographic data are available. Wilson et al. (2007) modeled the temporal and spatial changes of surface inundation over the Amazon floodplain with the first implementation of a 2D hydrodynamic model at a large scale. With the inclusion of bathymetric data, the LISFLOOD-FP model was used to analyse the hydraulic flood wave of the main stem of the Amazon River. A variety of other models exist (e.g., Hervouet and Van Haren, 1996) – all of them prove to be very data intensive and are usually accompanied with ground surveys. At the global scale, absence of reliable data had kept models fairly simple in their design. A coupling of the Soil, Biosphere, and Atmosphere (ISBA) land surface model and Total Runoff Integrating Pathways (TRIP) river routing model have allowed for surface inundation to be modelled (Decharme et al. 2008). The model was tested globally against remotely sensed inundated surface, flooding water volume and in situ discharge measurements (Decharme et al. 2012).

2.1.3 Combining Topography and Remote Sensing

In order to improve wetland classification accuracy, supplementing spectral data with ancillary topographic information was attempted by wetland remote sensing studies discussed below.

Masking

In some cases, topography is used as a filter to identify flood prone regions prior to a spectral based classification of vegetation/habitats within the identified areas. Akin to the topography based inundation and landform mapping approaches reviewed earlier, flood prone areas can be delineated in a variety of ways. Hamilton et al. (2007) included topography in their remote sensing object-oriented approach in order to discriminate floodplain areas using a relative elevation index. Alternatively, Islam et al. (2008) used topography to delineate wetlands/floodplains in two steps. A river network was generated from a DEM with a flow accumulation algorithm, and was then utilized to buffer valley bottoms at a certain distance of the rivers. Then, a slope threshold limited wetlands to areas flatter than a defined slope. Once the wetlands were delineated, they were further classified by their vegetation cover with data from the ETM+ optical sensor.

Classification

With the proper processing tools, topography can also be included as a variable along spectral bands in the classification scheme. This was the case Whitcomb et al. (2009) who included both SAR backscatter data with ancillary topographic data such as slope, elevation and distance to water bodies into a decision tree classifier. Bwangoy et al. (2010) also combined optical and SAR remote sensing data with a topographic information into a decision tree classifier to generate a per-grid cell probability map of each cells likelihood of being a wetland (i.e., flooded). Applying a cut-off probability value to the probability map produces the final flooded/not-flooded map from the probabilities. DEM topography was used to produce various first (i.e., slope) and second (i.e., curvature) order

topographic derivatives, some of them based on river networks of varying densities, that were all included in the decision tree classifier estimating the probabilities. The results of the Bwangoy et al.'s (2010) decision tree classifier showed local topography, as relative elevation to the closest stream, to be the most valuable discriminator of wetland cover. Nonetheless, all the 27 variables contributed to the classification tree procedure, reinforcing the idea that multi-source data are useful in characterization of wetland land cover (Bwangoy et al. 2010).

2.2 Global Wetland Inventories

Most of the EO approaches reviewed in the previous section have not designed to be expanded from their original extent to a global coverage. The methods that have been applied globally and have yielded global land cover and water body databases only offer incomplete portraits of wetland extent because of their focus on other landscapes or because of methodological limitations. As a result, the quality of global inventories is inferior to what has been achieved over smaller study regions. In the following section, spatially explicit inland wetland inventories of complete global coverage are chronologically reviewed from the late 1980s to present. Inventories attempting to represent the full extent of wetlands globally, such as the Ramsar database of wetlands of international importance or the Surface Water Body Database (Slater et al. 2006) were excluded from this review.

2.2.1 Early Global Inventories

The first exhaustive global natural freshwater wetland inventories were designed to account for natural wetland methane emission. By compiling navigational charts, Matthews and Fung (1987) estimated global wetland cover at 5300 thousand km^2 while Asselman & Crutzen (1989) arrived at 5700 thousand km^2. These estimates were approximately twice those from earlier global wetland area estimates (Lieth 1975, Whittaker & Likens 1975, Ajtay et al 1979) because of their broader wetland definition including seasonal and permanent freshwater

ecosystems (Spiers 2001). Both Matthews and Fung (1987) and Asselman & Crutzen (1989) inventories distinguished between five inland wetland classes, but the classification schemes differed, discriminating basic wetland types on the basis of vegetation presence/absence. Asselman & Crutzen (1989) considered rice paddies as an artificial wetland class and estimated its area at 1300 thousand km^2. Matthew & Fung's (1987) dataset provided fractional cover of different wetland types at the resolution of 1° x 1° while the one of Asselman & Crutzen (1989) was originally compiled at a coarser resolution of 2.5° latitudinal and 5° longitudinal and was interpolated to 1° x 1° by Stillwell-Soller et al. (1995). Independently, UNEP-WCMC's Global Wetland Map precisely delimited wetland extent from expert interpretation of navigational maps and imagery estimated the area of freshwater wetlands to 5600 thousand km^2 globally (Dugan, 1993).

2.2.2 GRoWI (1999)

The Global Review of Wetland Resources and Priorities for Wetland Inventory (GRoWI) is a database generated from a bottom-up aggregation of national inventories. GRoWI stems from an effort supported by Wetlands International and the Environmental Research Institute of the Supervising Scientist on behalf of the Bureau of the Ramsar Convention on Wetlands (Finlayson & Spiers 1999). The study had the objective of providing a baseline against which to assess trends in wetland degradation and conservation (Spiers 2001). The wetland area estimates from GRoWI, although not spatially represented, could be mapped with countries as the smaller unit.

The GRoWI standardized inventory data collection utilizes the Ramsar Convention on Wetlands' definition to compile inventories with different definition and classification systems (Spiers 2001). Over two thirds of the inventories compiled already employed classification systems closely resembling the Ramsar classification system (Finlayson et al. 1999). In total, the study reviewed 188 national inventories and 45 global to continental scale inventories or

areal estimates, most of them paper based and published in international and national NGO reports or books between 1980 and 1998 (Finlayson et al. 1999). The total global wetland area compiled through GRoWI accounted for 12792 thousand km^2, which is higher than that found by previous remote sensing inventories. Regardless, because of gaps in national inventories, this value is still considered an underestimate although it represents the highest wetland area estimate (Finlayson et al. 1999). Because of the various weaknesses of the approach, among them the lack of agreement on wetland definition, and many gaps and inaccuracies in the information, Finlayson et al. (1999) concluded that with information available at the time, it was not possible to provide an acceptable figure of the areal extent of wetlands at a global scale. Key recommendations ensuing from this review emphasized the importance of gathering standardized information on location, extent and biophysical characteristics of wetlands including variations of area and water regime.

2.2.3 GLWD (2004)

The Global Lakes and Wetlands Database (GLWD) is a spatially explicit global inventory depicting surface water features which was initially generated for hydrological and climate modeling but also supports ecological applications (e.g., Vyverman et al. 2007; Harrison et al. 2005;). The GLWD is divided into three levels: level 1 and 2 respectively represent vector shorelines of large and small lakes and reservoirs, and level 3 represents lakes, rivers, reservoirs and inland wetlands at a 30-arcsecond (~ 1 km at the equator) pixel resolution including the waterbodies of level 1 and 2. The GLWD-3 is divided into 12 classes that generally follow the definition of wetlands of Ramsar and exclude artificial (e.g., rice paddies) and degraded wetlands. GLWD- 3 constitutes an inventory of the maximum extent of wetlands and surface water comparable to other inventories previously mentioned, with an estimation of global wetland area between 8,300 and 10,200 thousand km^2, depending on interpretation of fractional wetland classes, but was reported as 9167 thousand km^2 in the Millenium Ecosystem Assessment (Finlayson et al. 2005). The GLWD global wetland area estimate

exceeded wetland estimates from previous global land cover maps but was in good agreement with estimates from the literature (Lehner & Döll 2004). The GLWD-3 was generated from the aggregation of various wetland and land cover maps generated between 1992 and 2000, but mainly based on UNEP's World Conservation Monitoring Center (WCMC 1993). The data sources were overlaid and the extent was defined as area covered by at least one of the sources. In defining classes, priority was given to those of WCMC over ArcWorld and DCW. The GLWD was compared to land cover datasets and global wetland estimates from literature but was not validated through ground truthing (Lehner & Döll 2004). Although the wetland extent of the map is static, the 12 types of wetlands offer the possibility of assigning some sort of temporality to each of them. Similar to GRoWI, the GLWD also suffers from problems of definition and inconsistency across regions from the data sources that constitute it (Finlayson et al. 2005).

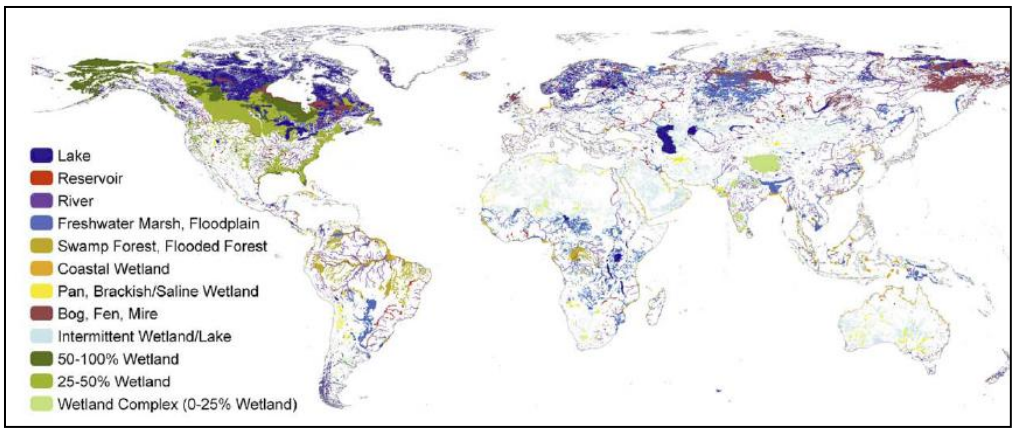


Figure 2.1: Global Map of GLWD - from Lehner & Döll, 2004

2.2.4 GSWED (2007, 2010)

The multi-satellite method capable of estimating global monthly surface water area developed by Prigent et al. (2007) and updated by Papa et al. (2010) described in section 2.1.1 was applied globally to produce a Global Surface Water Extent Dataset (GSWED). The data in GSWED is expressed as monthly composite surface water fraction of 0.25° x 0.25° cells (773 km^2 at the equator).

The multi-satellite method of Prigent et al. (2007) which was updated and the time record expanded to 12 year by Papa et al. (2010). The record revealed a decrease in total surface water extent over the time series (Papa et al. 2010). Globally, the mean annual maximum (i.e., the averaged maximum extent reached within a year) of GSWED stands at 10,100 thousand km^2 and the historical maximum on record at 16,400 thousand km^2. The historical maximum area figure from GSWED represents the largest natural and artificial surface water area estimate found in literature. This new method does not discriminate between surface water bodies, natural, or artificial wetlands (rice paddies). GSWED was used to estimate methane emissions (Bousquet et al. 2006; Ringeval et al. 2010), to study water storage variations in large water basins (Papa et al. 2008; Frappart et al. 2008) and to validate a global flooding model (Decharme et al. 2008).

2.3 Comparison of Inventories

As there are only a few trustworthy global inventories, inter-comparison amongst them is important to establish an agreed upon global wetland area estimate. In the comparisons below, disagreements can be explained by the different definitions, source material and methodology utilized by each inventory.

2.3.1 GLWD versus GRoWI

The Millenium Ecosystem Assessment (Finlayson et al. 2005) included a comparison of "the two best global wetland extent estimates": the GRoWI and the GLWD. Both these inventories use definitions of wetlands inspired from the Ramsar definition and provide different but comparable total estimates of natural inland and coastal wetlands. The divergence among total wetland area estimates suggests that there still exists a large range of uncertainty regarding global wetland extent. Arguments that the total area remains an underestimate based on missing information from national inventory data are said to be "well established" by the MEA (Finlayson et al. 2005). No approximation of the magnitude of this underestimation has ever been made for lack of information. At the regional scale, important discrepancies exist between GroWI and the GLWD, predominantly over Europe and the Neotropics (Table 2.1). Even though

GroWI's estimated that the area for the Neotropics is the largest among all regions, it is considered an acute example of underestimation partly caused by the exclusion of savannas as intermittent wetlands (Ellison 2004). The largest difference found in Europe is likely a result of an underestimation of wetland cover by GLWD due to its restrictive representation of ecologically significant wetlands, which have been extensively altered and managed in that region of the world.

Table 2.1: Estimates of inland (non-marine) wetland area for GRoWI and GLWD; Reproduced from Finlayson et al. (2005).

Ramsar's Geopolitical Region	**GRoWI: Global Review of Wetland Resources (Finlayson et al. 1999)**	**GLWD: Global Lakes and Wetlands Database (Lehner and Döll, 2004)**
	(thousand km^2)	
Africa	1213-1247	1314
Asia	2043	2856
Europe	2580	260
Neotropics	4149	1594
North America	2416	2866
Oceania	358	275
Total Area	**~12792**	**~9167**

Note on GRoWI: Not all wetland types are equally represented in the underlying national inventory data. Some countries lack information on some types of wetlands.

The gap between the two inventories for total area is also caused by a misinterpretation of what should be included as a wetland. In the comparison, the GroWI methodology included lake, reservoir, ponds and floodplains (often including rivers) as part of their national inventory aggregation, while classes of lakes and reservoirs were excluded from the GLWD area estimate. By including lakes and reservoirs classes (accounting for 2670 thousand km^2) into the GLWD area estimate, and so reconciling the definition of wetlands of the two inventories, the total wetland area of GLWD increases to 11846 thousand km^2 (while maintaining median fraction class areas), in much closer agreement with GroWI's estimate of 12798 thousand km^2. Moreover, the remaining gap can be reduced by modifying the interpretation of the three fractional wetland classes (50-100%, 25-

50%, wetland complex 0-25%) of the GLWD. The widely used estimate of 9170 thousand km^2 is based on mean values for fractional classes known to underestimate boreal and subarctic peatland area of Gorham (1991). By considering fractional wetlands classes as being equal to the upper limit of their ranges, the total wetland area of the GLWD again increases to 12798 thousand km^2, getting even closer to the estimate of GRoWI. Interestingly, the gross wetland map produced as a by-product of the GLWD and derived as the maximum area per 0.5° cell identified by various incomplete inventories (Matthews & Fung 1987, Stillwell-Soller et al. 1995, GLCC – Loveland et al. 2000, MODIS – Hodges et al. 2001) also provides a total area close to GroWI's, of 11711 thousand km^2.

2.3.2 GLWD versus GSWED

The development of the GSWED generated a new baseline inventory completely independent from the previous ones. Comparison of this new dataset with previous estimates allowed important conclusions about the quality and level of agreement between estimates. The GLWD, being the only spatially explicit global inventory, is a natural choice for comparison. This comparison has been done for large scale patterns by Papa et al. (2010) and at a finer scale by Adam et al. (2010). Unfortunately, problems of definition still complicate the interpretation of the comparisons.

Comparison of Papa et al. 2010

As a simple means of visual comparison, (Papa et al. 2010) plotted the latitudinal distribution of GSWED's mean annual maximum extent and the historical maximum extent over the 12-year record with all freshwater habitat classes of GLWD-3. The global wetland area (including lakes & reservoirs) calculated by Papa et al. (2010) for GLWD amounted to 12,800 thousand km^2, falling between the mean annual maximum and historical maximum GSWED. Visually, the latitudinal distribution of the GSWED was "largely parallel" to the one of GLWD. Papa et al. (2010) noted that the two datasets "agreed exceptionally well" between

latitudes 80N and 40N as well as below 15N, but that both GSWED estimates exceeded the values of the GLWD between 10N and 30N. These discrepancies between 10N and 30N were explained by Papa et al. (2010) by the inclusion of irrigated artificial wetlands in GSWED while the GLWD did not. Visual cross-referencing of the maps with the Digital Global Map of Irrigation Area from Siebert et al. (2006) by Papa et al. (2010) supports the artificial wetland explanation, particularly important in rice growing regions of Asia. The conclusions which can be taken from Papa et al.'s (2010) mostly qualitative assessment are few because both datasets disagree fundamentally in what they represent given that the method that generated GSWED cannot discriminate natural and artificial inundated areas. Furthermore, coastal cells for which inundated fraction is likely to be contaminated by oceans were filtered from the GSWED prior to comparison (Prigent et al. 2007; Papa et al. 2010). It can be assumed that a fraction of the inundation reported in these cells comes from coastal area water bodies, and that the total global inundated area is likely higher than the one reported. Regardless, for an appropriate comparison to be made, one would need to account for irrigation and artificial wetlands as done by Adam et al. (2010).

Comparison of Adam* et al. *2010

The evaluation by Adam et al. (2010) to assess the GSWED for validating flooding flow and storage processes in a global hydrological model is more comprehensive than Papa et al.'s comparison. Rather than using the raw GSWED inundation estimates, Adam et al. (2010) created a Naturally Inundated Areas (NIA) dataset by removing artificially inundated areas of irrigated rice agriculture from the GSWED estimate, using the monthly MIRCA2000 irrigation dataset (Portmann et al. 2010). Mean annual maximum and historical maximum area of NIA as well as the GLWD wetlands were aggregated to the 0.5° scale by Adam et al. (2010) for comparison. Unlike the comparison of Papa et al. (2010), coastal cells and large lakes absent from the NIA were removed from GLWD prior to comparison. Adam et al. (2010) expected the GLWD's representation of

maximum wetland extent to be larger than the extent of NIA, except in extreme cases. They found that in most 0.5° cells the mean annual maximum of NIA agreed well with the GLWD, but also found instances going against their expectations, where gross overestimations were made by NIA. Agreement between the two datasets improved when outliers were removed from NIA. The per-cell comparison of Adam et al. (2010) provides the best assessment of the GSWED local inundation estimates, even though it excludes fractional wetland classes of GLWD from the comparison.

2.4 Conclusion

Remote sensing satellite imagery has established itself as the most appropriate method for wetland inventorying because automated analysis and classification allows for delineating wetlands over large regions. The most promising approaches involve the use SAR L-band which benefits from cloud penetration and sub-canopy inundation detection capabilities as well as polarimetric response to accurately delineate inundated areas at high spatial resolution. Moreover, methods combining remote sensing data from different sensors together and with ancillary data sources, for instance topography, is increasingly popular. Sadly, many of the developments in remote sensing sciences are not easily transferrable to the global scale. Only the combination of passive microwave sensor data with other satellites was successfully applied globally to produce the GSWED. Although it suffers from a resolution too coarse to distinguish individual water bodies, the GLWED is the only observation based inventory of surface water available and offers many new exciting possibilities for global surface water studies. Before the GSWED, the only two reliable global wetland inventories, the GRoWI and GLWD had been compared extensively, highlighting the uncertainties over the total amount of wetlands existing worldwide.

3 Research Questions and Objectives

3.1 Gap in Knowledge

For a variety of reasons discussed in the literature review, such as incompatible wetland definitions and remote sensing limitations, spatial wetland inventorying efforts have fallen short of a truly complete and consistent global coverage. Consequently, the GLWD remains the best spatially explicit global wetland database available of spatial resolution capable of distinguishing among wetlands, despite its unverified level of accuracy. Because of GLWD's recognized limitations, its total wetland area estimates are questioned and considered underestimations of true wetland extent, the magnitudes of which are unknown. Additionally, the uncertainties and inconsistencies that GLWD inherited from its source data being untraceable make it impossible to correct for them. Likewise, the possibility of integrating temporal variation components is non-existent. These numerous limitations of the GLWD provide very little prospect for improvement, meaning that an alternative approach should be pursued to generate a superior baseline wetland inventory capable of fulfilling the needs of the research and conservation community.

3.2 Approach

The recent advances of global hydrographical data development and remote sensing methods offer an opportunity to improve upon current wetland/inundation inventories by synthesizing available datasets in novel ways. The opportunity lies within two particular datasets both possessing a global coverage: the Global Surface Water Extent Dataset (GSWED) inundated area estimates (Prigent et al. 2007; Papa et al. 2010) and HydroSHEDS topography-derived river network (Lehner et al. 2008). HydroSHEDS is a globally seamless dataset of high-resolution hydrography including river networks derived from an equally high-resolution DEM of the year 2000 (Lehner et al. 2008). It provides static information on the specific location and drainage size of linear river features, but does not provide information on the width and area covered by surface water bodies. GSWED - described in section 2.2.4 - provides information on the

inundated fraction of each cell's 0.25° x 0.25° surface area at a monthly temporal resolution over a 12 year record. Neither of these datasets alone possesses the required information to generate a high-resolution inundation map, but new information can emerge from an appropriate combination that preserves the complementary strengths of each dataset (Table 3.1). A combination preserving HydroSHEDS' superior spatial resolution and GSWED's superior temporal resolution can result in an inundation map that possesses unprecedented high spatial and temporal resolutions at once for a global dataset, constituting a major improvement over GLWD. Furthermore, synthesizing new datasets allows to diverge from recompiling existing inventories as done for GroWI and GLWD to produce a new continuous observation-based global wetland inventory.

Table 3.1: Comparison of HydroSHEDS and GSWED data, spatial and temporal resolutions. *(*HydroSHEDS' superior spatial resolution and GSWED's superior temporal resolution are italicized*.)*.

	HydroSHEDS	**GSWED**
Data type	Hydrographic and topographic information.	Observation of inundated area
Temporal resolution	Static – snapshot of year 2000	*Monthly – between 1993 and 2004*
Spatial resolution	*3, 15, 30 arcseconds (~ 90, 500, 1000 m at equator)*	0.25 x 0.25 (~27 km x 27 km at equator)

In essence, the combination consists of downscaling (i.e., refining spatial resolution) of the GSWED inundated area among the smaller pixels of HydroSHEDS to produce a higher resolution map. The downscaling needs to be accurate in how it spatially distributes inundated area at the finer resolution but should also ensure that the area of downscaled inundated pixels are equivalent the original estimate of GSWED. The hypothesis behind this combination is that the processes controlling wetland formation, such as surface inundation, local runoff or groundwater flows, can be captured with the hydrographic and topographic

information of HydroSHEDS alone and can help determine the location of inundation within the cells of GSWED.

This thesis develops and evaluates a method producing a global wetland and inundation map from the combination of HydroSHEDS and GSWED. In order to combine GSWED and HydroSHEDS in a way preserving each dataset's strengths, a downscaling procedure inspired by Bwangoy et al. (2010) was implemented over the continent of Africa in this thesis. The approach taken involves that an inundation/wetland prediction model based on hydro-topographic information is trained over certain regions where inundation location is known, to then generate a high-resolution probability map of wetland/inundation occurrence over the wider study region. The model in effect consolidates the hydrographic and topographic information of HydroSHEDS into a single per-pixel probability distribution. A threshold value is applied to the probability distribution, classifying all the pixels' exceeding the threshold as inundated, and the ones below as non-inundated. The threshold values are chosen to reproduce the same inundated area as dictated by GSWED estimates. The use of continuous probabilities distribution allows using different threshold values to replicate the inundated area of different times at the finer resolution and thus preserve GSWED's temporal resolution with a single probability distribution (Figure 3.1). Through the downscaling of inundated area at multiple dates, new information on duration and seasonality of inundation emerges from the process. With regards to the GSWED inundated area estimates, some of the caveats limitations exposed by Papa et al. (2010) and Adam et al. (2010) were expected to be transferred to the downscaled inundation map. In an attempt to fill the gaps and correct the biases of GSWED estimates, the GLWD was used to supplement the GSWED estimates in this work. The merger of GLWD and GSWED could provide a more complete baseline of maximum wetland extent than either individual dataset can, which would be reflected in the downscaled output. The downscaling process is tested in this thesis by mapping maximum wetland extent from GSWED and GLWD over Africa.

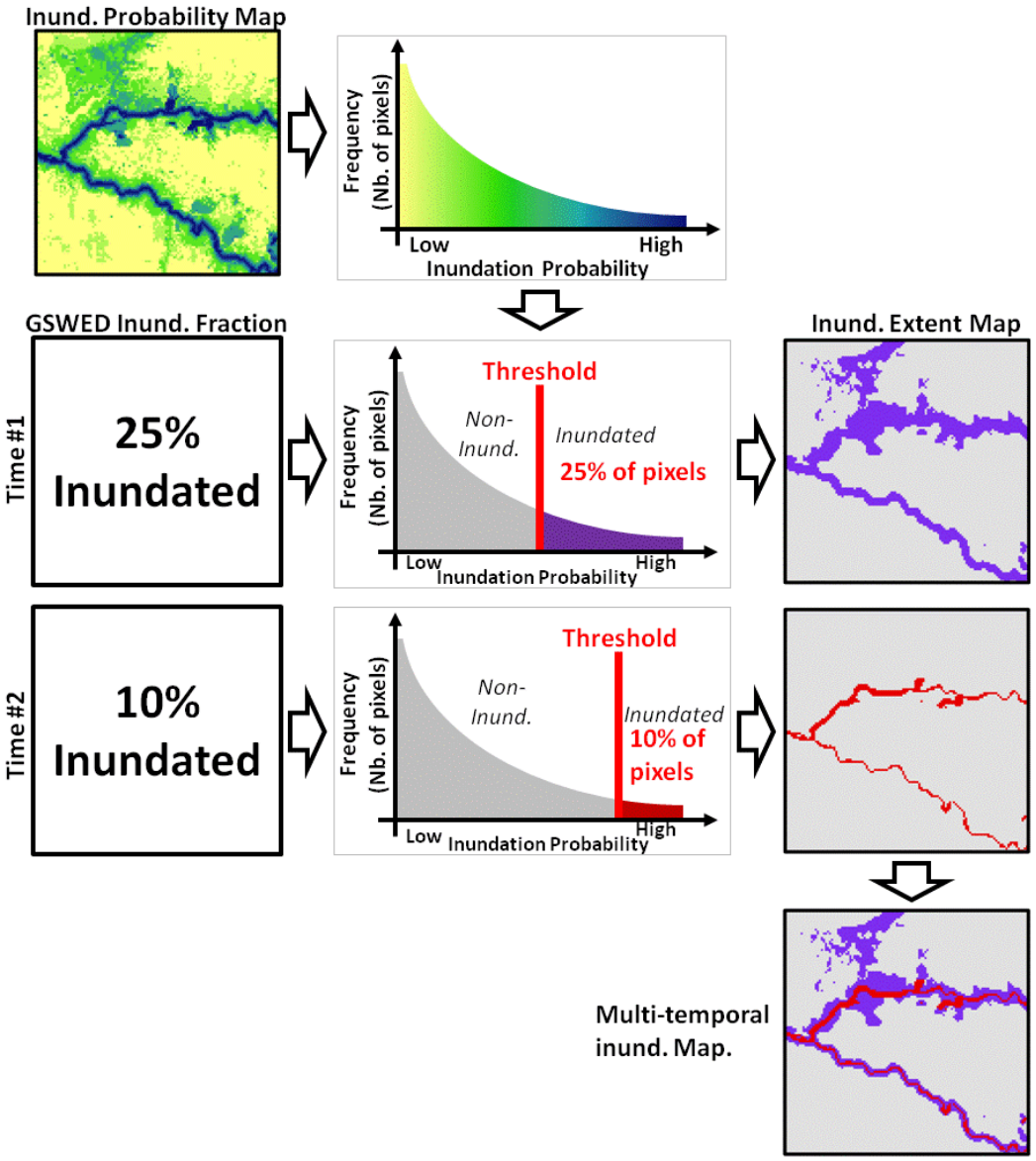


Figure 3.1: Conceptual diagram of a downscaling procedure generating two inundation maps from a single probability map. The downscaled maps at the two timeframes can then be compiled into an inundation duration map (*bottom*).

3.3 Research Questions

In developing a method capable of producing a downscaled global wetland/inundation map, two sources of uncertainty affect the map produced by the downscaling process: the spatial distribution and the total area of inundation. Through implementation of the downscaling process and the production downscaled inundation maps, this thesis answered two research questions, each addressing a source of uncertainty. The first question regards the fundamental concept behind the downscaling of GSWED and asks if the information contained in HydroSHEDS can be as appropriate predictors for the downscaling process:

> *How accurate can hydrographic and topographic data from HydroSHEDS alone predict spatial distribution of inundation/wetland extent if compared to regional maps?*

This research question was addressed through the accuracy assessment of the spatial distribution of downscaled inundated area, achieved by comparing the downscaled output against regional validation wetland maps. Comparison was accomplished only for equal inundated areas, as to eliminate the effect of different inundated areas on the measured accuracy metrics. To provide an accuracy benchmark for the downscaled inundation map, the spatial distribution of GLWD wetlands was also compared against some of the same reference datasets, in order to recognize improvements of the downscaled output over the leading global wetland map if any are present.

The second research question touches on the subject of the fusion of GLWD and GSWED data sources described in section 3.2, and the improvements stemming from it:

> *How well can the GLWD supplement the wetland/inundated area estimates of the GSWED?*

Addressing this question was achieved by developing and implementing a fusion procedure for the two datasets and comparing the fused estimates to the two sources. Analysis of the fusion output was done for the inundation estimates for

the spatial coverage as well as the total inundated area. Insights from the analysis can help conclude on GLWD's inherent capacity to complement GSWED and on the weakness of the fusion procedure implemented, to improve it in future implementations.

4 Data & Methods

The methodology presented in this section aims to generate a high-resolution inundation map from a downscaling of coarse resolution inundation estimates. The many elements of the methodology, as well as the four main steps in the process are shown in Figure 4.1. The numbers in the flow chart of Figure 4.1 refer to the section numbers where each step is described. As explained in the previous section, the method producing the inundation map relies on a model estimating inundation probability occurrence. The decision tree predictive model model (described and justified in section 4.4.1) is based exclusively on hydrographic and topographic information from HydroSHEDS and is trained with regional high-resolution inundation maps. The first step of the methodology is to extract the information relevant to surface water presence from HydroSHEDS by calculating 13 different hydrographic and topographic variables (section 4.3.1). The other data necessary for the predictive model are training data with inundated/non-inundated labels, taken from the regional reference datasets (section 4.2.3). The reference datasets are sampled and then randomly divided into training and validation subsets (section 4.3.2). The training subset is utilized to train the decision tree (section 4.4.1), which generates inundation class probabilities across the African continent (section 4.4.2). With the probability map, the inundation estimates from GSWED and from the fusion with GLWD (section 4.3.3.) were distributed among the high-resolution pixels based on each pixel's probability (section 4.4.3). Distribution of inundated area at the higher resolution was done with two distribution methods which threshold the probability values: the Moving Window Thresholding (MWT) and the Single Value Thresholding (SVT). The MWT developed specifically for this application is compared to the quicker alternative method of SVT. Once the distribution is accomplished, a majority filter is applied onto the inundation maps to eliminate isolated pixels (section 4.4.4).

Finally, assessment of the downscaling process – i.e., the spatial distribution of inundation at the finer resolution – is achieved through comparison of the

downscaled inundation maps against the validation subset from the reference datasets (section 4.5.1). The inundation estimates from the data fusion of GLWD and GSWED are analyzed to evaluate their suitability as inundation estimates for downscaling (section 4.5.2). The replication of the downscaled inundation map to the original inundation estimates by the downscaled inundation map is also assessed through quantification of their differences (section 4.5.2). Finally, the downscaled fusion estimates are compared to other datasets and inventories over selected case studies to evaluate their overall agreement (section 4.5.3). The methodology described here was implemented with the Geographic Information System (GIS) ESRI® ArcGIS version 10 software and Python geoprocessing scripts. ArcGIS 10 software was chosen over other software to remain compatible with other HydroSHEDS research applications. The decision tree module "TreeLearner" of the Orange Data Mining Software (v.2.0b) was embedded into the Python geoprocessing scripts (see Appendix).

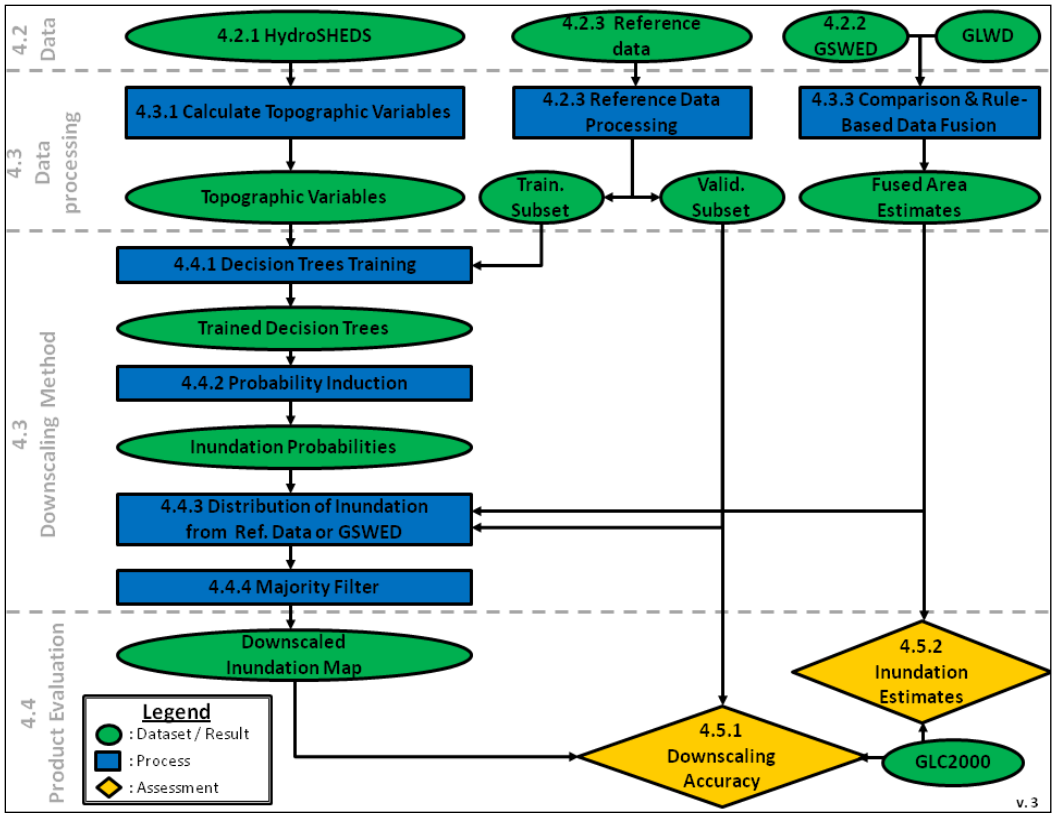


Figure 4.1: Flow chart of the methodology; the numbers in flow chart refer to the sections where they are described

4.1 Study Area

The area of this study covers the entire African continent. African wetlands, like other tropical wetlands are considered as biodiversity hot spots on the planet (Junk 2002), and have received increasing attention in the past decade (Grandi et al. 1996; Grandi et al. 2000; Rosenqvist & Birkett 2002; Murrayhudson et al. 2006; Seiler et al. 2009; Bwangoy et al. 2010; Rebelo et al. 2009; Rebelo et al. 2010). Africa's equatorial region also holds the Congo River, one of the few remaining unimpounded large river system with vast floodplains whose flooding patterns and ecology are still poorly understood (Keddy et al. 2009). Many remote regions of Africa are home to freshwater species yet to be inventoried (Abell et al. 2008). The remoteness of some of the continent's regions like the Congo has lead to the use of remote sensing to monitor and study wetlands of the continent. According to the GLWD, the African continent contains 1314 thousand km^2 of wetlands, of which approximately 40% consist of freshwater marshes/floodplains and 15% of lakes or intermittent wetlands/lakes. Africa, and particularly its Sub-Saharan part, is home to large inland wetland complexes such as the Sudd Marshes, Niger Inner Delta, Logone Floodplain, Lake Chad, Okavango Inland Delta as well as large coastal deltas of the Niger, Senegal and Zambezi rivers (Figure 4.2).

Vulnerable groups such as some rural sub-Saharan African communities are particularly dependent on the ecosystem goods and services provided by wetlands, primarily related to agriculture and fisheries (Neiland & Bene 2008; Rebelo et al. 2009). The African continent is setting for unprecedented development of its the water sector in the coming decades, particularly targeting the irrigation and hydropower sector (Economic Commission of Africa 2003). Managers of the region are therefore faced with dilemmas and conflicts of interest regarding the wetland resources of the continent (Rebelo et al. 2009; Darwall et al. 2011). However, knowledge of wetlands for the region that should support a sustainable management of the resource is inadequate (Taylor 1995; Finlayson et al. 1999). By producing a new wetland map for Africa to test the developed methodology,

this thesis hopes to contribute critical knowledge to support sustainable management of the wetlands of the region.

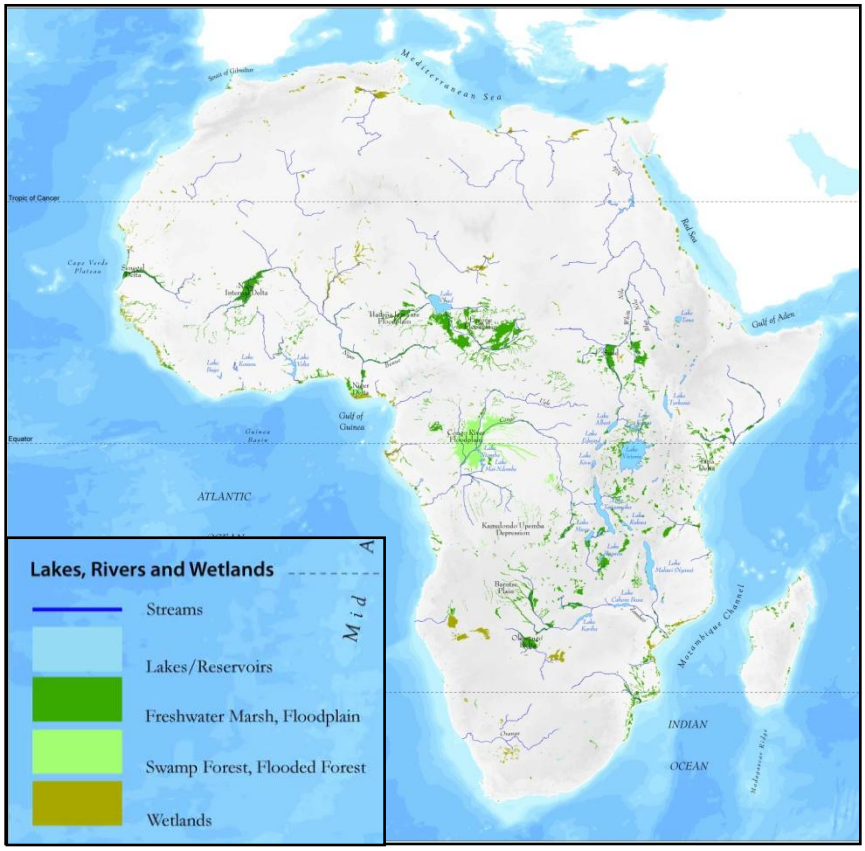


Figure 4.2: Maps of wetlands of Africa from the Africa Water Atlas of UNEP-WCMC (Dugan 1993). Note: similarities exist with the GLWD because the WCMC was used as an input to the production of the GLWD.

4.2 Data sources

Datasets of the methodology are separated into three groups based on their role in the methodology: HydroSHEDS (section 4.2.1), GSWED (section 4.2.2), and reference datasets (section 4.2.3).

4.2.1 HydroSHEDS

HydroSHEDS is a hydrographic data suite derived from the global Digital Elevation Model (DEM) of the Shuttle Radar Topographic Mission (SRTM) and is available at three different spatial resolutions: 3, 15 and 30 arcseconds between

60° of latitude North and South. For the current application, the 15 arc-seconds resolution (or 500 m at equator) was chosen as the scale of this study to maintain compatibility with other HydroSHEDS applications. At each resolution, HydroSHEDS offers three raster layers depicting topography which are used in the methodology: DEM, Flow Direction (FlowDir) and Flow Accumulation (FlowAcc) layers. HydroSHEDS also includes river networks and catchment outlines in vector format. In the context of this thesis, the appellation "pixel" is used to refer to the units of HydroSHEDS. To produce a river network from the SRTM topography, Lehner et al. (2008) applied a combination of two void-filling algorithms on the DEM before a standard GIS procedure filled spurious sinks on the coverage. Lehner et al. (2008) ensured correct river network typology by having the 3 arc-second DEM undergo hydrologic conditioning with modifications of the DEM such as deepening of open water surface, weeding of coastal zones, stream burning, valley moulding, and barrier carving, ensuring the correct topology of the river network. The resulting conditioned elevation could generate an appropriate drainage direction map and a flow accumulation map. This flow accumulation map was then used to generate a river network by using a contributing area threshold of 1000 pixels above which each pixel would be considered part of the river network. Cells possessing a contributing area above the threshold were considered part of the river network. Defining rivers with their contributing area overlooks the effect of climate in forming rivers (e.g. in an arid environment, rivers may only form with a larger contributing area in contrast to a humid environment where they may form with a smaller area). Lehner et al. (2008) upscaled their hydrologically conditioned 3 arcsecond resolution DEM and associated river network to 15 and 30 arcseconds resolutions (or 500 mand 1000 m at equator) and ensured transferability of river networks across scales by using stream burning techniques onto the upscaled DEMs. HydroSHEDS has not been systematically assessed but comparisons with local maps indicate that it is superior to other global river networks (HYDRO1k [USGS 2000], DCW [ESRI 1993]) in most areas (Lehner et al. 2008).

4.2.2 GSWED

The inundated area estimates for this thesis originate from the Global Surface Water Extent Dataset (GSWED) of Papa et al. (2010) described in section 2.2.4. Data of GSWED is expressed as the inundated fraction of each 0.25° × 0.25° cell (or ~27 km at equator; each cell is 773 km^2). Inundation fractions from natural (e.g., wetlands) and artificial (e.g., rice paddies) sources are measured for each month of the time series between January 1993 and December 2004, i.e., a total of 144 monthly values per cell. Global quality assessment of the data has been limited to comparison to static wetland and surface water maps andwas compared to regional SAR inundation extend data and precipitation data (Prigent et al. 2007; Papa et al. 2010; Adam et al. 2010). Assessment of the GSWED uncovered its tendency to underestimate low inundated fractions and over estimate high inundated fractions because of the method's inability to detect small patches of wet or dry land covering less than 10% of a cell (Papa et al. 2010). Following these observations, the uncertainty of the inundated extent measurements of GSWED is assumed to be approximately 10% (Frappart et al. 2010). In the context of this thesis, the appellation "cell" exclusively refers to the units of the GSWED. Cells of the GSWED have many roles in the implementation of the downscaling procedure, and have been used as the main processing unit of the methology.

Two versions of the GSWED data were used in the methodology. The first version of the dataset had coastline cells masked because of the contribution of the ocean to them as explained by Prigent et al. (2007). In order to map coastline areas, a second version of GSWED without the mask was acquired from Dr. Fabrice Papa (October 2011). Because the second version was acquired during the implementation of the methodology, it was only used for its inundation estimates data fusion with GLWD, while all previous steps were done with the first version of GSWED. The first version of GSWED was first acquired from the authors as a text file in which point coordinates were rounded to the second decimal, requiring some modifications to generate suitable cell coverage of 0.25°

average dimensions. The applied modifications included modifying center coordinates to locate them at equal distance as well as slightly modifying the dimensions of the cells to ensure coverage with no gap or overlap. Because of the unaligned distribution of point coordinates, the points could not be converted to raster and were instead converted to polygons (Figure 4.3). These steps could have introduced slight errors of geolocation to the cells – never more than 0.005° - which could be resolved with the more precise coordinates of the second version of the data in future work. Cells for which HydroSHEDS topographic data were absent were eliminated – 8 in total located along coastlines or over small islands. The African continent consists of 39544 land cells which were selected from the global dataset to replicate the same continent outline as used in HydroSHEDS.

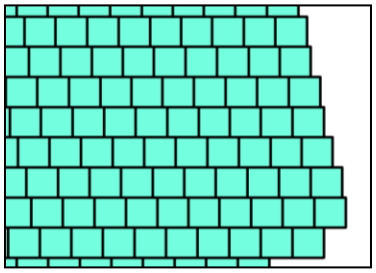


Figure 4.3: GSWED cell outline displaying longitudinal shift between rows of cells. The shift is more pronounced in higher latitude.

To avoid creating differences in geolocation between the two versions, the estimates of the second version were transferred, in the process, to the first version's geolocated cells, including the 1680 initially masked coastline cells. Within coastal cells, the oceanic area beyond the DEM's extent was subtracted from the inundated area of these cells to remove the contribution of the ocean. However, some of the coastal cells of unfiltered version of the GSWED data, still possess no estimates around large inland water bodies, like for example Lake Victoria (Figure 4.4).

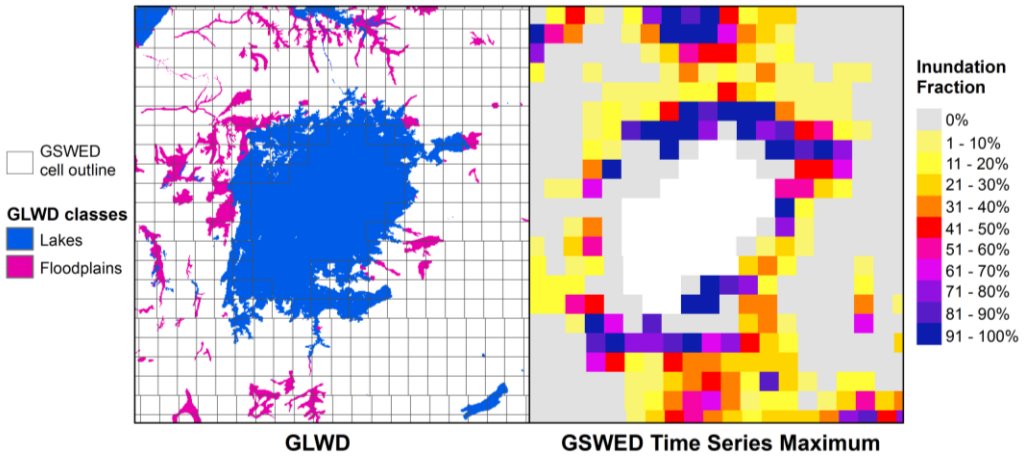


Figure 4.4: Comparison of GSWED and GLWD over Lake Victoria, highlighting the cells with zero estimates surrounding it in GSWED

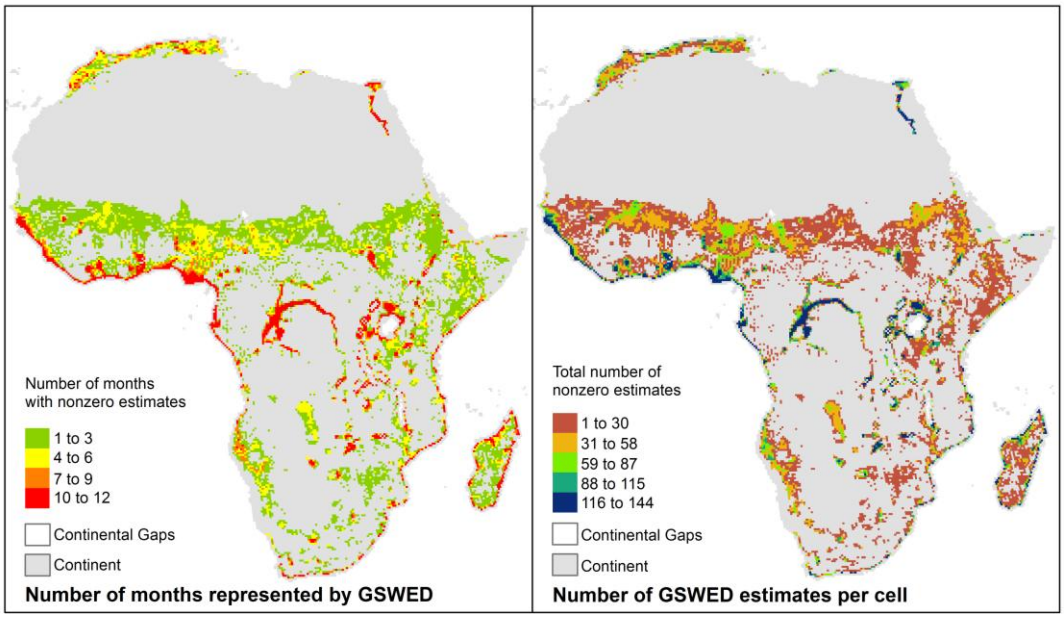


Figure 4.5: Number of different months with GSWED nonzero estimates and total number of nonzero GSWED estimates per cell across Africa.

Over the African continent, only 25.37% of all cells possess at least one inundation fraction value above zero (i.e., nonzero). In fact, only a small fraction (3.8%) of the cells actually possesses nonzero estimates for every month of the year. The cells with nonzero estimates for only part of the year (2 to 11 months) are also very rare (15.7% of all cells). Some of the numerous cells with inundation fraction of zero could be due to natural drying, or because the inundated area falls below detection level (~10%) (Prigent et al. 2007; Papa et al. 2010), or signal attenuation from dense canopies (Adam et al. 2010). Figure 4.5 shows the number of different months having at least one nonzero estimate as well as the total number of nonzero estimates across the continent. A good correspondence exists between the areas where many months are represented and high total number of nonzero estimates.

4.2.3 Reference Datasets

Reference datasets refer to a group of regional datasets assembled from various sources and used in this thesis for training and validation. These datasets generally consist of high resolution surface water, inundation or wetland extent maps produced from regional remote sensing studies. Because of their role for training and validating the predictive model, the characteristics of the reference data sources directly affect the performance of the model. Data sources therefore need to be selected adequately to represent the diversity of settings of wetlands and surface waterbodies found in Africa. Not all the reference data sources use the same definitions of wetlands/inundation. The admissible variety of depicted wetlands by the reference datasets ranges between permanent water bodies and ephemeral wetlands along the aquatic-terrestrial continuum. Temporary streams, floodplains and all other wetlands are found within this range. A uniform representation of wetlands along this continuum by the reference data sources ensures a consistent model output, and prevents a biased representation favouring certain wetlands types to be propagated to the model results. In the current implementation, quality of reference datasets was deemed secondary for testing

the mapping methodology. For a final inundation map product to be produced, only reference datasets of certain quality should be included. For example, the FAO Africover land cover data derived from manual interpretation of remote sensing imagery was not included as a reference dataset in the current implementation but should be in the future.

To test the method developed in this thesis, only two reference data sources have been collected and included in the methodology (Table 4.2). Among the collection, one dataset (CARPE) broadly represents all wetlands and the other (UMD) is limited to open water bodies. These two reference data sources were used because of their reasonably large coverage requiring limited processing. As explained in section 4.3.2, the data from these two sources were sampled and combined together to constitute both the training and validation subsets.

Table 4.1: Reference datasets collected; both data sources are present in both training and validation sets.

Data Source	Region	Wetland/Inundation Definition	Original Resolution	Validation
University of Maryland – Water Mask (UMD) (Carroll et al. 2009)	African Continent	Open water bodies from the SWBD (Slater et al. 2006) were combined with MODIS remote sensing data to fill its gaps to generate this improved global water mask (Carroll et al. 2009).	250 m	SWBD first validated by Rodrigez et al. 2006. Carroll et al. compared UMD with NLCD (Homer et al. 2001) over North America.
Central Africa Regional Program for the Environment (CARPE) – Congo Wetland Map (Bwangoy et al. 2010)	Central Congo Basin	General wetland definition, to distinguish from non-wetland upland forests. Map produced from thresholding of wetland probabilities from remote sensing and topographic indices. The threshold value was chosen to recreate wetland coverage found in Bwangoy et al's (2010) Congo training sites.	57 m	Validated by Bwangoy et al. with 6361 in situ reference points, and compared against other wetland maps of the region.

4.3 Data Processing

4.3.1 Hydrographic and Topography Variables

The use of exclusively topographic variables assumes that topography alone can be a good predictor of the presence of surface water without having to physically simulate hydrological processes. In total, 13 hydro-geomorphic variables (Table 4.2) carefully chosen for their relevance to inundation dynamics were calculated for every pixel of the African continent from the layers of HydroSHEDS. The predictor variables needed to be transferrable across regions, therefore, variables described as absolute values like latitude, longitude, slope orientation, river orientation and elevation cannot be used as predictors.

Table 4.2: Description of the 13 hydrographic and topographic variables computed from HydroSHEDS.

Hydrographic and topographic variables	Description
Slope	Slope of each pixel in degrees of inclination.
Distance to river network (x6)	The longitudinal distance of each pixel to the river networks following the flow path. River networks of different densities were produced with 500, 1000, 5000, 10 000, 25 000 and 50 000 up-cells.
Elevation above river network (x6)	The relative elevation of each pixel above the river pixel to which it flows. River networks of different densities were produced with 500, 1000, 5000, 10 000, 25 000 and 50 000 up-cells.

The first variable included was the slope of each pixel expressed as its inclination in degrees. Slope is relevant to inundation patterns because low inclination areas are more prone to water accumulation than steeper ones which can drain it more easily. Then, two variables, relative distance to river network and relative elevation above river network were repeatedly calculated for six river networks of different densities thus generating the other 12 variables. The six river networks of varying densities were produced with six thresholds of number of contributing cells: 500, 1000, 5000, 10000, 25000, and 50000. Higher threshold values signify that river reaches representing larger rivers with larger contributing areas are kept. Contributing area is a consistent manner of defining of river and,

provides a continuous, hence more detailed measure of size in comparison to stream order for instance. Bwangoy et al. (2010) argued that selection of an optimal contributing area threshold was poorly defined (Tarboton et al., 1992) and as a result used topographically-derived variables at different contributing areas. In their study, variables generated from river networks of varying densities were shown to generate complementary information for inundation mapping rather than duplicating information.

One of the repeated variables that are generated is the elevation of land pixels relative to the river network following the steepest path. The relative elevation to the river network is inspired from Bwangoy et al. (2010) but differs in the way relative elevation is measured. Bwangoy et al. (2010) calculated the elevation difference to the closest river with no regard to whether a cell flowed to that particular river reach or not, hence misrepresenting cases where the river reach closest to a pixel is not the one toward which it flows, and possibly introducing incorrect elevation difference values. This problem is corrected in the present study by considering flow directions in defining the river reach to which each pixel flows. The other repeated variable computed was the distance to the river network. The variable representing the distance from the river network has not previously been used in the literature to characterize inundation patterns, but it has an inherent relevance to floodplain setting where the main source of inundation comes from channel overbank flow. Together, these 13 variables provide clear insight into the fine-to-broad scale geomorphic setting of each cell. Second order variables such as curvature, convexity, and slope orientation were excluded from the set of variables, as they were shown to be of little explanatory power for wetland location by Bwangoy et al. (2010). For land cells flowing to the ocean or to inland sinks, the distance and relative elevation to the inland sink or ocean are used. These instances are more numerous for lower river network densities than for higher density ones.

4.3.2 Reference Data Processing

The sources of reference data need underwent some processing in order to be transformed into useable training and validation subsets (Figure 4.6).

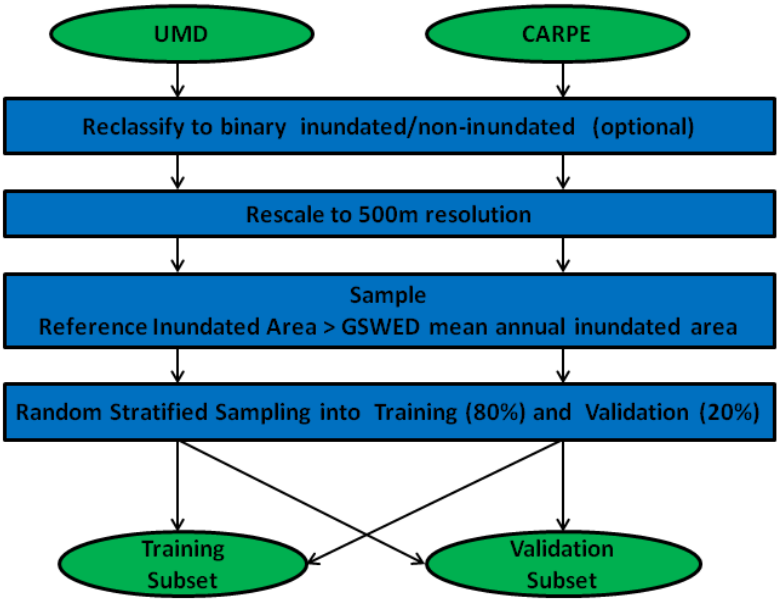


Figure 4.6: Flow chart of reference data processing from raw reference data sources to training and validation subsets.

Rescaling of Reference Data

Following the selection of data sources, reference datasets of two different spatial resolutions (57 m, 250 m) were upscaled to a standard 500 m raster format while preserving the total inundated/wetland area as in the original dataset (Figure 4.7). The number of inundated pixels of the original resolution was summed for each 500 m resolution pixel. The 500 m pixels bearing the number of originally inundated pixels are thresholded to recreate the total area covered by the original dataset. The example depicted in Figure 4.7 presents a perfect case where original and upscaled inundation maps display the exact same inundated area, but this is not always the case and sometimes small differences are introduced in the process.

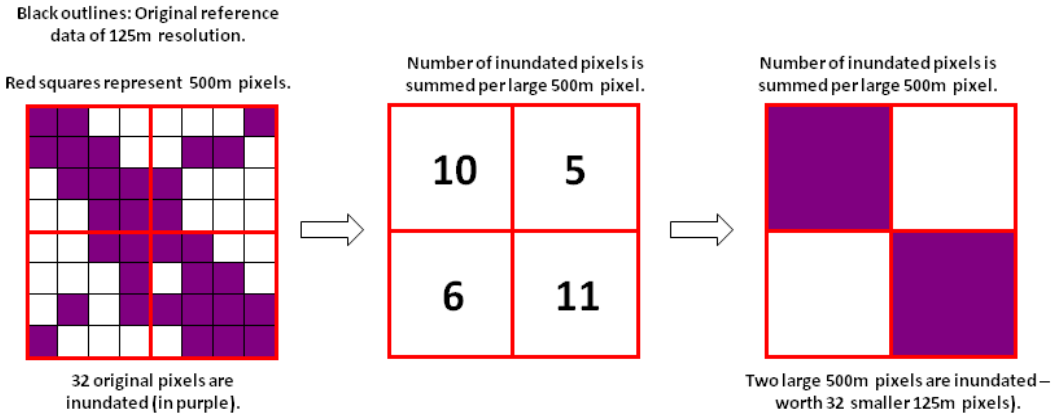


Figure 4.7: Diagram of an upscaling procedure for reference datasets preserving the total inundated area. This diagram uses hypothetical spatial resolution for the purpose of the example.

Reference Set Sampling

Within the extent covered by the rescaled reference data sources exist large areas with no inundation present. The inclusion of areas with unrepresentative level or absence of inundation would not improve the training data. To preserve the representativeness of the reference data and minimize the size of model input data, only particular areas truly useful for the predictive model within the coverage of the reference dataset and meeting certain criteria were selected. The selection of reference data was based on GSWED 0.25° cells, making the cells the basic unit of sampling. The selection of reference sample cells relies on the comparison of the inundated area from the reference data and GSWED. Only cells for which the reference inundated area was above the mean annual minimum of GSWED were selected. The decision to use GSWED cells as sampling units is supported by the fact that the downscaling and even validation are ultimately done over them. It is therefore crucial that training and validation areas be selected with the same units as for the downscaling to adequately capture the accuracy of the redistribution. Using smaller sampling units than the GSWED cells for selecting training/validation subsets would yield artificially high accuracy from the validation (because it is easier to correctly distribute inundation within a smaller area), whereas larger ones would lead to artificially low accuracy.

This selection has the objective of excluding samples where reference data are in contradiction with the GSWED estimates, and could misguide the predictive model. The result of this selection only preserved 790 reference sample units/cells for which the inundation extent is within the range of values of GSWED (502 from UMD, 288from CARPE). In selecting sample units, no consideration was given to the spatial location of each sample unit. Because one of the two reference sources included into the reference data group is limited to the Congo basin, that region is over-represented by the sampled units set, accounting for more than half the cells sampled, as illustrated by the cluster of sample units showed in Figure 4.4. The reference sample selection process is carried out for each individual reference data source, allowing for multiple data sources to be selected over the same sampling unit and to provide multiple snapshots of inundation at different water levels. In Africa, an overlap exists between the two reference sources of Table 4.2 over the Congo basin. In the current implementation of the methodology, no coastline cells were included in the selection process because they had been filtered in the first GSWED version used at the time of the sampling procedure. In the end, 502 cells were selected from the data source UMD which originally covered the entire continent and 288 were selected from the 1440 cells originally covered by the CARPE data source, totalling 790 selected cells altogether (Table 4.8).

Training & Validation Subsets

Selected reference sampled units were divided into training and validation subsets with a random stratified sampling procedure that attributed 80% of each data source's cells to training, and the remaining 20% to validation. The stratified sampling procedure preserved a proportional representation of each data source in the training and the validation subsets to maintain representativeness in both subsets. In total, 633 and 157 cells were respectively selected for training and validation (i.e., 1.60% and 0.40% of the total number of cells for Africa). The distribution of data source cell frequency among 10% inundation fraction ranges

(Table 4.3) shows that the representation of low inundation fraction is dominated by UMD while high inundation fraction is solely dependent on CARPE.

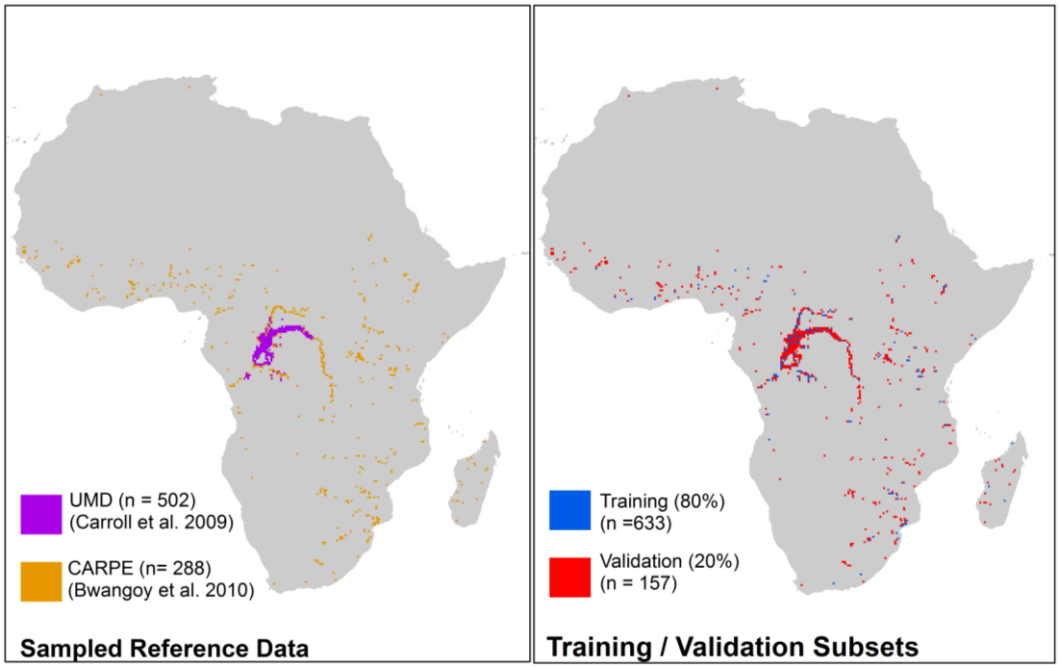


Figure 4.8: Data source and Training/Validation location of sampled cells over the African continent.

Table 4.3: Proportion of reference sources by 10% inundation fraction range for training and validation subsets

	Training (80%)			Validation (20%)		
Inundated fraction	**Total**	**nUMD**	**nCARPE**	**Total**	**nUMD**	**nCARPE**
0-10%	339	318	21	86	79	7
11-20%	67	47	20	17	13	4
21-30%	29	16	13	7	3	4
31-40%	19	6	13	6	1	5
41-50%	18	4	14	5	3	2
51-60%	15	4	11	5	1	4
61-70%	11	2	9	1	0	1
71-80%	26	3	23	5	0	5
81-90%	42	1	41	8	0	8
91-100%	67	1	66	17	0	17
Total	633	402	231	157	100	57

The sampled reference datasets are assessed for any bias they might have and that could affect the training and validation process. Possible bias of the reference sample can come from major discrepancies between the samples and the rest of the African continent. Misrepresentation of the African continent's wetlands by the training and validation set can be evaluated by looking at the relative abundance of each wetland type of GLWD within the cells selected for training and validation GLWD (Table 4.4). Although the GLWD is known to be inconsistent, a first order analysis of the wetland types provides insight into possible weaknesses of the sampled reference data. As expected, because of the characteristics of the reference data sources as well as the sampling procedure, classes such as Reservoirs, Rivers and Swamp/Flooded Forest are over-represented by training and validation samples, while Floodplains and Intermittent Lake/Wetland are markedly underrepresented. As mentioned, Reservoir and Rivers are overrepresented because the majority of the sampled reference data represents permanent water bodies. Underrepresentation of Intermittent Lake/Wetland class is caused by the absence of that class in the Congo basin, outside of which the other reference dataset only captures permanent waterbodies. Most of the occurrences of Intermittent Lake/Wetland class are located in the Deception Valley of Botswana and in the Sahara. The same can be said of floodplains, which are chiefly located in the Okavango Delta, Barotse Plain, Sudd Marshes, Logone and Hadejia Jama'are Floodplain as well as in the Niger Internal Delta and Senegal River Delta. The under-representation of Lake, Coastal Wetland and Brackish/Saline wetland classes is understandable when considering that the sampling rule excluded all coastal cells (along ocean or inland lake) where these classes are located. Overall, the sampled reference data can be considered to possess a bias favouring rivers and permanent water bodies. Table 4.5 confirms that training & validation subsets are equivalent among themselves in many aspects.

Table 4.4: Comparison of sampled training & validation sets to the African Continent for GSWED and GLWD

Variables	Training Set	Validation Set	African Continent
Number of Cells (0.25º)	633	175	39544
Reference Data (CARPE & UMD)			
Average Inund. Fraction:	27.90%	26.36%	-
Maximum Inund. Fraction:	100.00%	100.00%	
Minimum Inund. Fraction:	0.05%	0.08%	
GSWED MAMax			
Average Inund. Fraction:	14.70%	13.72%	3.93%
Maximum Inund. Fraction:	96.98%	90.42%	100.00%
Minimum Inund. Fraction:	0.01%	0.02%	0.00%
GLWD classes			
Lake			
Reservoir			
River	7.17%	8.60%	14.05%
Floodplain	5.54%	4.05%	2.44%
Swamp/Flooded Forest	10.02%	12.63%	2.86%
Coastal Wetland	21.56%	21.51%	43.99%
Brackish/Saline Wetland	50.99%	52.95%	11.11%
Bog, Fen, Mire	0.84%	0.22%	4.68%
Intermittent	3.71%	0.02%	7.72%
Lake/Wetland	0.00%	0.00%	0.00%
50-100% Wetland	0.18%	0.02%	13.16%
25-50% Wetland	0.00%	0.00%	0.00%
Wetland Complex (0-25%)	0.00%	0.00%	0.00%

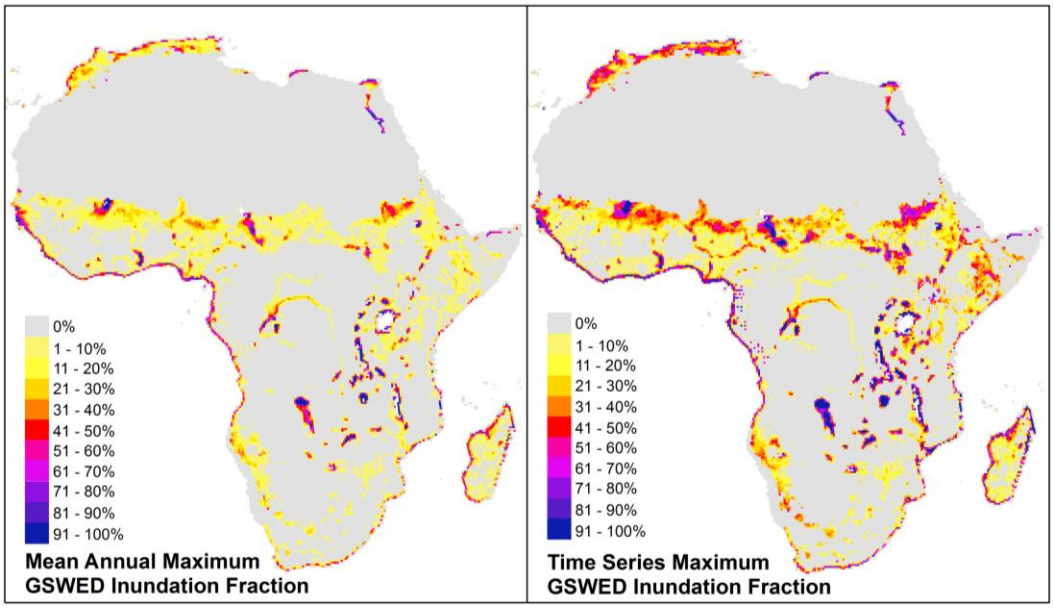


Figure 4.9: GSWED Mean Annual Maximum Inundation Extent over the continent of Africa from GSWED (Papa et al. 2010)

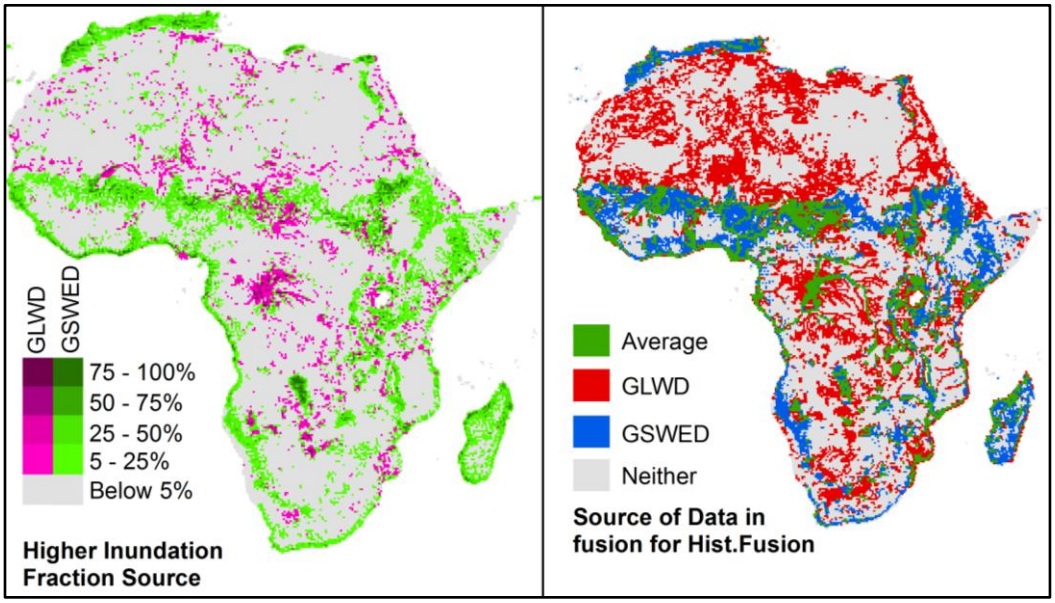


Figure 4.10: Comparison of data sources for fusion, in terms of inundation fraction and source used in the fusion

4.3.3 GSWED Temporal Aggregation

With so many temporal gaps present in the GSWED data, inundation at the monthly time step remains very incomplete for many regions. To overcome these gaps, the monthly estimates of the time-series were aggregated to create the Mean Annual Maximum (MAMax) and Time Series Maximum (TSMax) estimates for each cell (Figure 4.7). The MAMax temporal aggregation creates an averaged composite coverage of yearly maximum inundated area unaffected by extreme outlier estimates. The TSMax is generated from the highest value found among the estimates of the time series for each cell to illustrate the maximum inundation observed over 12 years, but is vulnerable to outliers and overestimation. Following temporal aggregation, the number of HydroSHEDS pixels in each cell is computed and the inundated fraction is converted into an inundated number of pixels. There is on average 3600 pixels per cell for the African continent.

4.3.4 Fusion of GLWD and GSWED

To further address the shortcoming of GSWED, i.e., complement and fill some of its gaps, it was combined with the wetland extent of the GLWD (previously discussed in section 2.2.3). Because the GLWD represents the maximum extent of wetland, the fusion can only illustrate the maximum extent of wetlands and inundation which is particularly important because it provides an ecological baseline against which current degraded wetland extents can be evaluated. To achieve the maximum inundation extent possible, the GLWD was merged with the time series maximum (TSmax) of GSWED. The fusion consisted in converting the areal coverage from all the classes of GLWD into a fraction of the 0.25° cells, and then combining it with TSmax using simple decision-rules to the GSWED time series maxima. For cells where there is a nonzero inundation area for either GLWD or TSmax, the nonzero value is used. In cases, where both sources possess a nonzero value, the average of the two is used. The choice of averaging the two sources where possible, instead of using the maximum value for instance, comes from a wish to generate a conservative wetland area estimate as a starting point for further work on the fusion process. Moreover, preliminary testing of a maximum based fusion yielded unreasonably high inundation estimates. The difference of inundation fraction between the two data sources and the source of data used in the fusion is shown in Figure 4.10. Because of its more distributed inundated area, GLWD was used in ~30% of the cells of the continent, twice as much as those based on GSWED estimates or an average. The fusion of the GSWED with the GLWD generated a maximum inundation extent (MaxFusion) larger and spatially more distributed coverage than the GLWD or the GSWED alone (Figure 4.11).

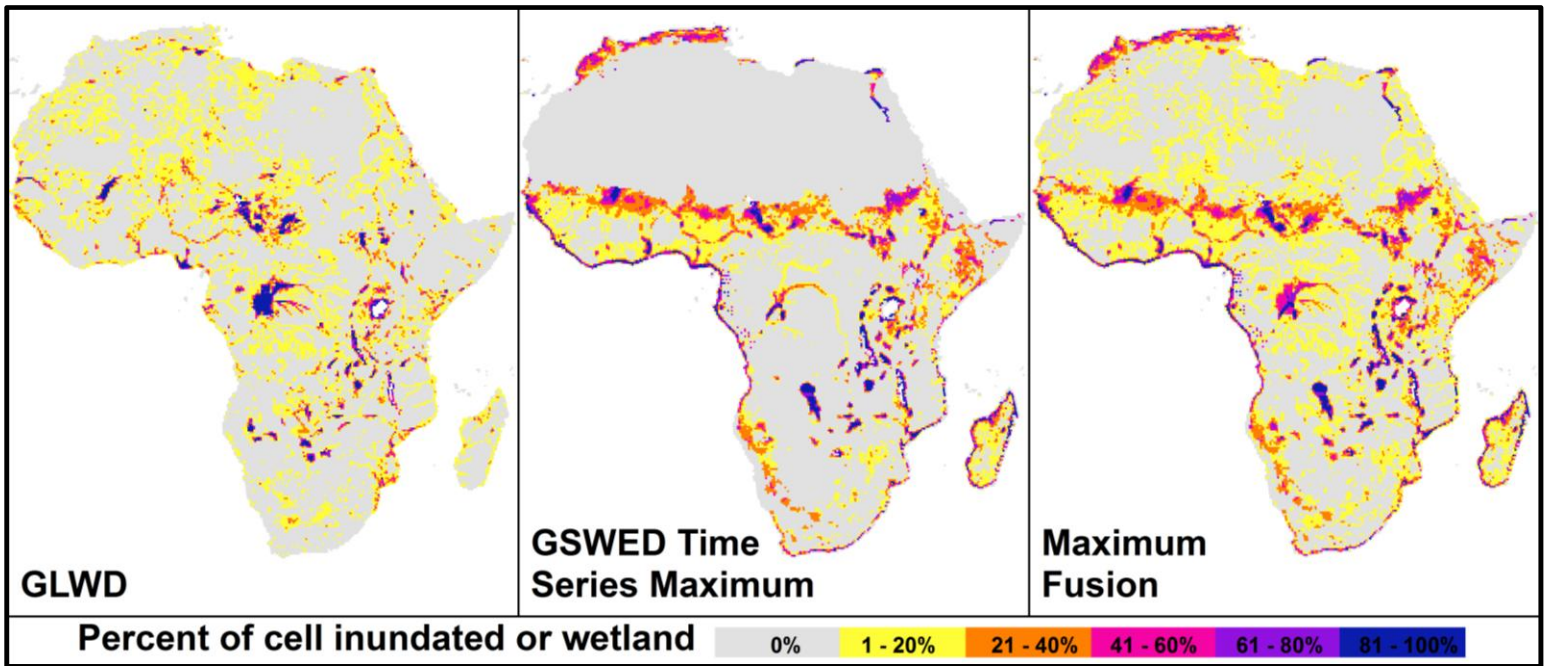


Figure 4.11: Per-cell inundated fraction of the two fusion inputs: GLWD and GSWED TSmax, and the resulting output map MaxFusion. The output MaxFusion map inherits the wider coverage from GLWD and the inundation patterns of the Sahel from GSWED.

4.4 Downscaling Method

The downscaling methodology refers to the conversion of inundation extent from the coarse resolution cells of GSWED to the finer pixel density of HydroSHEDS. The methodology consists in the succession of the three following steps: inundation probability induction with decision trees, distribution of inundated area among pixels, and majority filter applied onto downscaled inundation map.

4.4.1 Decision Tree Training

A decision tree learner (Breiman et al. 1984) was used as a predictive model of per-pixel inundation likeliness expressed as inundation class membership probability. The decision tree based its decisions on hydrographic and topographic variables and was trained on the sampled reference data. The use of decision trees is inspired by the application by Bwangoy et al. (2010) to produce wetland class membership probability over the Congo basin. Because decision trees work well with large datasets and large number of variables with little parameterization, their use for machine learning and data mining has been widespread (Provost & Domingos 2003). Moreover, decision trees are flexible in the type of data they can integrate (continuous, discrete, categories), their processes are transparent and retraceable (i.e., white box model), and they are appropriate to capture non-linearity and interaction among predictor variables (Elith et al. 2008).

The use of decision tree classifiers has been commonplace in remote sensing land-cover mapping (Hansen et al., 1996; Michaelson et al., 1994) even at the global scale (DeFries et al. 2000; Hansen et al. 2000; Friedl et al. 2002) and was also proven adequate specifically for wetland and riparian area mapping specifically (Hess et al. 1995; Simard et al. 2002; Baker et al. 2006; Ordoyne & Friedl 2008; Whitcomb et al. 2009; Bwangoy et al., 2010). Decision trees use a set of hierarchical univariate rules that successively split input data into two output subsets purer than the input group - known as nodes (Breiman et al. 1984).

Decisions at each split are chosen by the tree induction algorithm to maximize information gain, or Gini index (Foster & Domingo 2003).

Data in each node is further split, and tree induction is maintained, until a stopping condition is reached and induction is halted for each node (they become end nodes). This process of successive decisions splits data into end nodes which can then be classified, or as done by Bwangoy et al. (2010), used to calculate class membership probabilities. Probabilities are useful in situation where cost/benefit analysis is necessary and where certainty of a decision is necessary, such as for medicinal applications like cancer detection (Cieslak & Chalwa 2008). Class membership probabilities are calculated for each end node of a tree from the training data as the fraction of a class within all instances found in that end node (Figure 4.12). By splitting the input data into increasingly distinct subsets of pixels, the decision trees partition data into relatively homogenous groups to which probabilities can be attached. In the present application, sampled training pixels are split by the decision tree, and the inundation class membership probability of the end nodes is calculated from the number of inundated pixels in the training data. However, conventional decision trees algorithms intrinsically focus on maximizing classification accuracy and minimizing the size of the tree, which can lead to poor probability estimates in some cases (Provost & Domingo 2003). Large trees with even superfluous numbers of leaf nodes are more appropriate for probability estimation than smaller ones (Provost & Domingos 2003). There is also a risk of over-fitting the data in large trees, which can lead to extreme probability estimates due to small sample sizes of the end nodes (Provost & Domingos 2003). Some corrections, such as "smoothing" and "Laplace transform" can account for those biases from small sample sizes (Provost & Domingos, 2003).

In this project's use of decision trees, inundation class probabilities are used to divide pixels into inundated/non-inundated pixels with different exceedance threshold values to recreate the various levels of inundation. To properly recreate inundation estimates of GSWED, the estimated probabilities should be well varied

and evenly distributed over their range of values. By being evenly distributed (i.e., an even concentration of probabilities along their possible range), the probabilities can properly recreate all of the different inundated area estimates among the pixels, while a large variety of probability values makes it possible to closely replicate the inundated area. These particular requirements from the estimated probabilities for this application are not conventional for decision tree induction. Equal distribution and large variety of probability values are not contradictory and both can be reached with careful choice of termination condition which control the number of end nodes. To ensure production of a robust probability scheme, bagging of the training set was employed, since it is known to improve probability induction from decision trees (Provost & Domingos, 2003). Bagging, or bootstrap aggregation, is a method of generating and aggregating multiple predictors from bootstrap replicates of a learning set (Breiman 1996). In the context of decision tree induction here, and similar to Bwangoy et al. (2010), bagging involved the induction of multiple trees independently from randomly selected subsets, and aggregated by averaging the output inundation class membership probabilities for each pixel.

Five trees were grown from five bagged subsets of 50000 pixels, representing a potential maximum of 250,000 pixels (the actual maximum number of pixels sampled can be smaller because of the possibility of pixels being selected twice). Tree growth was terminated at nodes below the arbitrary number of 1000 pixels, possibly yielding a maximum of 50 end nodes per tree. Choice of the tree growth stopping condition was chosen after trial and error. For the bagged subset selection, the 2,316,120 pixels within the 633 training cells were grouped together in order for a stratified random sampling to be selecting individual pixels and not cells. As a result of the sampling technique, the subsets were assembled from pixels located in any cells to ensure maximum subset heterogeneity. The maximum sampled 250,000 pixels represent 10.79% of the total training pixels and 0.17% of the total 146,893,243 pixels of the continent. The pixel probabilities of the five trees were averaged to produce a total of 44251 different

probabilities. The size and number of trees of the decision trees were defined after a limited sensitivity analysis showed that larger and more numerous subsets did not significantly improve accuracy (see recommended improvements, section 6.3.4). In this project`s thresholding of pixel probabilities to turn them into inundated/non-inundated classes, only the ranking – the relative size in comparison to others – and not the absolute probability of each pixel influences the distribution of the inundated area. As explained in the next section, the distribution of inundation among the pixels depends on their rank among the other pixels located in a given cell. Favorable to this practice, it has been shown that surprisingly good rankings can be generated from questionable probability estimates (Margineantu and Dietterich 2001), and that probability-bagged can generate excellent rankings (Provost & Domingo 2003).

With the trees populated with the training subset, inundated class membership probabilities can be predicted for all pixels of the rest of the continent, including the training and validation subsets. These pixels without inundated/non-inundated labels are ran through the decision trees defined from the training data based on their topographic and hydrographic variables. Pixels are attributed the probability values of the end nodes to which they belong, and are then averaged to generate the final probability. To estimate probabilities for the pixels of the entire continent, pixels were selected and processed one cell at a time, but merged back together to form a continent wide mosaic. The probabilities of pixels in different regions of the continental mosaic can be compared because they have been generated from the same bagged decision trees.

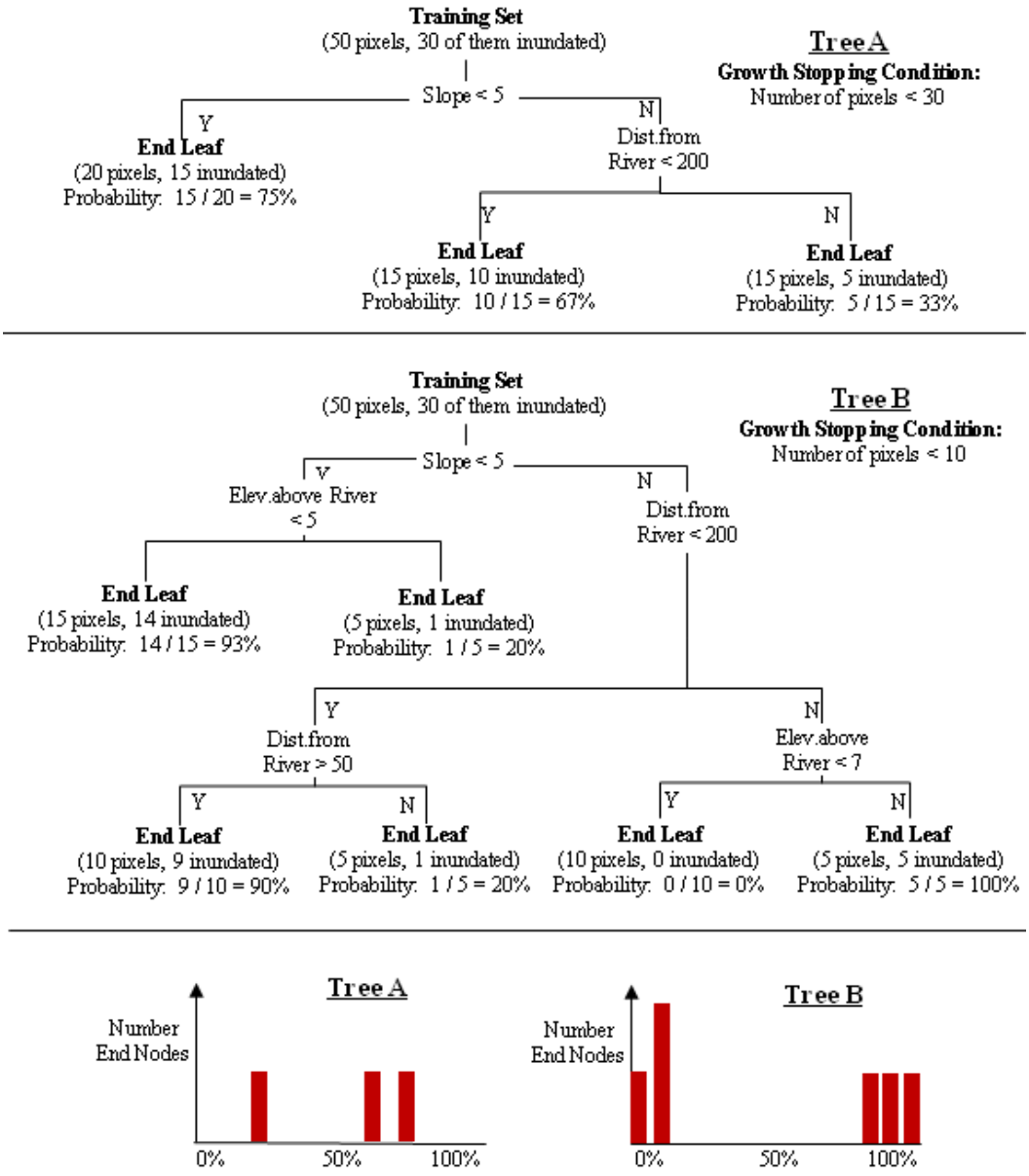


Figure 4.12: Examplar decision trees and probability induction process of two trees illustrating the two possible faults of probability distributions. Tree A is a small tree with a low variety of probability values, while the larger Tree B has higher variety of values but unequal distribution. The wanted probability distribution for the downscaling process must be both equally distributed (Tree A) and possess a large enough variety of values (Tree B), and thus attempts to stike a balance between the two examples here.

4.4.2 Distribution of inundated area

The process of distributing inundated area among pixels consists of thresholding the probability distributions to recreate a given inundated area estimate at the finer resolution. In Bwangoy et al. (2010) the thresholding of probabilities recreated the fraction of wetland observed in training regions to their entire study area whereas in the present thesis, the threshold values are chosen to recreate the inundated area estimate within the outline of the coarse resolution GSWED cells.

Edge Effects

Thresholding the pixels of multiple adjacent cells individually can sometime have the unwanted result of producing noticeable features (i.e., unwanted linear contours) in the resulting inundation map at the border between adjacent cells. These straight lines, referred to as "edge effects" are particularly prevalent along the borders of cells where the magnitude of inundated area differs greatly between these adjacent cells - i.e., a very inundated cell next to a very dry cell (Figure 4.13). To prevent such "edge effects" from occurring, allotment of inundated area among pixels was performed with the Moving Window Thresholding method (MWT). The MWT method is a novel concept, designed as part of this thesis with the specific goal to eliminate edge effects from the downscaled map while remaining as close as possible to the original estimates. As a quick alternative to longer processing time of MWT, the Single Value Threshold method (SVT) was tested alongside the MWT.

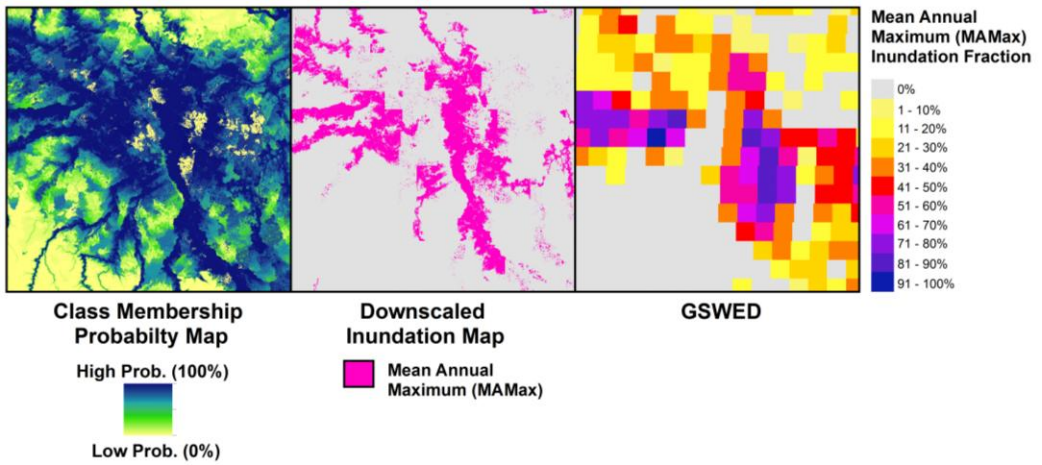


Figure 4.13 : Map of Sudd Marshes displaying the "edge effects" and the original GSWED estimates. The straight features absent from the probability map appear from the downscaling of inundation estimates. The features appear along the edge of cells with contrasting high/low inundation fractions.

Moving Window Thresholding (MWT)

The Moving Window Thresholding (MWT) method applies the thresholding concept to a window on multiple cells at a time. A moving window is constituted of a group of cells, in this case eight, centered around a central processing cell (nine cells altogether). As processing moves from one central processing cell to the next, so does the moving window. This method attempts to eliminate edge effects by reallocating inundated area within the entire window's extent while disregarding borders between cells of the window. In the MWT procedure, the thresholding of probabilities is executed for the moving window as a whole. As a result, the process allocates inundated area of the whole window (Figure 4.5). Following the thresholding of the entire area of a window, only the central cell of the window is extracted and the process is then repeated for the next central cell, producing an inundated area spatial distribution that is smoothed across the landscape. As a means to control the quantity of reallocation among the cells that defines the degree to which the original per-cell estimate is replicated in the downscaling movement of inundated area among cells, a Residence Ratio (RR) was introduced in the MWT. The RR is a measure of the proportion of the estimated inundated area from the window's center cell that should remain within that cell.

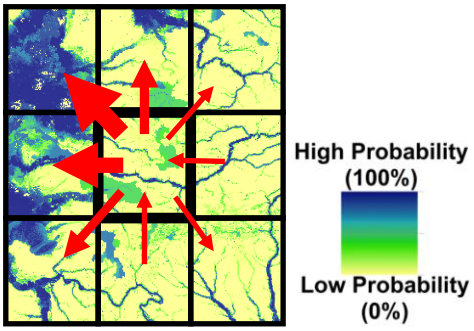


Figure 4.14: Conceptual example of the Moving Window Thresholding method (MWT); The black outlines represent a window of 9 cells. The red arrows going from the center cell to the outer cells represent the direction and magnitude of the exchange with the center cell. The reallocation of inundation generally goes from cells with lower probability to cells with higher probability. Note that the arrows show the possible reallocations between these cells, but the actual reallocation depends on the inundation area estimates for each cell.

With the inclusion of RR, the MWT process is carried out in multiple steps. First, based on the inundated area of each cell and the RR, the number of "residence inundated pixels" is defined and thresholded within each cell. Second, the overall number of inundated pixels and the "residence inundated pixels" are summed for the entire window. The difference between the two represents the number of "mobile inundated pixels" which can be reallocated anywhere within the window. Then, the remaining non-inundated pixels of the window are thresholded at once to replicate the number of "mobile inundated pixels" previously determined. Pixels of the central cell are extracted and collated to the center cells from other windows. The result is no overall change in total inundated area, but only a distribution across a larger area. Changes in total inundated area from this method are not directly controlled, but instead depend on the assumption that overestimations and underestimations of inundated area across the landscapes should even themselves out over large areas. Because of this assumption, slight changes in total inundated area across the continent are expected.

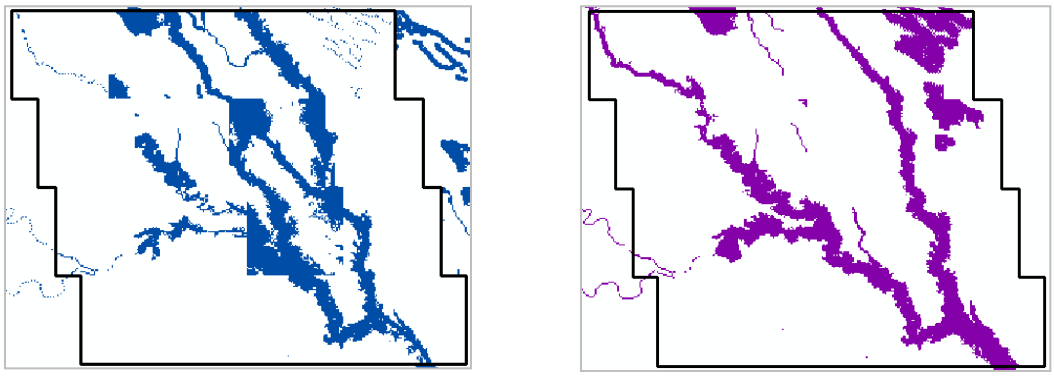


Figure 4.15: Downscaled inundation map from the moving window thresholding (MWT) with Resident Ratio values of 1.00 (left) and of 0.00 (right). The right panel displays little or no "edge effect" compared to the one on the left, from the reallocation of inundated area that strongly alters the overall pattern of inundation.

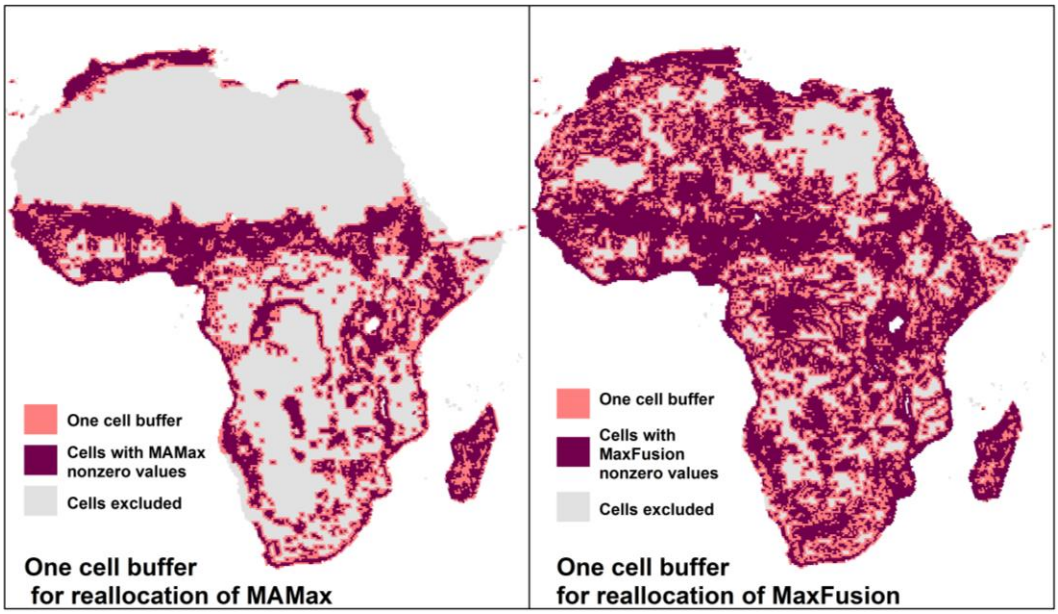


Figure 4.16: Thresholded area of the continent Area used for the displacement analysis for MAMax (left) and MaxFusion (right)

The MWT process does not affect the downscaling within each single cell, because it should instead be thought of as an alteration to the inundated area estimate. The modification of the original estimates to eliminate edge effects is defended with the argument that the original estimates possess sizeable uncertainties due to their geolocation, the accuracy of the estimates and the spillover effect among cells. The reallocation process can be considered as correcting for these uncertainties by reallocating the estimates among the cells on the basis of topography. Catherine Prigent (personal communication, December 2011) agreed that 15-20% of the inundation could be distributed to neighboring cells to adjust for edge effects (slightly higher than the 10% error reported by Papa et al. 2010). The current iteration of the moving window process uses a RR of 20%, meaning that 80% of the inundated area reported by each cell can be reallocated elsewhere, much higher than prescribed by C. Prigent. The moving window radius was of 0.25 decimal degrees, approximately a one cell buffer, which averages 9 cells per window (including center cell). The high reallocation rate used is justified by previous testing with lower values which resulted in unacceptable numbers of edge effect instances.

Single Value Threshold (SVT) Method

With the Single Value Threshold method, a single probability cutoff that replicates the inundated area summed for the entire continent is selected and applied on the entire continent. With the SVT, redistribution of the inundated area has no barriers and is free to be reallocated anywhere within the area where the cutoff is used. The area defined for the SVT was the same processed by the MWT (with a radius of 0.22°), for comparison purposes. The area consists of a one cell buffer around the nonzero estimate cell, which is the area effectively mapped by the MWT. For MAMax, a buffer consisting of 17778 cells surrounding the 9946 cells with nonzero inundation estimates was used, while analysis of MaxFusion consisted of a buffer of 33156 cells surrounding the 20,050 cells with MaxFusion estimates (Figure 4.16). The result is an inundation map equivalent in total area as the MWT output, but not forced locally by the

inundation estimate by the inundated fraction of individual cells. The SVT requires much less computation power than the MWT because it does not need to process each cell individually.

4.4.3 Majority Filter

The current implementation of the decision trees generates a probability distribution containing some outlier values affecting the distribution of inundated area. To counteract these outlier values, or noise, present in the probability, a low-pass 3 × 3 majority filter was applied onto the downscaled inundation maps, giving the filters' central pixel the value of the majority of pixels of the filter. The majority filter has the effect of reducing the number of isolated pixels and simplifying edges of inundation features, overall making the map more esthetic. By modifying the spatial distribution of inundation, the filter has the potential to also modify the total inundated area. Application of the filter is assessed alongside unfiltered downscaled inundation maps for comparison.

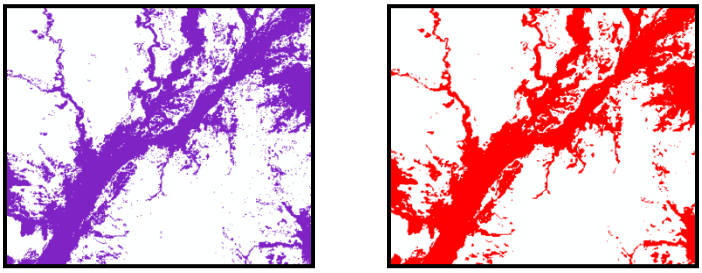


Figure 4.17: Representation of the main stem of the Congo River before (left) and after (right) the application of the majority filter

4.5 Product Evaluation

Evaluation of the downscaled inundation maps is twofold to assess the two sources of error of the output: a) assessing the accuracy of the downscaling procedure and, b) assessing the errors in the original and downscaled inundated area.

4.5.1 Downscaling Accuracy

The validation of the downscaling procedure was achieved by comparing downscaling outputs with the reference data over the 157 validation subset cells. To assess the spatial distribution alone without the effect of different class abundance, comparisons must be made between two equal inundated areas. This has the effect of only assessing allocation disagreement and excluded quantity disagreement. Quantity disagreement introduced by the thresholding process is assumed to be negligible. To compare equivalent areas, the inundated area from the reference validation set was aggregated for each cell and then downscaled by the thresholding process. The thresholding process was applied onto single cells and not over a moving window, because the validation should reflect the ability of the algorithm to redistribute an exact inundated area correctly within a single cell's space. The same procedure was also applied over the training set in order to obtain the training accuracy.

Different accuracy metric extracted from confusion matrixes were computed to quantify the performance of the downscaling procedure. The training and validation sets were divided into 10% inundation fraction ranges, each assessed independently. Three accuracy metrics were utilized: Producer's Accuracy, Overall Accuracy and Kappa Index of Agreement. The use of the Producer's Accuracy, i.e., the level of agreement as the percentage of correctly classified inundated pixels, is appropriate without the User's Accuracy because of the equal class proportion in the downscaled and reference maps. Because of the equal number of omission (false negative) and commission errors (false positive), the user's and producer's accuracy are equivalent for a same class, assuming that there is no difference in class proportion. Producer's Accuracy was calculated only for the inundation class as the other class found in the confusion matrix simply represents the absence of inundation.

Equation 1 $Producer's\ Accuracy = \frac{TP}{TP+FN}$

Equation 2 $User'sAccuracy = \frac{TP}{TP+FP}$

Where True Positive: Number of correctly classified inundated pixels
False Negative: Number of incorrectly classified non-inundated pixels
False Positive: Number of incorrectly classified inundated pixels

The producer's accuracy, however, is impacted by the inundation class fraction within a validation cell. To assess the quality of the map considering both classes, the Overall Accuracy'metric was used. The Overall Accuracy is calculated as the number of correctly classified pixels for both classes and will provide a better accuracy assessment across the varying levels of inundation assessed.

Equation 3 $Overall\ Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$

Where TP: True Positive: Number of correctly classified inundated pixels
TN: True Negative: Number of correctly classified non-inundated pixels.
FN: False Negative: Number of incorrectly classified non-inundated pixels
FP: False Positive: Number of incorrectly classified inundated pixels

The Kappa Index of Agreement (KIA) represents the percentage of accuracy improvement over the level of agreement that would be reached by chance alone (Cohen, 1968). Although now standard component of accuracy assessment in remote sensing, KIA has been criticised for its combined assessment of quantity and allocation errors into a single metric (Pontius & Millcones 2010). In this present case however, because quantity disagreement has been excluded from the validation, the KIA, like the other metrics, only capture the accuracy of spatial distribution within each cell. Moreover, KIA can be biased in areas with uneven class distribution, for example very low and high inundation fractions (Foody 2004).

Equation 4 $Kappa\ (KIA) = \frac{Overall\ Accuracy - Random\ Accuracy}{1 - Random\ Accuracy}$

Equation 5 $Random\ Accuracy = \frac{(TN+FP)\times(TN+FN)+(FN+TP)\times(FP+TP)}{Total \times Total}$

See above for symbol meanings and Overall Accuracy definition.

Accuracy metrics were calculated for outputs from the MWT and SVT distribution methods, as well as with and without the majority filter, to evaluate their combined effect on downscaling accuracy. Because they are the result from collating the downscaled maps from adjacent cells, edge effects cannot be captured by the accuracy metrics calculated for individual cells. As a result, presence of edge effects has to be evaluated visually.

4.5.2 Inundated Area Estimates

Accuracy of the downscaled inundated area can depend on two sources: the data fusion of the GLWD and GSWED and the reallocation by distribution methods (MWT and SVT). The two sources of error are assessed separately. Prior to downscaling, total inundated area estimates from data fusion must be reliable. During downscaling, inundation estimates must be accurately replicated locally by the downscaled map. Data fusion estimates are assessed by looking at changes in cell distribution among inundation fraction ranges of 10% as well as spatial visual assessment of the regions where changes occurred. This analysis of the data fusion is the only means of evaluating the data fusion method's results in the absence of validation data. Analysis of the fusion should provide information to direct future improvements to the process. The distribution methods were evaluated in terms of the errors regarding the reallocation of inundated area between cells as well as changes in total inundated area. Reallocation of inundation between cells is measured as the coefficient of variation of Root Mean Squared Error (CV-RMSE) calculated as the RMSE per cell, divided by the average inundation, yielding a percentage value change from original average estimates. The change in total inundated area is quantified as percentage change from the original total inundated area.

4.5.3 Regional Case Studies

Evaluating the representation of some of Africa's largest wetlands by the downscaling procedure was accomplished through comparison of the downscaled MaxFusion with the GLWD (Lehner and Döll 2004) and GLC2000 (Mayaux et al.

2004). Only global datasets were considered for the comparison in order to be consistent between sites and replicable globally in the future. This comparison differs from the accuracy assessment in section 4.4.1 because it compares the output map as a final product compounding both sources of error from the downscaling process and the inundated area estimates.

The GLC2000 land cover map of Africa mostly based on SPOT-4 optical sensors in combination with other data sources is divided into 27 land cover classes, among which four depict inland wetlands: swamp forest, swamp bushland or grassland, waterbodies and irrigated cropland (Mayaux et al. 2004). The mangrove class is the sole representative of coastal wetlands in the dataset. The flooded forest and swamp classes were derived from SAR mosaic of JERS and ERS sensors at high and low water seasons (from GRFM), and was classified based on training data from the Congo floodplain by Mayaux et al. (2002). For comparison with the downscaled map, the GLWD and GLC2000 were reclassified to a wetland/non-wetland system. The GLC2000 was used as an input in the GLWD for constituting the grouped class of "bog, fen, mire". The absence of this class over the African continent frees the comparison between the datasets from circular analysis. Five regional case studies depicting temporary wetlands were selected: Okavango Inland Delta, Sudd Marshes, Nile River floodplain, Congo River floodplain and the Zambezi River delta. Case studies were visually chosen and outlined based on similarity of wetland area in GLC2000 and GLWD. For example, regions of the Niger River basin like the Niger inland delta showed insufficient agreement between GLWD and GLC2000 for inclusion into the case studies. Regions with exclusively permanent wetlands such as rivers and lakes were not considered as case studies because of the lack of variety of representation of these settings. The outline of the Congo and Sudd case studies were chosen to replicate those of Bwangoy et al. (2010) and Rebelo et al. (2011), while the outlines of the other three cases studies were arbitrarily defined to include the wetland areas of all data sources. Thus, case studies were not selected for representativeness of wetlands across the continent. Agreement between the

different representations of the wetlands was captured with the Kappa Index Agreement as well as through comparison of the total inundated area.

5 Results & Interpretation

The downscaled inundation maps of the entire extent of the African Continent HydroSHEDS DEM are the main results of this study. These continent wide maps were generated for Mean Annual Maximum (MAMax) and Maximum Fusion (MaxFusion) temporal aggregations. The downscaling procedure of the two temporal aggregations was repeated using the two alternate distribution methods (MWT and SVT) for comparative purposes, cumulating into four continent-wide inundation maps. All four inundation maps were generated from a single probability map (Figure 5.1). The probability map generated from the decision tree probability induction process shows areas that are topographically prone to inundation but not necessarily inundated in reality. The major exception to this is the Congo floodplain which displays surprisingly low probabilities. The processes behind the generated map displayed in Figure 5.1 are analyzed in depth in the sections below, starting with a downscaling accuracy assessment carried out over the training and validation cell subsets as covered in section 5.1. The modifications of the inundation estimates by the data fusion and the downscaling processes are evaluated in section 5.2. A conclusion from the analysis of this section is that the majority filter improves the overall accuracy, and as a result, a majority filter was applied to the downscaled maps of Figure 5.1. Following the assessment of the downscaling accuracy and inundation estimates, the representation of MaxFusion wetland extent was compared to other global datasets over several study sites in section 5.3. The continent downscaled inundation map are finally shown and visually assessed in section 5.4.

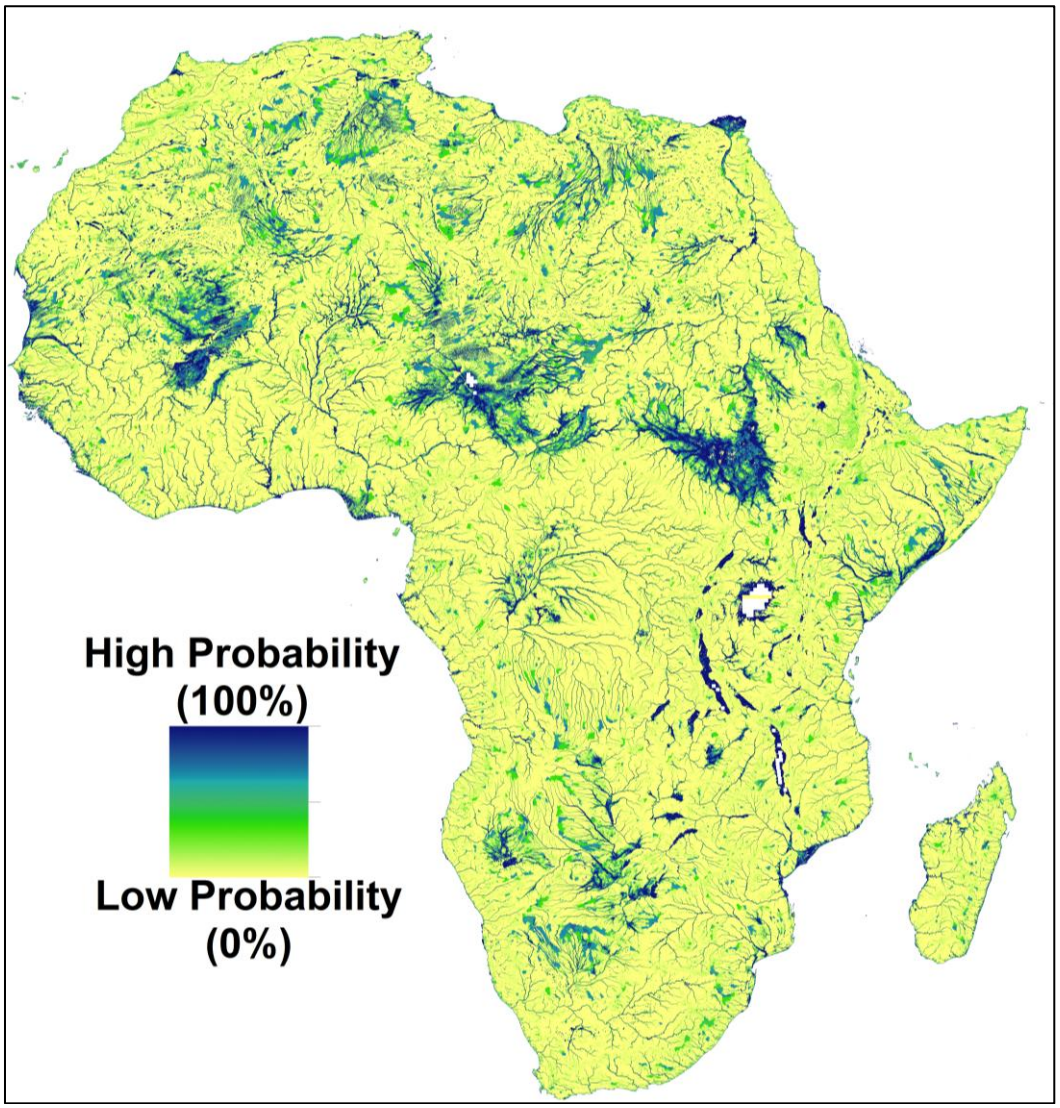


Figure 5.1: Per-pixel probabilities map over the continent of Africa. High probability areas are topographically likely to be inundated, but not necessarily are.

5.1 Downscaling Accuracy

The accuracy of the downscaling procedure is evaluated by comparing the spatial distribution of downscaled inundated extents with the reference distribution of equal inundated area over sampled areas. Producer's Accuracy, Overall Accuracy and KIA metrics are presented for training and validation sets in Tables 5.1 and 5.2 for both, with and without the majority filter. The validation set is completely independent from the probability induction, therefore providing a reliable estimate of the model's accuracy. Accuracy assessment over the training set extends the area of assessment and evaluates whether the decision tree model is over-fitted to the training data. Because only 10.79% of the pixels of the training subset were sampled for training the decision trees, the training set is also largely independent from the decision tree, and can act as a validation.

To assess performance of the downscaling at different inundation levels, the training and validation sets were divided into 10% inundation fraction ranges for each of which accuracy metrics were calculated and are shown in Tables 5.1 and 5.2. The tables show that the ten inundation fraction ranges are unequally represented among the ten ranges, displaying an over-representation of low and high inundation fraction ranges (0 to 10%, 11 to 20%, 81 to 90%, 91 to 100%) while moderate inundation fraction ranges (21 to 80%) are under-represented by both training and validation subsets. The fact that the moderate ranges are under-represented is particularly important for the validation subset because it makes the validation of these ranges less reliable than for other ranges. Downscale products with and without the majority filter (Figure 5.3) are validated and compared to assess the effect of the majority filter on the spatial distribution accuracy.

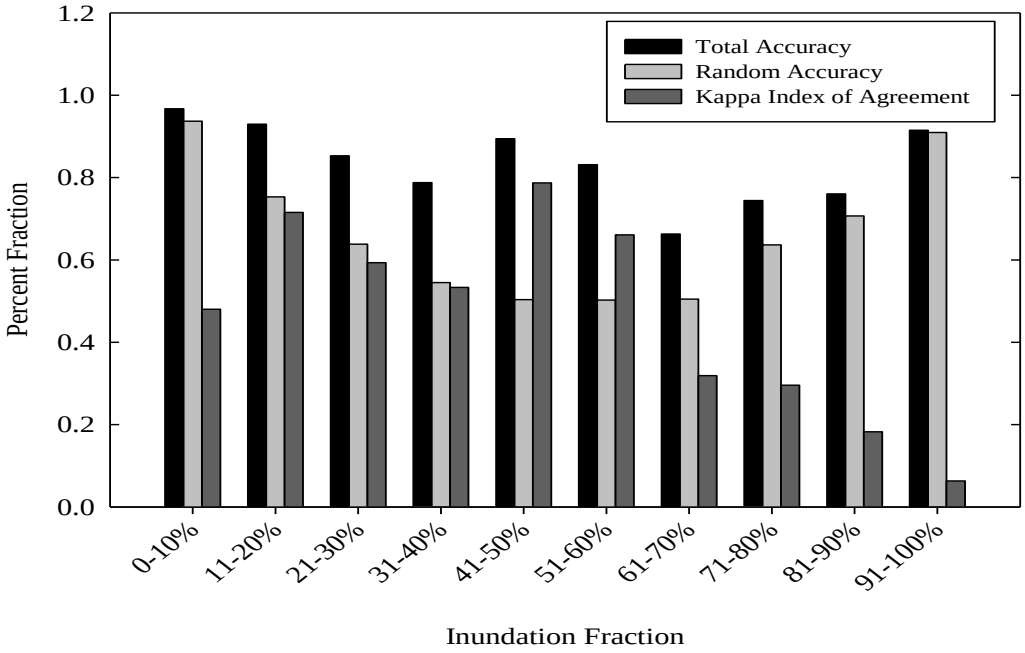
Table 5.1: Accuracy metrics over the validation Subset with and without majority filter

Inundated fraction	Number of Cells (% of total)	Producer's Accuracy		Overall Accuracy		Kappa Index of Agreement	
		Pre-MJ	Post-MJ	Pre-MJ	Post-MJ	Pre-MJ	Post-MJ
1-10%	87 (55.4%)	46.5%	45.7%	96.7%	96.9%	48.0%	49.4%
11-20%	14 (8.9%)	77.3%	77.4%	93.0%	93.2%	71.5%	72.4%
21-30%	7 (4.5%)	69.5%	69.2%	85.3%	85.7%	59.3%	60.1%
31-40%	7 (4.5%)	69.5%	69.7%	78.8%	79.3%	53.3%	54.4%
41-50%	5 (3.2%)	88.6%	88.9%	89.4%	90.1%	78.7%	80.0%
51-60%	6 (3.8%)	83.6%	83.5%	83.1%	83.5%	66.1%	66.9%
61-70%	1 (0.6%)	64.8%	62.8%	66.3%	65.8%	31.9%	31.6%
71-80%	6 (3.8%)	82.7%	83.7%	74.4%	74.8%	29.6%	29.7%
81-90%	7 (4.5%)	84.2%	85.1%	76.0%	76.6%	18.3%	18.2%
91-100%	17 (10.8%)	94.9%	96.0%	91.5%	92.5%	6.3%	5.8%
Total (Avg.: 27.9%)	**157**	**83.7%**	**84.3%**	**92.0%**	**92.3%**	**79.1%**	**80.1%**

Table 5.2: Accuracy metrics over the training set with and without majority filter

Inundated fraction	Number of cells (% of total)	Producer's Accuracy		Overall Accuracy		Kappa Index of Agreement	
		Pre-MJ	Post-MJ	Pre-MJ	Post-MJ	Pre-MJ	Post-MJ
1-10%	339 (53.6%)	44.0%	43.1%	96.7%	97.0%	44.4%	45.9%
11-20%	67 (10.6%)	69.2%	69.0%	91.9%	92.3%	66.2%	67.4%
21-30%	29 (4.6%)	54.6%	53.7%	90.7%	91.0%	49.8%	50.8%
31-40%	19 (3.0%)	77.2%	76.9%	83.4%	83.7%	64.0%	64.5%
41-50%	18 (2.8%)	76.2%	75.9%	78.8%	79.1%	57.1%	57.6%
51-60%	15 (2.4%)	79.7%	80.1%	78.4%	78.9%	56.4%	57.6%
61-70%	11 (1.7%)	80.7%	81.2%	74.7%	74.9%	44.5%	44.8%
71-80%	26 (4.1%)	85.4%	85.9%	77.2%	78.2%	40.6%	40.3%
81-90%	41 (6.5%)	87.9%	89.2%	79.0%	79.6%	19.7%	19.1%
91-100%	67 (10.6%)	96.5%	93.5%	92.8%	93.5%	8.3%	7.3%
Total (Avg.: 26.36%)	**632**	**77.8%**	**78.1%**	**91.7%**	**92.0%**	**69.7%**	**71.0%**

Figure 5.2: Histogram of accuracy metrics for inundation fraction 10% ranges



The three accuracy metrics vary differently for inundation fraction ranges (Figure 5.2). The same behaviours are exhibited for both subsets as well as with and without the majority filter. Producer's Accuracy, the fraction of correctly classified inundation pixels, steadily increases with inundated fraction. Overall accuracy which considers the correct classification rate of both inundated and non-inundated classes is highest for very uneven ranges, where opportunities for misclassification are fewer. The overall accuracy slightly decreases for more even ranges (41-50%, 51-60%) where confusion is likelier. When comparing extreme symmetrical ranges (1-10% and 91-100%), low fractions appear to be slightly better. This divergence in accuracy for symmetrical class distributions indicates that the model is better at correctly mapping inundations than dry land, which is understandable considering that the model is designed to redistribute inundation, and that the "dry land" in fact represents the absence of inundation. KIA also possesses an increasing trend until the 41-50% fraction range, where it starts decreasing to reach very low values for higher inundated fraction ranges. For

very low and very high inundation fraction ranges (0 to 10% and 91-100%), total and random accuracy are almost of equal value, at around 95%. The equivalence of the two accuracy measures signifies that the models classification accuracy is as good as a random distribution's accuracy for very asymmetric class abundance. With such high and similar values as input to the KIA ratio calculations, small differences between total and random accuracy can lead to large variations of KIA. This makes KIA unstable in uneven class proportions, just as pointed by Foody (2004). This unevenness of the classes could be caused by the predominance of low water settings ($n_{(0\text{-}10\%)}$ = 339 cells ; $n_{(91\text{-}100\%)}$ = 67 cells) from the UMD reference dataset in the training subset. Furthermore, features of low inundation fraction settings, such as rivers and lakes, are believed to be more geomorphically distinct than rarely inundated areas. Distinctiveness of these areas would cause the model to be capable of improved discrimination and redistribution in these areas.

The slight improvement (~1%) of accuracy brought by the majority filter is present among both training/validation and accuracy metrics. This improvement from the majority filter can be found for most inundation fraction ranges, except for a few ranges where it has caused a decrease in accuracy. The most consistent decreases in accuracy exist for very high inundation fraction cells (91-100%), which are reduced by ~1-2% for both metrics. A decrease in accuracy was also expected for the 0 – 10% range, where alterations by the filter have a relatively large importance in comparison with the areas mapped, but was found only for the Producer's Accuracy. Despite similarities, the metrics of the training and validation subsets display some divergence. The most plausible explanation for this difference is that a group of cells with distinct inundation patterns were randomly selected in the training and validation subsets, although they were shown to be equivalent in many ways (Table 4.4). The two reference data sources (UMD & CARPE) represent the same proportion of both of the subsets, and therefore do not offer alternative explanations for the discrepancy. This explanation implies that a larger validation sample would decrease the accuracy to

the lower but more representative level of the training set. On the whole, accuracy of the spatial distribution in the validation subset was found to be 84% for the producer accuracy, an overall accuracy of 92% and a KIA of 80%, indicating a moderate agreement. The downscaling procedure performance is acceptable in comparison to the Bwangoy et al.'s (2010) producer accuracy of 92%, overall accuracy of 88.6% and a KIA of 68%, especially considering the model's simplicity and its exclusive use of hydro-geomorphic variables.

5.2 Inundated Area Estimates

Misrepresentation of inundated areas of the downscaled map can originate from the coarse estimates or from the reallocation done during the downscaling process. Inundated area errors affect both distribution and total inundated area of the continent. In the case of MAMax, the estimates from the GSWED were left unmodified for lack of means to correct them. Conversely, the fabricated estimates of MaxFusion can be modified, and are thus assessed in this section. The other source of error, from the distribution process affects both MAMax and MaxFusion by locally reallocating inundation and making the downscaled map unfaithful to the original estimates. Reallocation error of the two distribution methods (MWT and SVT) is assessed in this section to orient future work.

5.2.1 Inundated Area from Data Fusion

Another way to address some of the uncertainties of the GSWED estimates, besides temporal aggregation, is through the data fusion with the GLWD. Because the GLWD represents a maximum extent of wetlands that cannot be tied to any return period, the fusion of GLWD with GSWED estimates causes the latter to lose its temporal dimension. The merger of the two datasets with a simple rule-based fusion was designed to fill the gaps of the GSWED, reduce its biases, and generate an overall larger and more complete total inundated area. The average of the two estimates where data from both GLWD and GSWED are available can cause the inundated area of some cells to decrease. The inundation fraction estimates from the fusion as well as original GLWD and GSWED are

plotted for 10% inundation fraction ranges in Figure 5.6. By merging the GLWD to the GSWED estimates, the number of cells with no inundation (0%) of the GSWED declined by over 4000 cells, while the 0 – 10% group did not increase significantly from the GLWD's number, suggesting that the cells having left the 0% group have been moved to several of the higher groups. The reduction of 0% cells is the most important modification to fill the gaps and increase the map coverage. The largest increase of number of fusion cells compared to both sources is found in the 11-20% range. The number of 91-100% inundated cells increased marginally, as both datasets are expected to adequately capture large inundated areas. Overall, the data fusion increases the number of cells with estimates, producing a more complete coverage and reduces the under-representation of moderately inundated areas by GSWED. However, it is not clear whether the inclusion of these two data sources adversely creates "edge effects" in the distribution process.

Over the entire continent, the total inundated area of the downscaled MAMax cumulated 1339 thousand. The GLWD (with lakes and reservoirs) and GSWED total areas of 1499 and 2249 thousand km^2 respectively were combined to generate MaxFusion estimates totally 2692 thousand km^2. The MAMax total area is found to be in good agreement with the GRoWI and GLWD estimates although it includes artificial wetland while these previous estimates did not. The area of MaxFusion is close to double the area estimates, providing the largest total wetland area from any inventory over the African continent. As a point of comparison for the MaxFusion area, a previous version of the fusion procedure taking the maximum value of either source in each cell generated a total area of 3136 thousand km^2 which was deemed unreasonably high.

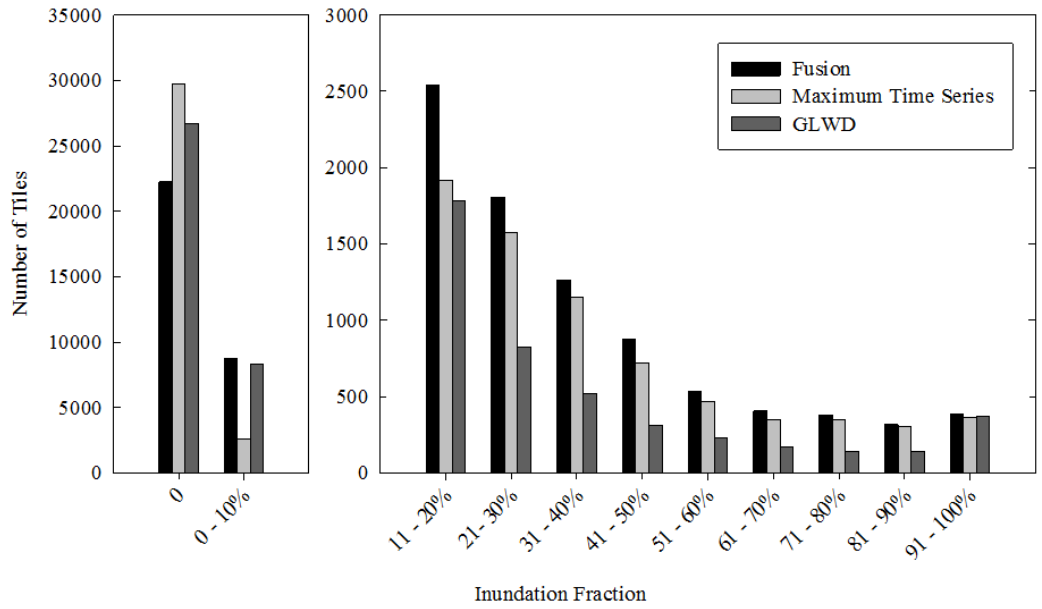


Figure 5.3: Histogram of number of cells per 10% inundation fraction range. The 0 and 0-10% ranges contain many more cells than the other ranges, and have been plotted on a different scale.

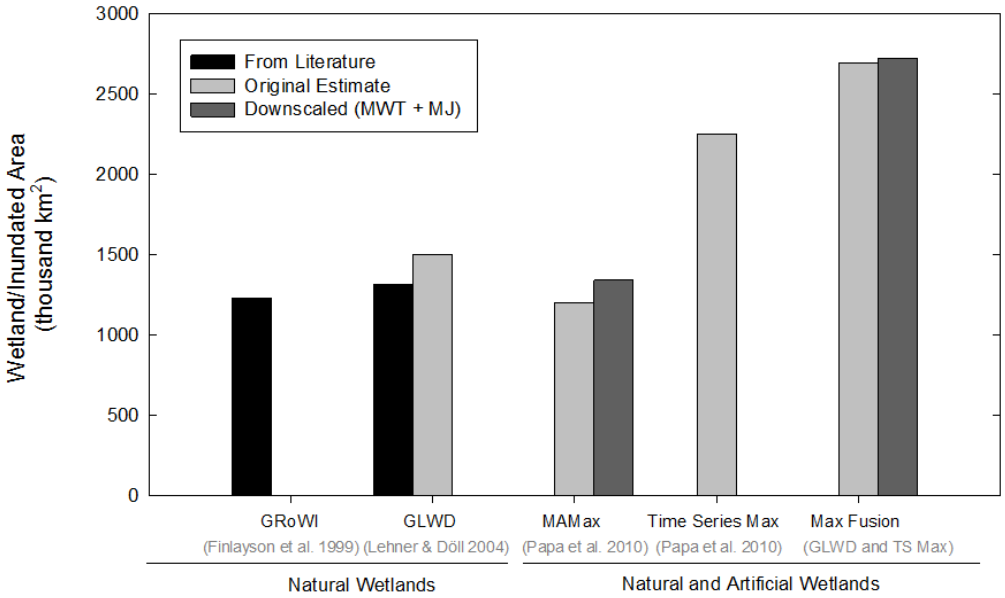


Figure 5.4: Total inundated wetland area comparison from various sources

5.2.2 Reallocation of Inundated Area

The moving window thresholding process was included in the methodology instead of a single cell thresholding to reduce the occurrence of "edge effects". Edge effects are eliminated through a reallocation of a portion of the inundated extent of a cell among its adjacent cells. Reallocation of inundated area also makes the downscaled inundated area inconsistent with the original inundated area locally within each cell. These errors contravene to the downscaling procedure remaining faithful to the original GSWED estimates. Through this reallocation, the moving window thresholding has the potential side effect of slightly modifying the total original area inundated from the cells. By attempting to replicate the number of inundated pixels of original estimate as close as possible, a residual is sometime introduced due to multiple pixels of same probability. Theoretically, the inundated area created and removed from the thresholding should cumulatively produce a map of roughly equal inundated extent, but it is possible that a marginal difference between the original and redistributed estimates is introduced. Measures of alteration of the per-cell and total inundated area estimates are reported in Table 5.3 for both temporal aggregations and for both distribution methods over the continent and the validation set. The analysis of the displacement and calculation of RMSE was limited to the extent included in the distribution process and where inundation can be found, which represents a 1-cell buffer around cells with nonzero inundation estimates.

The overall change in total inundated area is not noticeably different for the MWT and SVT methods, fluctuating within 2% of over and under-representation. While the total change in inundated area is acceptable considering the scale of the study area, the displacement of inundation among cells varies wildly between the distribution methods. The relative importance of reallocation measured as CV-RMSE is in the order of 120% for MWT and a much higher 200% for Single Threshold Value. The much higher RMSE values of the SVT method are due to the absence of barriers to reallocation. In the Single Value Threshold, inundated

area originally found in a corner of the continent can be reallocated anywhere else in the one cell buffer, creating ample of opportunities for reallocation. The large areas of reallocation of inundation data can also, to some extent, be explained by the problem in geolocation between large water bodies and coastlines. Spatially, the SVT method redistributes inundation away from lower probability regions to higher probability areas with no inundation. These patterns are clearly shown in Figures 5.5 and 5.6. For the MWT, local reallocation achieved to eliminate edge effects shows no large overall pattern. Areas of high reallocation around the Nile River and large rift lakes are due to the disagreement in geolocation between the coarse estimates and the probability map.

Table 5.3: Accuracy metrics evaluating the displacement (Per-Cell CV-RMSE) and alteration of total area (Total Change) for MWT and SVT distribution methods with and without the majority filter.

		MWT	**MWT + Maj.Filt**	**SVT**	**SVT + Maj.Filt**
MAMax	**Total Change (%)**	2.6%	2.1%	0.4%	-0.4%
(n_{cells} = 17,777)	**Per-cell CV-RMSE (%)**	125.8%	129.8%	215.9%	218.9%
MaxFusion	**Total Change (%)**	3.3%	1.2%	-0.9%	-1.9%
(n_{cells} = 33,156)	**Per-cell CV-RMSE (%)**	107.2%	110.6%	194.1%	196.7%

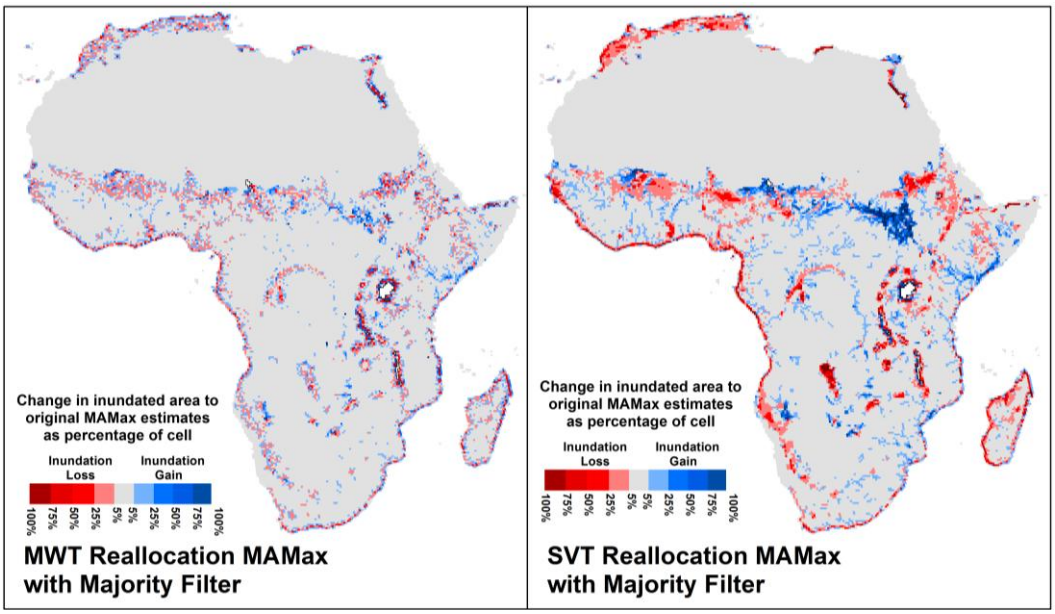


Figure 5.5: Reallocation rates of Mean Annual Maximum (MAMax) estimates of GSWED for the MWT and SVT methods.

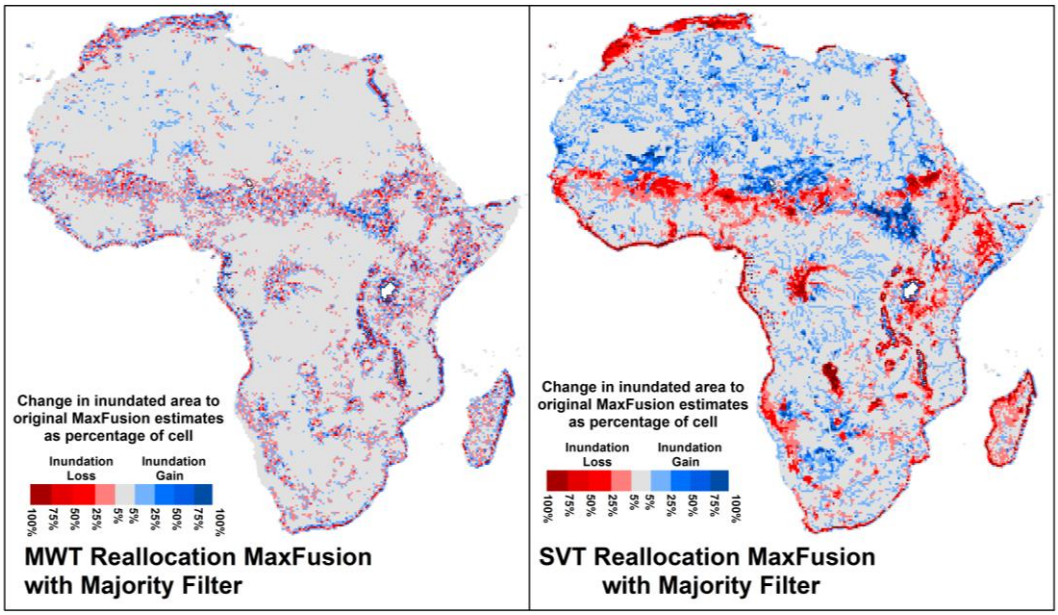


Figure 5.6: Reallocation rates of MaxFusion estimates for the MWT and SVT distribution methods.

5.3 Regional Case Studies

Wetland service provision and biodiversity increase with wetland area (Keddy et al. 2009), making the largest wetlands the most ecologically critical features on the map that must be accurately represented. An assessment of the representation of maximum wetland extent by the downscaled MaxFusion map of some of Africa's largest wetlands was accomplished through comparison to other datasets – the GLWD and GLC2000 (Mayaux et al. 2004). Based on the accuracy assessments of the prior sections, the compared MaxFusion map was produced from the best downscaling method, resulting from the MWT and majority filter. Five regional case studies depicting temporary wetlands were selected (Figure 5.7): Okavango Inland Delta, Sudd Marshes, Nile River floodplain, Congo River floodplain and the Zambezi River delta.

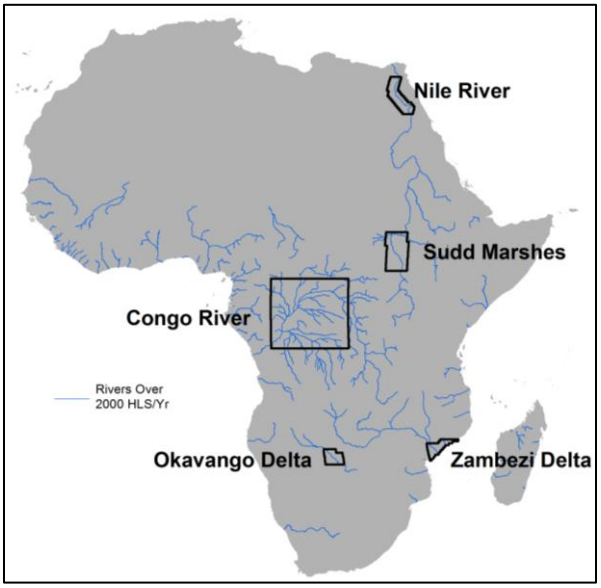


Figure 5.7: Regional cases studies outline across Africa

For each of the five case studies, the downscaled MaxFusion and the GLWD were compared to the GLC2000. The wetland area of GLC2000 and GLWD redistributed by the downscaling process was also compared to the GLC2000. The comparison of the redistributed GLC2000 and GLWD over the study regions provides accuracy benchmarks of the downscaling process by excluding the effect

of different inundated areas on the accuracy. In an effort to provide evidence that the downscaling process can generate a map superior to GLWD, the redistribution of the GLWD area provided an accuracy baseline for the downscaling process at equal area with the GLWD. Redistribution of the GLWD and GLC2000 was done with the identical process used to produce the MaxFusion map of the comparison. The five wetland representations are shown in Figure 5.8 over the study sites. Agreement between the wetland representations is evaluated with the Kappa Index Agreement (Table 5.4) as well as the total inundated area (Table 5.5), and details of each study site are discussed in Table 5.6. Assuming that the total inundated area is not significantly altered by the redistribution process, as confirmed by section 5.3.2, areas of the GLC2000 and GLWD are reported only once. The wetland areas of the sites are compared to the ones in the literature wherever possible.

In terms of area, the wetland area depicted by the GLC2000 is, in most cases, lower than the one from GLWD and MaxFusion. The wetland representation of the GLC2000 relying on two composite snapshots of the years 1994 and 1996, likely does not capture the maximum wetland extent as does the GLWD. The MaxFusion was intended to depict maximum wetland extent, and accordingly proves to indeed have a higher area than of GLWD, with the main exception of the Congo floodplains where fusion method produced MaxFusion estimates lower than GLWD. MAMax areal estimates are systematically lower than all other estimates. GLWD, GLC2000 and MaxFusion estimates are overall equal or lower than area estimates from the literature. There exists little agreement between the total areas provided for these three sources for most study sites.

In terms of spatial distribution accuracy, the redistribution of the GLC2000 by the downscaling process reaches the highest levels of accuracy except in the case of the Okavango. The redistribution of the GLWD leads to an improvement in accuracy over its original version in three of the five sites. Decreases in accuracy from the original GLWD are experienced for the two wetlands with the most

complex inundation patterns - the Okavango and Sudd Marshes. Across the five study sites on the continent, the agreement between MaxFusion and GLC2000 is moderate and in most cases weaker than for GLWD, but its KIA values were more consistent across sites than for GLWD. The MaxFusion map surpasses the levels of agreement of the original GLWD only for the Nile and Zambezi study sites.

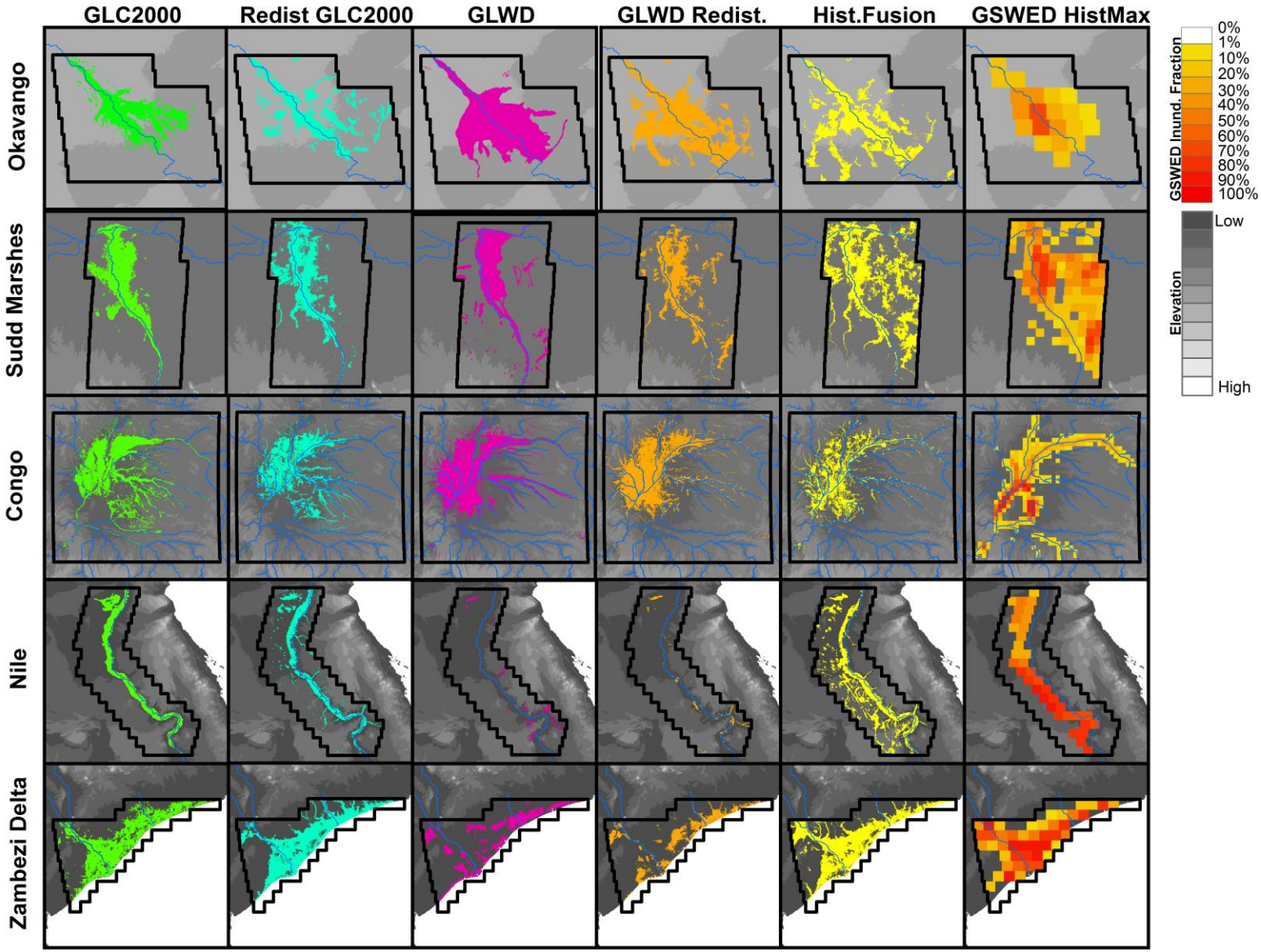


Figure 5.8: Comparison of wetland representation over the five study sites

Table 5.4: Kappa Index of Agreement for the data sources against the original GLC2000 over the five case studies

	Redist. GLC2000	GLWD	GLWD Redist.	Max Fusion
Okavango	41.6%	51.2%	40.1%	37.2%
Sudd	63.7%	58.3%	50.5%	34.9%
Congo	61.6%	51.4%	52.8%	41.6%
Nile	69.6%	3.9%	14.4%	47.6%
Zambezi	59.7%	25.0%	28.9%	52.8%

Table 5.5: Inundated area, in km^2, for the four different data sources and from literature over the five case studies

	GLC2000	GLWD	Max Fusion	GSWED MAMax.	Others
Okavango	8,528	14,969	12,573	3,596	28,000 a
Sudd	31,331	32,961	57,589	24,216	41,334 b
Congo	143,256	198,179	134,061	44,796	360,000 c 189,000 d
Nile	12,475	2,877	20,812	18,772	-
Zambezi	16,660	8,055	18,309	7,396	-

a – Junk et al. 2006, b – Rebelo et al. 2011, c – Bwangoy et al. 2010, d – Keddy & Fraser 2005

Table 5.6: Comments on comparison of total inundated area and spatial distribution of area for the study sites

Cases	Description	Comment on total inundated area	Comment of spatial distribution
Okavango Delta	Distal part of the endorheic Okavango drainage basin whose unimodal flow pulse drives the extraction and contraction cycle of flooding (Murray-Hudson et al. 2006).	The areas of both the GLC2000 and the GLWD are lower than the estimate provided by Junk et al. (2006) of which the source is unknown. The area of MaxFusion realistically sits between that of GLC2000 and GLWD.	Not well replicated by any of the redistributed maps. Accurate representation of the delta is caused by HydroSHEDS only including confluence processes and not river bifurcation (Lehner et al. 2008).
Sudd Marshes	One of the largest wetlands in the world (Rebelo et al. 2009). Because of its remoteness and difficulty of access as well as the complex flow dynamics from multiple inflows, the intra and inter annual variations of the Sudd Marshes remain poorly mapped and studied.	The representation of Sudd marshes by the GLC2000 and GLWD are similar in their distribution and total wetland area. The area of MaxFusion stands at nearly the double of the ones of GLC2000 and GLWD because of it also represents inundated areas distinct from the main floodplain.	The Sudd depression wetlands extent wider than what is included in the study site and captured by the MaxFusion map. Over the larger region of the Sudd depression, the GLWD and MaxFusion map product do not agree with GLC2000 as well as they did over the case study outline.
Congo River	Delineation of the large forested lowlands annually oscillating between terrestrial and aquatic phases has been a particular focus in recent years (De Grandi et al. 2000, Rosenqvist et al. 2007; Vancutsem et al. 2009).	The inundated area of MaxFusion is lower than that of GLWD because the GSWED estimates do not capture the inundation beneath the forest canopy. Due to this, the total area from MaxFusion agrees well with the one of GLC2000. GLWD's area estimate agrees with Keddy and Fraser's (2005) figure, but both are dwarfed by the area reported by Bwangoy et al.'s (2010).	In spite of being the largest wetland among the study sites, the Congo is reasonably well represented by each of the sources in terms of the KIA value. In fact, even the accuracy of the representation of the Congo floodplain was slightly improved through the redistribution of the GLWD.
Nile River	The Nile River floodplain allowed early civilizations to emerge and prosper. With time however, dam construction has tamed the river and the naturally inundated floodplains have been converted to permanent agriculture irrigated from the Nile's waters.	The total inundated area of MaxFusion is larger than GSWED and GLWD due to an error of geolocation, causing a cumulation of both instead of the average from the two. Disagreement between the GLC2000 and the GLWD is due to the presence of artificial wetlands included in GSWED and MaxFusion but excluded in GLWD.	The redistribution effectively aggregated the small wetland patches of the GLWD into the floodplain, thus improving the accuracy.
Zambezi Delta	The Zambezi river delta is portrayed by the GLC2000 as flooded shrubland or grassland and mangroves. Although mangroves are not inland wetlands, their areas overlapping with the HydroSHEDS DEM is mapped nonetheless.	GLWD underestimates the extent of wetlands along the coast and in the delta, for an unknown reason. It is one of the rare areas where the GLC2000 possesses an estimate larger than the GLWD.	GLWD's patchy wetland distribution is clumped together by the redistribution, improving the overall accuracy. The high accuracy of the MaxFusion map, confirms that the downscaling behaves well in coastal delta settings although those were not part of the training subset.

5.4 Continental Map Description

Considering the insights gained in the accuracy assessments covererd in the previous sections, the downscaled inundation maps of MAMax and MaxFusion from both distribution methods MWT and SVT, are presented in Figures 5.9 and 5.10. All four maps have been processed with the majority filter because of the ubiquitous improvements in accuracy it brings.

5.4.1 Visual Assessment

Visually, the distribution of MAMax inundation extent across the continent resembles the original GSWED, with most large features such as lakes and wetland complex present. Comparing the inundation maps of Figure 5.1 with the probability map of Figure 5.2 reveals that most high probably areas of the continent are covered by inundation, often by both MAMax and MaxFusion. As concluded in section 4.4.2, noticeable differences are present in the distribution of inundated lands between the MWT and SVT maps, due to the reallocation limits between the two redistribution methods. Some regions of the continent displaying substantial inundation in the MWT map can show completely different patterns in the SVT map, and vice-versa. The most visible examples of these differences are found in the Sahara and Kalahari deserts where the SVT depicts considerably more inundated area than the MWT. Also visible on the maps are the inland gaps, named "Missing Cells", where GSWED cells were absent. These gaps, almost exclusively present over large water bodies (Lakes Chad, Victoria, Tanganyika and Malawi), are assumed to be completely and permanently inundated.

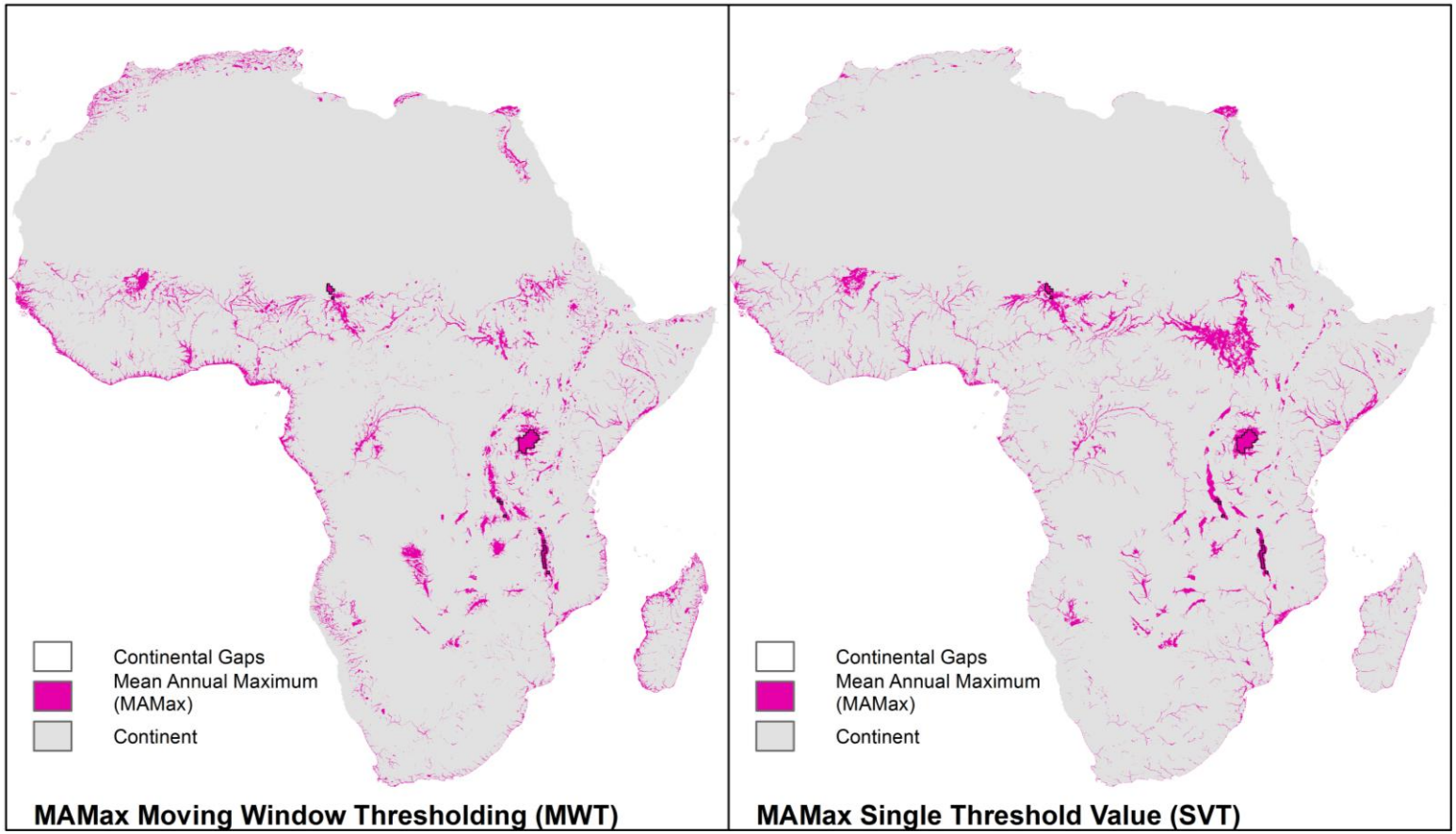


Figure 5.9: Downscaled MAMax inundation maps of the african continent for Moving Window Treshold (MWT) and Single Value Theshold (SVT) methods.

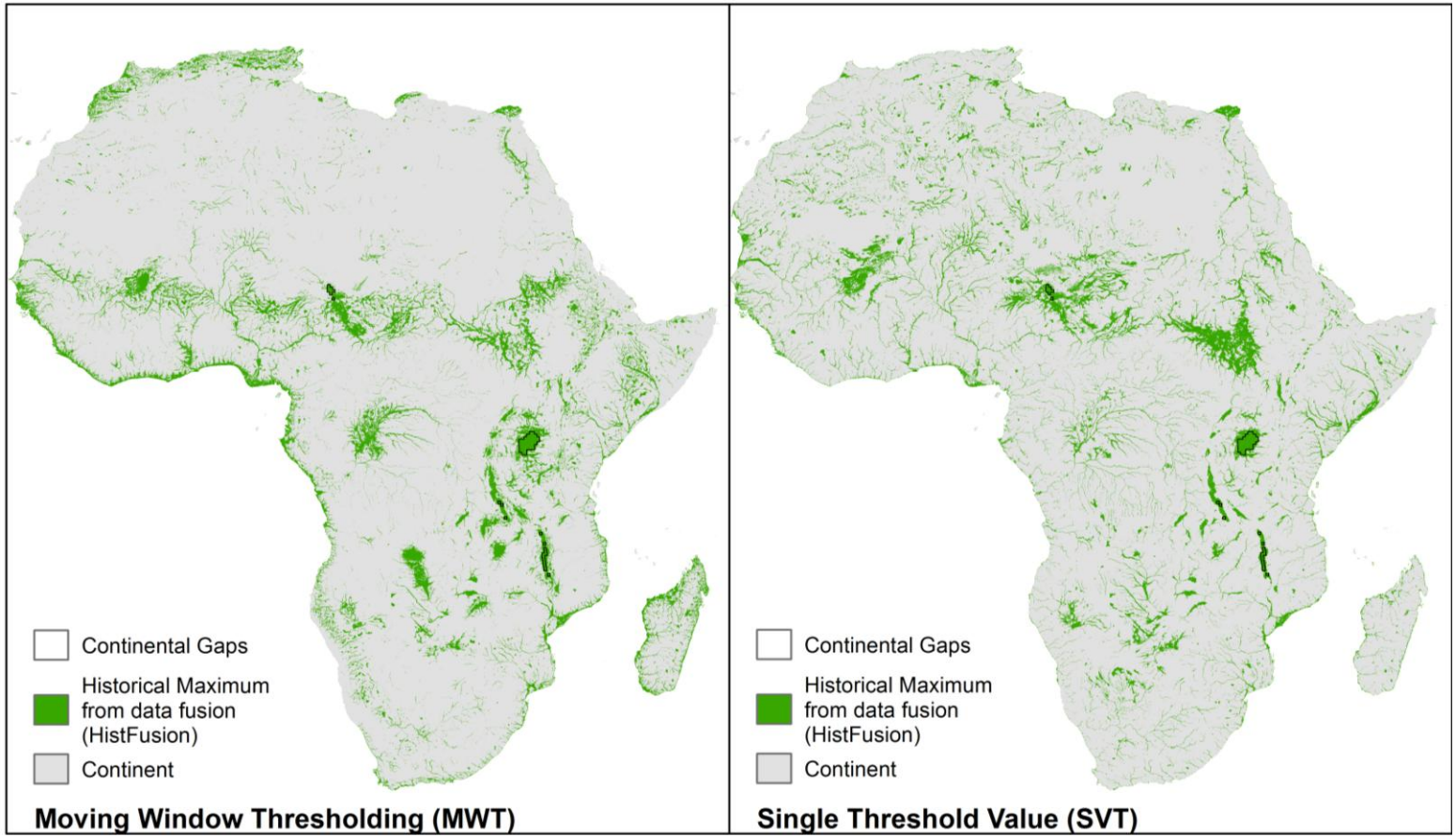


Figure 5.10: Downscaled MaxFusion inundation maps of the african continent for Moving Window Treshold (MWT) and Single Value Theshold (SVT) methods.

5.4.2 Observable Defects

Within the continent-wide map, some visually detectable erroneous features are recurrent over the landscape. These defects are not necessarily captured by the accuracy assessments, as they can be the result of a combination of factors. The most typically observable errors are described in this section.

Edge Effects

Although the MWT was implemented to eliminate "edge effects", some of these features persist in the final map. Remaining edge effects are mostly found as disconnected river segments, for example in the case of the Nile and the Senegal rivers, as well as at the border between Sahel and Sahara where the MWT process is insufficient to prevent edge effects from appearing. Edge effects are most prevalent in the MAMax map suggesting that patchiness of GSWED estimates causing edge effects is somewhat corrected by its fusion with the GLWD. Although the Single Value Threshold generates a much smoother map, some edge effects do still appear, likely caused by the one cell buffer to which the thresholding was limited.

Low Probability Patches

In some upstream regions, patches of pixels, possessing a similar but generally low probability value sometimes leads to the entire patch being defined as inundated by the thresholding procedure at once (Figure 5.11). The patches are caused by insufficient variety in the probability values distribution for these upstream areas. It is expected that larger decision trees with more varied inputs would diversify the probability values of these pixel in the patches.

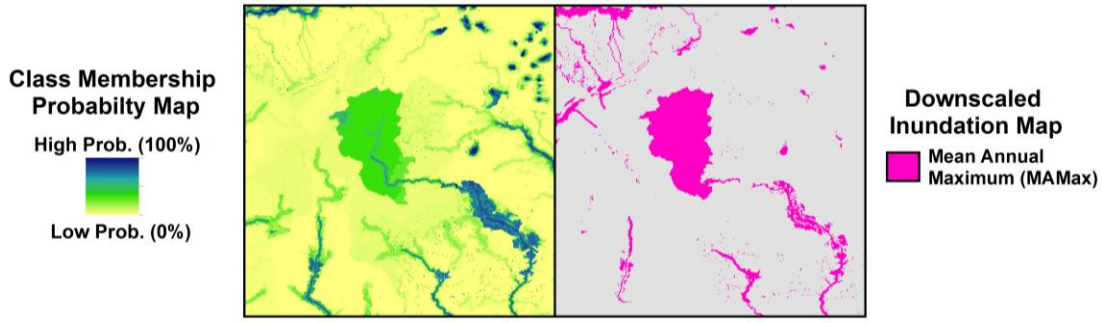


Figure 5.11: Low probability patch in probability map (left) and inundation map (right)

Coastal Regions

Probabilities located in coastal cells display strange regular linear features that no algorithm correction could fix (Figure 5.12). Most of the pixels forming these linear features possess the same probability value, which allowed the alternate solution of reclassifying the pixels to a lower probability value commonly found around coastal regions. Regardless of the correction, some remaining linear features affect the redistribution of inundated areas in those regions and appear within them. This behaviour was not present in the probability raster of other continents such as Europe and South America.

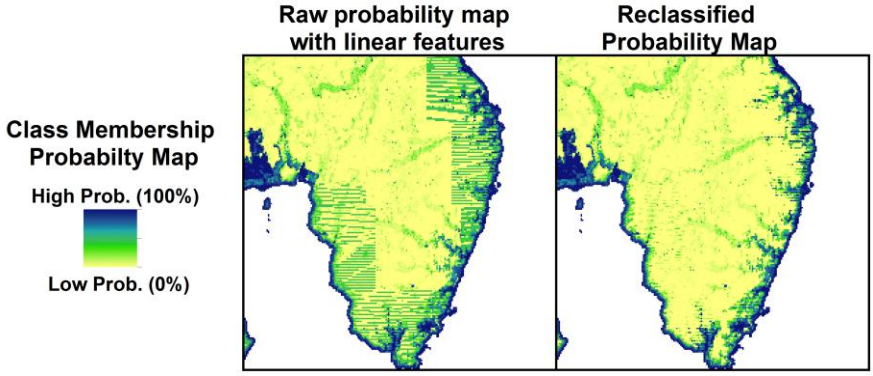


Figure 5.12: Example of coastal linear features before (left), and after (right) reclassification of probability values

6 Discussion

6.1 Assessment of methodology

The accuracy assessments presented in the previous sections provide a comprehensive evaluation of the sources of error in the methodology. The downscaling accuracy assessment, although applied on the limited area of the validation and training subsets, performed well. The level of accuracy reached is deemed to be sufficient for a global dataset, but should be obtained over larger validation extents to be truly robust. Reallocation of inundation among cells by the MWT appropriately reduced the presence of edge effects, while remaining truthful to the original estimates. The SVT method created displacements of inundation twice as large as MWT, often across large distances, which created downscaled outputs too different from the original estimates to be trusted. The SVT could perhaps prove useful over smaller, more climatically homogenous regions. The data fusion of the GLWD and GSWED proved to be capable of extending the coverage of GSWED and correct for its bias toward very low and very high inundation fractions. The majority filter slightly improved accuracy. Some visible recurring errors – for instance the edge effects - prevent the downscaled map from being fully reliable. The level of displacement of inundated area by the MWT is likely close to the minimum necessary to eliminate most of the edge effects and make it visually pleasing. Overall, the accuracy assessments and comparison results have not clearly proven that the current iteration of the downscaled MaxFusion wetland extent map is superior to GLWD over the African continent.

Going forward, the downscaling process output holds many fundamental advantages over the GLWD in terms of possible refinements to the downscaling procedure and input data which may materialize in future iterations of the map products. Moreover, the early stage of development of GSWED data opens the prospect of anticipated newer improved GSWED-type data to be downscaled in future iterations. More reliable GSWED estimates could be used without the data fusion, and would allow for mapping inundation at a finer temporal resolution

with confidence, which is not currently the case. Another advantage is the products' compatibility with the HydroSHEDS river network that makes it attractive for an array of future applications described in section 7.

6.2 Limitations of HydroSHEDS

HydroSHEDS, from which the hydro-geomorphic variables were calculated, contains inherent unaccountable errors carried to the output results. A design flaw of the methodology is the use of the static topography of the year 2000 of HydroSHEDS to downscale inundation estimates ranging between 1993 and 2004. Although topography does not drastically change over yearly to decadal time scales, the topography captured by the SRTM of rivers, lakes and reservoirs can be affected by the date and seasonal timing at which the SRTM snapshot was taken. Lakes and rivers observed by SRTM during their high water levels have their coastlines and floodplains masked by the surface water measured by the interferometer. As a result, the submerged topography exposed only at low water levels is unavailable to appropriately direct redistribution of inundation among these areas. Any reservoirs built after the SRTM snapshot would appear in the GSWED estimates but would not be in the DEM. These instances would not be well represented with the current method but could be resolved with the use of a more recent DEM around these particular areas. Likewise, the effect of vegetation canopy on the DEM and probability induction is acknowledged but not investigated for this thesis.

6.3 Recommended Improvements

In light of the limitations of the results of this thesis, various improvements to the methodology are prescribed for future iterations to increase accuracy and extent the application of this method beyond the African continent.

6.3.1 Hydro-Geomorphic Variables

An analysis of the contribution of each hydro-geomorphic variable to the decision trees was not accomplished as part of this study. The relative contribution of

variables was calculated from the number of instances where each variable was used in the splitting rule with the improvement of that split to the model. Bwangoy et al. (2010) and Elith et al. (2008) were able to identify which of their predictor variables were most useful by calculating the relative contribution of each variable (Friedman & Meulman, 2003). The Orange Data Mining classification tree of the software tool library does not include a built-in module generating relative contribution of each variable. The migration of the algorithm to another platform possessing an implementation of Friedman & Meulman's (2003) equation, such as R's gbm module, could make evaluating the contribution of each variable possible. Calculating the relative contribution could justify the removal of trivial variables and the inclusion of additional relevant ones (e.g., local flow accumulation, sink type), which could improve the model's predictive capacity. Moreover, benefitting from monthly discharge values attached to HydroSHEDS river reaches (Lehner, in prep.), distance and elevation to rivers of certain size could be recomputed from the downscaled global hydrological model WaterGAP outputs. Using monthly discharge has the potential of improving the downscaling performance, particularly at monthly time scales.

6.3.2 Optimization of MWT

The MWT process involves a trade-off between authenticity to the original estimates and presence of edge effects. The adjustment of the Residence Ratio parameter of the MWT allows managing this trade-off, but an exhaustive sensitivity analysis in search of an optimal value for the Residence Ratio parameter value was not undertaken. An optimal Residence Ratio should strike a balance between rigidity regarding the original estimates and presence of edge effects. In addition to the Residence Ratio, the size of the moving window (i.e., number of cells in the window) also affects output map product by defining the area over which reallocation of inundated area can occur. Larger window can cause a more pronounced departure from the original inundation estimate distribution, but does not directly increase reallocation. An optimization of the MWT process could consist of a combination of the Residence Ratio and Moving

Window size. It is expected that a larger window in combination with a stricter residence ratio can generate fewer edge effects while remaining true to the original estimates, by reallocating less inundated area from more cells to correct edge effects.

6.3.3 Additional Reference Datasets

Including more reference datasets in the training set would increase the area and diversity of inundation/wetland representations and settings. Increasing the diversity of wetlands represented would improve the model's accuracy over non-tropical regions, not currently represented in the training data. Additional existing reference data sources deemed adequate for use are listed in Table 6.1.

Table 6.1: Additional reference data sources which could enhance representation of area and diversity for iterations of the inundation map product

Data Source	Region	Methodological Comment
Dual-Season Inundation (Hess et al., 2003)	Central Amazon Basin (18° × 8° region)	From SAR observations, high (May – Aug 1996) and low (Aug – Sep 1995) water levels snapshots.
Wetland Map (Hess et al., 2003)	Entire Amazon Basin	Classified from SAR and topographic data (< 500m elevation).
Wetlands GIS of the Murray-Darling Basin Series 2.0 (Murray Darling Basin Authority	Murray Darling River Basin	Maximum wetland extent over a ten-year period (1983-1993), from unsupervised classification of Landsat MSS imagery and additional wetland data.
Africover) (FAO 2004)	Central Africa (Sudan, Tanzania, RDC, Kenya, Uganda, etc.)	Manualy interpreted from Landsat imagery.

6.3.4 Training & Validation Subsets

The five decision trees of this project were trained with 50,000 pixels, representing less than 1% of the land pixels of the continent. Increasing the area covered by the subsets would make the extrapolation of the training more robust and validation more representative. Increasing the area can be done through the

inclusion of additional reference datasets (previous section) or by loosening the sample selection rule.

6.3.5 Decision Tree Size

Preliminary analysis of a previous version of the algorithm has shown that increasing training sample size and number of trees did not considerably improve accuracy over the validation set (Table 6.2). The effect of the size and number of the trees should be tested with a wider range of reference data sources.

Table 6.2: Preliminary comparison of tree and training set size on spatial distribution accuracy measured with the Kappa Index of Agreement (KIA)

	Bagging #1 5 trees 50,000 pixels per tree 1000 pixel end rule	**Baggin #2** 10 subsets 100,000 subsets 1000 pixel end rule
UMD	60.4 %	60.3 %
CARPE	69.8 %	69.4 %

6.3.6 Geolocation of GSWED Cells

The cells from the first version of the GSWED dataset were limited to only two decimals coordinates, creating a coverage with large gaps and overlaps. The process of modifying the cell coordinates to produce an even coverage may have introduced geolocation errors of the cell center and outline. The second unfiltered version GSWED data acquired from Dr. Fabrice Papa (October 2011) possesses more detailed coordinates and should improve geolocation of cells.

6.3.7 Seeded Region Growing

An alternative to the majority filter to correct for noise in the probability map consists of distributing the inundated area using a seeded region growing (SRG) method instead of the MWT. A seeded region growing process is an object-oriented method in which an object is grown from defined seeds and whose growth is limited to areas connected to the object. In the context of flooding and

inundation mapping, the SRG process intuitively recreates the dynamics of the flood wave rising, leaving the channel and growing towards the uplands, and ensures that flooded areas are contiguous in some way to the river channels. By avoiding cases where low-lying sinks disconnected from the river network are considered as wetlands to be classified as flooded, the SRG can theoretically improve classification accuracy from a simple thresholding procedure. Although the SRG improves classification as compared to the MWT thresholding, both methods should be extensively compared to reach a conclusion on the value of SRG. A SRG process was developed as part of this project, but the MWT method was preferred over the SRG based on on preliminary analysis (not included) in which SRG showed only marginal improvement in accuracy despite the longer processing time it required.

The SRG procedure that was designed but not used in this thesis was inspired by the one of Straumann & Purves (2008), used for delineation of valleys as defined as flat and low-lying landforms, and in which the growing procedure was initiated at thalwegs (i.e., lowest points of a valley or river). In their implementation, their growing process iterated until no more pixels, with a slope below a certain threshold, were detected. In the adaptations of their growing procedure, growth is seeded at the thalweg and is driven by the probability ranking generated from the decision tree. The region iteratively grows to include the highest ranking pixels directly adjacent to pixels previously recognized as inundated. This growing process is repeated until the area covered by the inundated region is equivalent to the GSWED inundated area estimate of the cell. The SRG, just like the thresholding can be used as part of the moving window reallocation.

7 Conclusions & Future Work

7.1 Conclusions

This project was able to address the two questions defined in the research objectives section. The first question asked whether hydro-geomorphic data alone could serve as an accurate predictor of wetland extent when compared to regional maps. Results from this project show that the set of hydro-geomorphic variables can indeed reasonably well represent inundated areas when evaluated against regional data. When compared to the GLWD, the downscaled map showed lower but more consistent agreement with the GLC2000. The second question addressed the validity of fusing GLWD and GSWED to generate a 'reliable' global wetland area estimate. Results show that the combination of GLWD and GSWED can generate a reliable maximum inundation/wetland extent, although this is achieved at the expense of the temporal resolution of the GSWED.

Aside from supplying the wetland research community with a new wetland map for the continent of Africa, this project contributed wetland mapping methodological advances. The most noteworthy methodological contribution is to demonstrate the achievable level of inundation mapping accuracy by using exclusively hydro-geomorphic variables. The Moving Window Thresholding (MWT) method, developed for this particular project, also constitutes a methodological contribution of this study. The spatially-consistent wetland map from this project for the continent of Africa, and eventually globally, represents a new asset to study the global freshwater system. Moreover, being based on the HydroSHEDS data, this method could be included in a variety of future studies and applications, such as the ones described below.

7.2 Future Work

7.2.1 Hydro-Geomorphic Classification

A classification based on hydroperiod and topographical characteristics possesses the clear advantage of being globally consistent, unlike the current Ramsar classification scheme which fails to properly reconcile wetland classes from different regions of the globe. In fact, the Scientific & Technical Review Panel (STRP) of the Ramsar Convention has identified the development of a hydro-geomorphically based wetland classification as a research gap (Davidson & Finlayson, 2007). A possible follow-up application of such a downscaled map product would consist of using globally available physical variables, such as hydroperiod and topography to produce a global hydro-geomorphic wetland map. The result would divide the downscaled inundation map into a few hydro-geomorphic classes attached to homogeneous habitat patches, and would constitute an essential step toward a globally consistent classified wetland inventory.

7.2.2 Global Ramsar Site Delineation

The spatial database of Ramsar sites, storing geographical location and characteristics of each Ramsar sites suffers from the discrepancies among sites worldwide because of the lack of a standardized national inventory. Downscaled inundation maps could be used to delineate Ramsar sites in a consistent way globally through a region growing process (SRG) seeded at the site's point coordinates growing outward from it. At the minimum, Ramsar sites are represented by a single point coordinate falling somewhere within the designated site extent. Often, several additional characteristics of the wetland are inventoried in the database, such as surface area, average elevation, wetland and vegetation type, which could be used to delineate the wetland surrounding the points

7.2.3 Future Global Inundation Maps

An application that would benefit immensely from the downscaled map's relation with HydroSHEDS would be one that attempts to establish relationships between

hydrograph components (e.g., peak flow, low flow) and habitat types and abundance. Large uncertainties regarding wetlands response to climate change exist. Statistical regression models between downscaled discharge and inundated area built for each river reach documented in HydroSHEDS and based on long-term averages could be used to predict the inundated area under past and future discharge scenarios. The objective of this work would be to produce wetland/freshwater habitat maps showing river flow and forecasts for 2050 and 2100 in response to climate change and anthropogenic water use projections. This would improve current understanding of the past changes undergone by freshwater habitats due to humans as well as of future changes likely to occur. The production of these maps relies on the scenarios simulations of future discharge under different climate change and anthropogenic disturbance scenarios, provided by a global hydrological model like WaterGAP (Alcamo et al., 2003).

Literature

Abell, R., Thieme, M.L., Revenga, C., Bryer, M., Kottelat, M., Bogutskaya, N., Coad, B., Mandrak, N., Balderas, S.C., Bussing, W., Stiassny, M.L.J., Skelton, P., Allen, G.R., Unmack, P., Naseka, A., Ng, R., Sindorf, N., Robertson, J., Armijo, E., Higgins, J.V., Heibel, T.J., Wikramanayake, E., Olson, D., López, H.L., Reis, R.E., Lundberg, J.G., Sabaj Pérez, M.H. & Petry, P. 2008. Freshwater ecoregions of the world: A new map of biogeographic units for freshwater biodiversity conservation, Bioscience, vol. 58(5), pp. 403-414.

Adam, L., Dol1, Prigent C., & Papa F., 2010. Global-scale analysis of satellite-derived time series of naturally inundated areas as a basis for floodplain modeling. Advances in Geosciences, 27, pp.45-50.

Alcamo, J., Döll, P., Henrichs, T., Kaspar, F., Lehner, B., Rösch, T. & Siebert, S. 2003. Development and testing of the WaterGAP 2 global model of water use and availability, Hydrological Sciences Journal, vol. 48(3), pp. 317-338.

Alsdorf, D.E. & Lettenmaier, D.P., 2003. Tracking fresh water from space. Science, 301, pp.1491-94.

Alsdorf, D., Bates, P., Melack, J., Wilson, M. & Dunne, T. 2007. Spatial and temporal complexity of the Amazon flood measured from space, Geophysical Research Letters, vol. 34(8).

Ajtay G.L., Ketner, P. & Duvigneaud, P. 1979. Terrestrial primary productivity and phytomass. In The global carbon cycle, eds B Bolin, ET Degens, S Kempe & P Ketner, Scope 13, Wiley, Chichester, New York, Brisbane, Toronto, 129–181.

Aselmann, I. & Crutzen PJ. 1989. Global distribution of natural freshwater wetlands and rice paddies, and their net primary productivity, seasonality and possible methane emissions. Journal of Atmospheric Chemistry 8, 307–358.

Baker, C., Lawrence, R., Montagne, C. & Patten, D. 2006, "Mapping wetlands and riparian areas using landsat ETM+ imagery and decision-tree-based models", Wetlands, vol. 26, no. 2, pp. 465-474.

Bousquet, P., Ringeval, B., Pison, I., Dlugokencky, E.J., Brunke, E., Carouge, C., Chevallier, F., Fortems-Cheiney, A., Frankenberg, C., Hauglustaine, D.A., Krummel, P.B., Langenfelds, R.L., Ramonet, M., Schmidt, M., Steele, L.P., Szopa, S., Yver, C. & Ciais, P. 2010. Source attribution of the changes in atmospheric methane for 2006-2008, Atmospheric Chemistry and Physics Discussions, vol. 10(11), pp. 27603-27630.

Breiman, L., Friedman J. H., Olshen, R. A., & Stone, C. J. 1984. Classification and regression trees. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software.

Brown, C.G., Sarabandi, K., & Pierce, L.E. 2005. Validation of the shuttle radar topography mission height data. IEEE Transaction in Geoscience and Remote Sensing, 43(8), 1707−1715.

Bullock, A. & Acreman, M.C., 2003. The role of wetlands in the hydrological cycle. Hydrology and Earth System Sciences, 7(3), pp.358-389.

Bwangoy, J.-.B., Hansen, M.C., Roy, D.P., Grandi, G.D. & Justice, C.O. 2010. Wetland mapping in the Congo Basin using optical and radar remotely sensed data and derived topographical indices, Remote Sensing of Environment, vol. 114(1), pp. 73-86.

Carroll, M.L., Townshend, J. R., DiMiceli, C. M., Noojipady, P. and Sohlberg, R. A. 2009. A new global raster water mask at 250 m resolution, International Journal of Digital Earth, vol.2(4), pp. 291 – 308.

Cieslak, D.A. & Chawla, N.V. 2008, Learning decision trees for unbalanced data.

Cohen, J. 1968. Weighted kappa: Nominal scale agreement provision for scaled disagreement or partial credit, Psychological bulletin, vol. 70(4), pp. 213-220.

Costa, M.P.F., Niemann, O., Novo, E. & Ahern, F. 2002. Biophysical properties and mapping of aquatic vegetation during the hydrological cycle of the Amazon floodplain using JERS-1 and Radarsat, International Journal of Remote Sensing, vol. 23(7), pp. 1401-1426.

Costa, M.P.F. & Telmer, K.H. 2007. Mapping and monitoring lakes in the Brazilian Pantanal wetland using synthetic aperture radar imagery, Aquatic Conservation: Marine and Freshwater Ecosystems, vol. 17(3), pp. 277-288.

Darras, S., Michou, M., Sarrat, C., 1998. A first step towards identifying a global delineation of wetlands. The IGBP-DIS wetland data initiative and the Ramsar convention. IGBP-DIS Working Paper No. 19, Toulouse, France.

Darwall, W.R.T., Holland, R.A., Smith, K.G., Allen, D., Brooks, E.G.E., Katarya, V., Pollock, C.M., Shi, Y., Clausnitzer, V., Cumberlidge, N., Cuttelod, A., Dijkstra, K.-.B., Diop, M.D., García, N., Seddon, M.B., Skelton, P.H., Snoeks, J., Tweddle, D. & Vié, J.-. 2011, Implications of bias in conservation research and investment for freshwater species, Conservation Letters, vol. 4(6), pp. 474-482.

Davidson, N. & Finlayson, M., 2007. Earth Observation for wetland inventory , assessment and monitoring. Earth, 228, pp.219-228.

Decharme, B., Douville, H., Prigent, C., Papa, F. & Aires, F. 2008. A new river flooding scheme for global climate applications: Off-line evaluation over South America, Journal of Geophysical Research, vol. 113(11).

Decharme, B., Alkama, R., Papa, F., Faroux, S., Douville, H. & Prigent, C. 2012, "Global off-line evaluation of the ISBA-TRIP flood model", Climate Dynamics, vol. 38(7-8), pp. 1389-1412.

DeFries, R.S., Hansen, M.C., Townshend, J.R.G., Janetos, A.C. & Loveland, T.R. 2000, A new global 1-km dataset of percentage tree cover derived from remote sensing, Global Change Biology,vol. 6(2), pp. 247-254.

De Grandi, G.F., Mayaux, P., Malingreau, J.P., Rosenqvist, Å., Saatchi, S. & Simard, M. 2000. New perspectives on global ecosystems from wide-area radar mosaics: Flooded forest mapping in the tropics, International Journal of Remote Sensing, vol. 21(6-7), pp. 1235-1249.

De Grandi, G., Philippe, Rauste, Y., Rosenqvist, A., Simard, M. & Saatchi, S.S. 2000. The global rain forest mapping project JERS-1 radar mosaic of tropical africa: development and product characterization aspects, IEEE Transactions on Geoscience and Remote Sensing, vol. 38(5), pp. 2218-2233.

Dudgeon, D., Arthington, A.H., Gessner, M.O., Kawabata, Z., Knowler, D.J., Lévêque, C., Naiman, R.J., Prieur-Richard, A., Soto, D., Stiassny, M.L.J. & Sullivan, C.A. 2006. Freshwater biodiversity: Importance, threats, status and conservation challenges, Biological reviews of the Cambridge Philosophical Society, vol. 81(2), pp. 163-182.

Dugan P. 1993. Wetlands in danger - A World Conservation Atlas, UNEP-WCMC. Oxford University Press, New York.

Durieux, L., Kropáček, J., de Grandi, G.D. & Achard, F. 2007. Object-oriented and textural image classification of the Siberia GBFM radar mosaic combined with MERIS imagery for continental scale land cover mapping, International Journal of Remote Sensing, vol. 28(18), pp. 4175-4182.

Economic Commission for Africa. 2003. The Africa water vision for 2025: equitable and sustainable use of water for socioeconomic development. UN-Water/Africa, Addis Ababa.

Elith, J., Leathwick, J.R. & Hastie, T., 2008. A working guide to boosted regression trees, Journal of Animal Ecology, vol. 77(4), pp. 802-813.

Ellison, A.M., 2004. Wetlands of Central America, Wetlands Ecology and Management 12, pp. 3–55.

Environmental Systems Research Institute (ESRI) (1993), Digital chart of the world, 1:1M, Redlands, Calif.

Falkenmark, M., Finlayson, M., Gordon, L.J., 2007. Water for Food, Water for Life: A Comprehensive Assessment of Water Management in Agriculture: Chapter 6 - Agriculture, water, and ecosystems: avoiding the costs of going too far, London: Earthscan, and Colombo: International Water Management Institute.

Finlayson, C.M., Davidson, N.C., Spiers, A.G. & Stevenson, N.J. 1999. Global wetland inventory - Current status and future priorities, Marine and Freshwater Research, vol. 50(8), pp. 717-727.

Finlayson, M. & D'Cruz, R., Davidson N. 2005. Millenium Ecosystem Assessment - Chapter 20: Inland Water Systems.

FAO – Food and Agriculture Organization. 2004. The Africover Initiative, Food and Agricultural Organisation.

Frappart, F., Papa, F., Güntner, A,, Werth, S. , Ramillien, G., Prigent, C., Rossow, W.B., & Bonnet, M.-P., 2010. Interannual variations of the terrestrial water storage in the Lower Ob' basin from a multisatellite approach. Hydrol. Earth Syst. Sci., 14, pp. 2443-2453.

Friedl, M.A., Brodley, C.E., & Strahler, H. 1999. Maximizing land cover classification accuracies produced by decision trees at continental to global scales. IEEE Transactions on Geoscience and Remote Sensing, 37(2), pp. 969−977.

Friedl, M.A., McIver, D.K., Hodges, J.C.F., Zhang, X.Y., Muchoney, D., Strahler, A.H., Woodcock, C.E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F. & Schaaf, C. 2002. Global land cover mapping from MODIS: Algorithms and early results, Remote Sensing of Environment, vol. 83(1-2), pp. 287-302.

Friedman, J.H. & Meulman, J.J. 2003. Multiple additive regression trees with application in epidemiology. Statistics in Medicine, vol. 22, pp. 1365–1381.

Gallant, J.C. & Dowling, T.I., 2003. A multi-resolution index of valley bottom flatness for mapping depositional areas. Water Resources Research, vol. 39(12), pp.1-4.

Gordon, L.J., Finlayson, M. & Falkenmark, M., 2009. Managing water in agriculture for food production and other ecosystem services. Agricultural Water Management.

Gallant, J.C. & Dowling, T.I. 2003, A multiresolution index of valley bottom flatness for mapping depositional areas, Water Resources Research, vol. 39(12), pp. ESG41-ESG413.

Gurnell, A.M., 1997. The Hydrological and Geomorphological Significance of Forested Floodplains. Global Ecology and Biogeography Letters, 6(3), pp.219-229.

Hansen, M.C., Defries, R.S., Townshend, J.R.G. & Sohlberg, R. 2000. Global land cover classification at 1 km spatial resolution using a classification tree approach, International Journal of Remote Sensing, vol. 21(6-7), pp. 1331-1364.

Hansen, M.C., Roy, D.P., Lindquist, E., Adusei, B., Justice, C.O. & Altstatt, A. 2008, A method for integrating MODIS and Landsat data for systematic monitoring of forest cover and change in the Congo Basin, Remote Sensing of Environment, vol. 112(5), pp. 2495-2513.

Hamilton, S., Kellndorfer, J. & Lehner, B., 2007. Remote sensing of floodplain geomorphology as a surrogate for biodiversity in a tropical river system (Madre de Dios, Peru). Geomorphology, 89(1-2), pp.23-38.

Hervouet, J.-M., Van Haren, L., 1996. Recent advances in numerical methods for fluid flows. In: Anderson, M.G.,Walling, D.E., Bates, P.D. (Eds.), Floodplain Processes.Wiley, Chichester, pp. 183–214.

Hess, L.L. & Melack, J.M. 1994. Mapping wetland hydrology and vegetation with synthetic aperture radar, International Journal of Ecology & Environmental Sciences, vol. 20(1-2), pp. 197-205.

Hess, L.L., Melack, J.M., Novo, E.M.L.M., Barbosa, C.C.F. & Gastil, M. 2003 Dual-season mapping of wetland inundation and vegetation for the central Amazon basin, Remote Sensing of Environment, vol. 87(4), pp. 404-428.

Hodges, J.C.F., Friedl, M.A., Strahler, A.H., 2001. The MODIS global land cover product: new data sets for global land surface parameterization. Proceedings of the Global Change Open Science Conference, Amsterdam.

Homer, C., Huang, C., Yang, L., Wylie, B., and Coan, M., 2001. National Land Cover Database. USGS Eros Data Center. Digital Media.

Hughes, R. H., and Hughes, J. S., 1992. A Directory of African Wetlands, IUCN: Gland, Switzerland. United Nations Environment Programme: Nairobi. World Conservation Monitoring Centre: Cambridge.

Islam, M.A., Thenkabail, P.S., Kulawardhana, R.W., Alankara, R., Gunasinghe, S., Edussriya, C. & Gunawardana, A. 2008. Semi-automated methods for mapping wetlands using Landsat ETM+ and SRTM data, International Journal of Remote Sensing, vol. 29(24), pp. 7077-7106.

Junk, W.J. 2002. Long-term environmental trends and the future of tropical wetlands", Environmental Conservation, vol. 29(4), pp. 414-435.

Jung, H.C., Hamski, J., Durand, M., Alsdorf, D., Hossain, F., Lee, H., Azad Hossain, A.K.M., Hasan, K., Khan, A.S. & Zeaul Hoque, A.K.M. 2010. Characterization of complex fluvial systems using remote sensing of spatial and temporal water level variations in the Amazon, Congo, and Brahmaputra rivers, Earth Surface Processes and Landforms, vol. 35(3), pp. 294-304.

Keddy, P.A., Fraser, L.H., Solomeshch, A.I., Junk, W.J., Campbell, D.R., Arroyo, M.T.K. & Alho., C.J.R., 2009. Wet and wonderful: The world's largest wetlands are conservation priorities. BioScience, vol. 59(1), pp. 39-51.

Kirkby, M.J., Hydrograph modelling strategies, 1975. Progress in Physical and Human Geography. Edited by: R.F. Peel, M.D. Chisholm, and P. Haggett, Heinemann, London, pp. 69–90.

Lehner, B. & Döll, P., 2004. Development and validation of a global database of lakes, reservoirs and wetlands. Journal of Hydrology, vol. 296(1-4), pp.1-22.

Lehner, B., Verdin, K. & Jarvis, A., 2008. New Global Hydrography Derived From Spaceborne Elevation Data. Eos, vol. 89(10), pp.93-93.

Lieth H 1975. Primary production of the major vegetation units in the world. In Primary productivity of the biosphere, eds H Lieth & RH Whittaker, Ecological Studies 14, Springer, New York, Heidelberg, Berlin, pp. 203–215.

Loveland, T.R., Reed, B.C., Brown, J.F., Ohlen, D.O., Zhu, J., Yang, L., Merchant, J.W., 2000. Development of a global land cover characteristics database and IGBP DISCover from 1-km AVHRR data. International Journal of Remote Sensing, vol. 21(6-7), pp. 1303–1330.

Lowry, J., Hess, L., Rosenqvist, A., 2009. Mapping and Monitoring Wetlands Around the World Using ALOS PALSAR: The ALOS Kyoto and Carbon Initiative Wetlands Products, Innovations in Remote Sensing and Photogrammetry, Springer Berlin Heidelberg, pp. 105-120

MacKay, H., Finlayson, C.M., Fernández-Prieto, D., Davidson, N., Pritchard, D. & Rebelo, L.-M. 2009. The role of Earth Observation (EO) technologies in supporting implementation of the Ramsar Convention on Wetlands, Journal of environmental management, vol. 90(7), pp. 2234-2242.

Manfreda, S., Sole, A., Fiorentino, M., 2008. Can the basin morphology alone provide an insight into floodplain delineation? Flood Recovery, Innovation and Response I, vol. 118, pp.47-56.

Margineantu, D., & Dietterich, T. G. 2001. Improved class probability estimates from decision tree models. In C. Holmes (Ed.), Nonlinear Estimation and Classification. The Mathematical Sciences Research Institute, University of California, Berkeley.

Matthews E., Fung I., 1987. Methane emission from natural wetlands: Global distribution, area, and environmental characteristics of sources. Global Biogeochemical Cycles, vol. 1(1), pp. 61–86.

Matthews, G. V. T. 1993. The Ramsar Convention; its History and Development. Ramsar Convention Bureau, Gland, Switzerland.

Mayaux, P., De Grandi, G.F., Rauste, Y., Simard, M. & Saatchi, S., 2002. Large-scale vegetation maps derived from the combined L-band GRFM and C-band CAMP wide area radar mosaics of Central Africa, International Journal of Remote Sensing, vol. 23(7), pp. 1261-1282.

Mayaux, P., Bartholomé, E., Fritz, S. & Belward, A., 2004. A new land-cover map of Africa for the year 2000", Journal of Biogeography, vol. 31(6), pp. 861-877.

McCartney, M.P. & Vladimir, S., 2010. Blue paper - Water Storage in an Era of Climate Change.

Melack, J.M. & Hess, L.L., 2004. Remote Sensing of Wetlands on a Global Scale. SIL news, 42(May), pp.1-12.

Mialon, A., Royer, A. & Fily, M. 2005. Wetland seasonal dynamics and interannual variability over northern high latitudes, derived from microwave satellite data, Journal of Geophysical Research, vol. 110(17), pp. 11-19.

Mitsch, W.J. and Gosselink. J.G., 2007. Wetlands, 4th edition, Inc., New York, 582 pp.

Moghaddam, M., McDonald, K., Cihlar, J. & Chen, W. 2003. Mapping Wetlands of the North American Boreal Zone from Satellite Radar Imagery, International Geoscience and Remote Sensing Symposium (IGARSS), pp. 261.

Murray-Hudson, M., Wolski, P. & Ringrose, S., 2006. Scenarios of the impact of local and upstream changes in climate and water use on hydro-ecology in the Okavango Delta, Botswana. Journal of Hydrology, vol. 331(1-2), pp.73-84.

Naiman, R.J. & Dudgeon, D. 2011, Global alteration of freshwaters: Influences on human and environmental well-being, Ecological Research, vol. 26(5), pp. 865-873.

Neiland A.E., Bene C., 2008. Tropical river fisheries valuation: background papers to a global synthesis. The WorldFish Center, Penang, Malaysia.

Nel, J.L., Roux, D.J., Abell, R., Ashton, P.J., Cowling, R.M., Higgins, J.V., Thieme, M. & Viers, J.H. 2009. Progress and challenges in freshwater conservation planning, Aquatic Conservation: Marine and Freshwater Ecosystems, vol. 19(4), pp. 474-485.

Ordoyne, C. & Friedl, M.A. 2008. Using MODIS data to characterize seasonal inundation patterns in the Florida Everglades, Remote Sensing of Environment, vol. 112(11), pp. 4107-4119.

Papa, F., Prigent, C., Aires, F., Jimenez, C., Rossow, W. B. & Matthews, E., 2010. Multi-Satellite Remote Sensing of Global Surface Waters Extent, 1993-2004. Journal of Geophysical Research, vol. 115(D12111), pp.1-17.

Portmann, F.T., Siebert, S. & Döll, P., 2010. MIRCA2000-Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. Global Biogeochemical Cycles, vol. 24(1), pp.GB1011.

Prigent, C., Papa, F., Aires, F., Rossow, W.B. & Matthews, E., 2007. Global inundation dynamics inferred from multiple satellite observations, 1993-2000. Journal of Geophysical Research, vol. 112(12).

Provost, F. & Domingo, P., 2003. Tree Induction for Probability-based Ranking. Machine Learning 52:3

Rebelo, L.-M, McCartney, M.P. & Finlayson, M., 2010. The application of geospatial analyses to support an integrated study into the ecological character and sustainable use of Lake Chilwa. Journal of Great Lakes Research.

Rebelo, L.-M., McCartney, M.P. & Finlayson, M., 2009. Wetlands of Sub-Saharan Africa: distribution and contribution of agriculture to livelihoods. Wetlands Ecology and Management.

Richey, J.E., Melack, J.M., Aufdenkampe, A.K., Ballester, V.M. & Hess, L.L. 2002, Outgassing from Amazonian rivers and wetlands as a large tropical source of atmospheric CO2, Nature, vol. 416(6881), pp. 617-620.

Rebelo, L.M., Senay, G.B. & McCartney, M.P. 2012, Flood pulsing in the Sudd wetland: Analysis of seasonal variations in inundation and evaporation in South Sudan, Earth Interactions, vol. 16(1), pp. 1-19.

Ringeval, B., De Noblet-Ducoudré, N., Ciais, P., Bousquet, P., Prigent, C., Papa, F. & Rossow, W.B. 2010. An attempt to quantify the impact of changes in wetland extent on methane emissions on the seasonal and interannual time scales, Global Biogeochemical Cycles, vol. 24(2).

Robinson, C.T., Tockner, K. & Ward, J.V. 2002. The fauna of dynamic riverine landscapes, Freshwater Biology, vol. 47(4), pp. 661-677.

Rosenqvist, A. & Birkett, C.M., 2002. Evaluation of JERS-1 SAR mosaics for hydrological applications in the Congo river basin. International Journal of Remote Sensing, vol. 23(7), pp.1283-1302.

Rosenqvist, A., Shimada, M., Chapman, B., McDonald, K., De Grandi, G., Jonsson, H., Williams, C., Rauste, Y., Nilssdn, M., Sango, D. & Matsumoto, M. 2004. An overview of the JERS-1 SAR global boreal forest mapping (GBFM) project, International Geoscience and Remote Sensing Symposium (IGARSS), pp. 1033.

Rosenqvist, A., Finlayson, M., Lowry, J.B. & Taylor, D., 2007. The potential of long-wavelength satellite-borne radar to support implementation of the Ramsar Wetlands Convention, 244, pp.229-244.

Schroeder, R., Rawlins, M.A., McDonald, K.C., Podest, E., Zimmermann, R. & Kueppers, M. 2010. Satellite microwave remote sensing of North Eurasian inundation dynamics: Development of coarse-resolution products and comparison with high-resolution synthetic aperture radar data, Environmental Research Letters, vol. 59(1).

Scott, D. & Jones, T., 1995. Classification and inventory of wetlands: A global overview. Plant Ecology, vol. 118(1), pp.3–16.

Seiler, R., Schmidt, J., Diallo, O. & Csaplovics, E. 2009, Flood monitoring in a semi-arid environment using spatially high resolution radar and optical data, Journal of Environmental Management,vol. 90(7), pp. 2121-2129.

Siebert, S., J. Hoogeveen, and K. Frenken. 2006. Irrigation in Africa, Europe and Latin America - Update of the Digital Global Map of Irrigation Areas to Version 4. Frankfurt Hydrology Paper 05. Institute of Physical Geography, University of Frankfurt, Frankfurt am Main, Germany and Food and Agriculture Organization of the United Nations, Rome, Italy.

Simard, M., De Grandi, G., Saatchi, S. & Mayaux, P. 2002. Mapping tropical coastal vegetation using JERS-1 and ERS-1 radar data with a decision tree classifier", International Journal of Remote Sensing, vol. 23(7), pp. 1461-1474.

Sippel, S.J., Hamilton, S.K., Melack, J.M. & Novo, E.M.M. 1998. Passive microwave observations of inundation area and the area/stage relation in the Amazon River floodplain, International Journal of Remote Sensing, vol. 19(16), pp. 3055-3074.

Slater, J.A., Garvey, G., Johnston, C., Haase, J., Heady, B., Kroenung, G. & Little, J. 2006. The SRTM data "finishing" process and products, Photogrammetric Engineering and Remote Sensing, vol. 72(3), pp. 237-247.

Spiers, A., 2001. Wetland inventory: Overview at a global scale. In Endorsements for the conference were received from the Convention on Biological Diversity, the Convention to Combat Desertification, the Convention on the Conservation of Migratory Species of Wild Animals, the Ramsar Convention on Wetlands, the UN Economi. Citeseer, p. 23.

Spiers A.G. 2001. Wetland inventory: Overview at a global scale. Wetland inventory, assessment and monitoring: Practical techniques and identification of major issues. Proceedings of Workshop 4 - 2nd International Conference on Wetlands and Development, Dakar, Senegal, 8-14 November 1998

Springate-Baginski, O., Allen, D. & Darwall, W., 2009. An Integrated Wetland Assessment Toolkit: A guide to good practice, IUCN

Stillwell-Soller, L.M., Klinger, L.F., Pollard, D., Thompson, S.L, 1995. The Global Distribution of Freshwater Wetlands. NCAR Technical Note TN-416 STR, National Center for Atmospheric Research, Boulder, CO.

Straumann, R.K. & Purves, R.S., 2008. Delineation of Valleys and Valley Floors, pp.320-336.

Tarboton, D. G., Bras, R. L., & Rodriguez-Iturbe, I., 1992. A physical basis for drainage density. Geomorphology, vol. 5, pp. 59−76.

Taylor ARD, Howard GW, Begg GW (1995) Developing wet- land inventories in southern Africa: a review. In: Classifi- cation and Inventory of the World's Wetlands, Finlayson CM, van der ValkAG(eds) Advances in vegetation science 16. Kluwer Academic Publishers, Dordrecht, pp 57–79

Tockner, K. & Stanford, J.A. 2002, Riverine flood plains: Present state and future trends, Environmental Conservation, vol. 29(3), pp. 308-330.

U.S. Geological Survey , 2000, HYDRO1k Elevation Derivative Database, Cent. for Earth Resource. Observation and Sciences, Sioux Falls, S. D.

Vancutsem, C., Pekel, J., Evrard, C., Malaisse, F. & Defourny, P. 2009. Mapping and characterizing the vegetation types of the Democratic Republic of Congo using SPOT VEGETATION time series, International Journal of Applied Earth Observation and Geoinformation, vol. 11(1), pp. 62-76.

Ward, J.V., Tockner, K. & Schiemer, F. 1999. Biodiversity of floodplain river ecosystems: Ecotones and connectivity, River Research and Applications, vol. 15(1-3), pp. 125-139.

Ward, J.V. & Stanford, J.A., 1995. Ecological connectivity in alluvial river ecosystems and its disruption by flow regulation. Regulated Rivers, vol. 11, pp.105-120.

Whitcomb, J., Moghaddam, M., McDonald, K., Kellndorfer, J. & Podest, E. 2009, Mapping vegetated wetlands of Alaska using L-band radar satellite imagery, Canadian Journal of Remote Sensing, vol. 35(1), pp. 54-72.

Whittaker RH & Likens GE 1975. The biosphere and man. In Primary productivity of the biosphere (Ecological Studies 14), eds H Lieth & RH Whittaker, Springer, New York & Heidelberg, Berlin, pp. 305–328.

Wilson, M.D., Bates, P., Alsdorf, D., Forsberg, B., Horritt, M., Melack, J., Frappart, F. & Famiglietti, J. 2007. Modeling large-scale inundation of Amazonian seasonally flooded wetlands, Geophysical Research Letters, vol. 34(15).

Wood, A.P. & Halsema, G.E.V., 2008. Scoping agriculture – wetland interactions: Towards a sustainable multiple-response strategy, Rome.