The Global Wetland Extent: Towards a high-resolution global-level inventory of the spatial extent of vegetated wetlands

Contents

Acknowledgement: The Data on wetlands was produced by DHI GRAS on behalf of the United Nations Environment Programme. The European Commission and ESA provided the Sentinel data. The USGS and NASA provided the Landsat imagery. Google provided the Google Earth Engine.

Citation: Tottrup, C., Druce[1], D., Tong, X[1]., Barvels, E. [1], Christensen, M. [1], Grogan, K. [1], Huber, S. [1] & Crane, S.[2] (2020): The Global Wetland Extent: Towards a high-resolution global-level inventory of the spatial extent of vegetated wetlands. UN Environment, Nairobi, Kenya.

Authors affiliation: [1]DHI GRAS, Agern Allé , 2970 Hoersholm, Denmark; [2] UNEP, Nairobi, Kenya.

Corresponding author: Christian Tottrup: cto@dhigroup.com

Introduction

Over 1 billion people rely entirely on the services provided by wetland ecosystems. Healthy and functional natural wetlands are intrinsically linked with human livelihoods, well-being and sustainable development. However, wetlands are facing major threats, caused by conversion for agriculture or other commercial development, overfishing, tourism, pollution and climate change, just to name a few. Thus, there is an urgent need to strengthen, and reinforce, national policies and legal frameworks to help countries to protect and restore critical wetland ecosystems. Past efforts, however, have been hampered by the lack of data on the locations, types and sizes of wetland resources. This kind of data and information is crucial to measure the effectiveness of policy, legal and regulatory mechanisms and essential for tracking progress towards the Sustainable Development Goals.

As part of the 2030 Agenda for Sustainable Development, UN Member States committed to achieving Sustainable Development Goal (SDG) 6, ensuring water and sanitation for all. Within SDG 6, target 6.6 specifically aims to protect and restore water-related ecosystems, including mountains, forests, wetlands, rivers, aquifers, and lakes. Monitoring progress on target 6.6 is based on one global indicator, indicator 6.6.1, "Change in the extent of water-related ecosystems over time", which can provide the necessary data for countries to take action. Water-related ecosystems include five categories: vegetated wetlands, rivers and estuaries, lakes, aquifers, and artificial waterbodies.

Despite the importance of wetlands, and unlike other critical ecosystems (cf. forests, mangroves and inland water bodies), the extent and dynamics of wetland ecosystems remains ill defined, characterized and modeled. The global map of vegetated wetlands as presented in this paper is an attempt to bridge this data gap by developing a consistent wetland monitoring mechanism based on satellite Earth Observation data and envisioned to become the blueprint for future routine monitoring of vegetated wetlands.

The dynamics and large diversity of wetlands is a challenge for global scale mapping, and as of today, only coarse resolution data exists on a global level. However, the recent availability of the Sentinel missions from the Copernicus programme of the European Commission now provides global satellite imagery at the required temporal and spatial resolutions to derive information on wetland extent at relatively high spatial resolution. In addition, the Copernicus data streams are free and open to the public and covered by an operational service guarantee for many years to come, hence providing a strong incentive for using such data in combination with data from long-term archives (e.g., NASA/USGS Landsat missions) for wetland inventorying.

This document presents the results as well as the data and methods behind the first high-resolution global-level inventory of the spatial extent of vegetated wetlands. The primary objective of the inventory has been to provide countries lacking a national wetland inventory with a first cut extent map of vegetated wetlands which in cohort with other geospatial datasets (e.g. distribution of open water surfaces) can be used as the baseline for monitoring progress towards SDG target 6.6.

The mapping included the entire land surface except for Antarctica and a few small islands. Only vegetated wetlands were mapped according to the following definition: "Vegetated wetlands include areas of marshes, peatlands, swamps, bogs and fens, the vegetated parts of flood plains as well as rice paddies and flood recession agriculture". Close to 4 million satellite images amounting to 2.8 petabyte of data were analyzed and classified as wetland or non-wetland using an automated machine learning model implemented in Google Earth Engine (GEE) - a cloud platform that provides the necessary combination of large-scale computing resources with instant access to planetary scale archives of the required input satellite imagery.

Results

The classified global vegetated wetland extent at 10x10-meter pixel resolution is shown in Figure 1. Continental and national wetland extents can also be extracted as well as more detailed subsets. For instance, Figure 2 and Figure 3 show the wetland extents for Sub-Saharan Africa and Uganda, respectively, while **Error! Reference source not found.** shows a selection of detailed subsets over widely known wetlands in the US, Europe, Africa and South America.

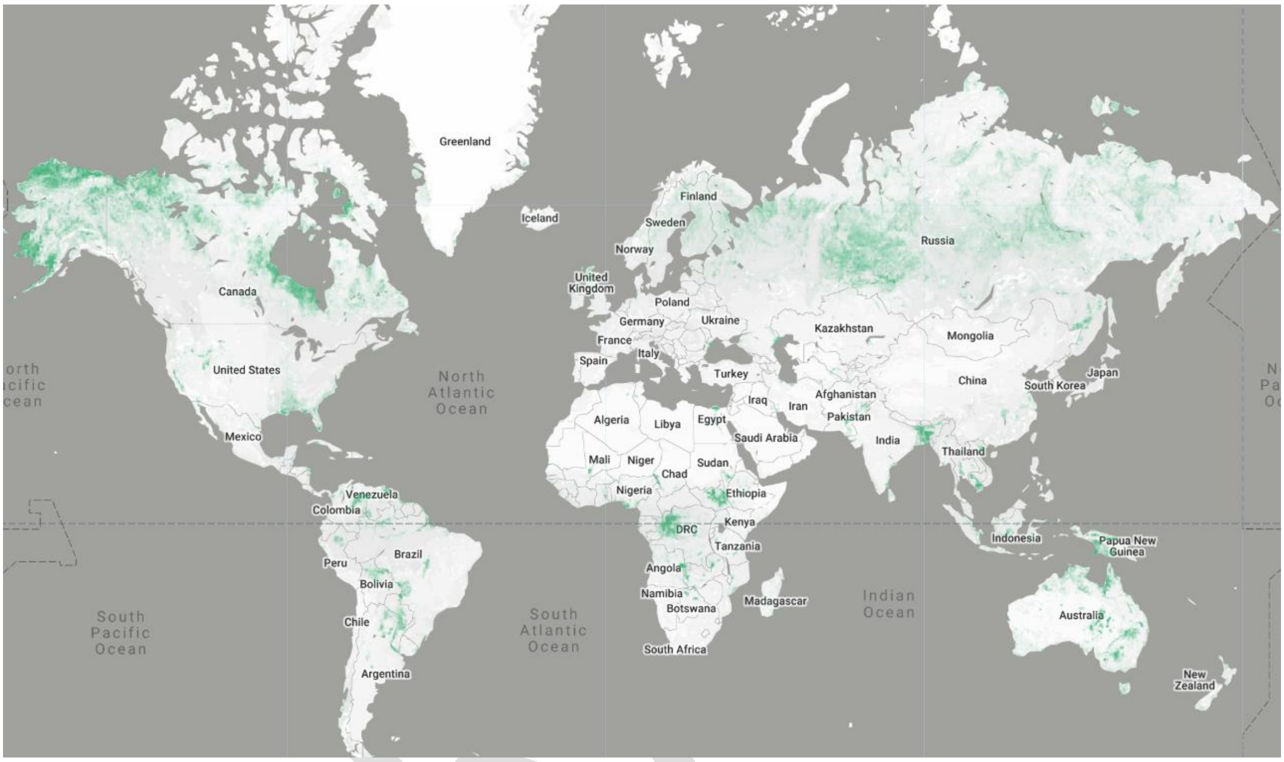

Figure 1. Global extent of vegetated wetlands at 10-meter spatial resolution.

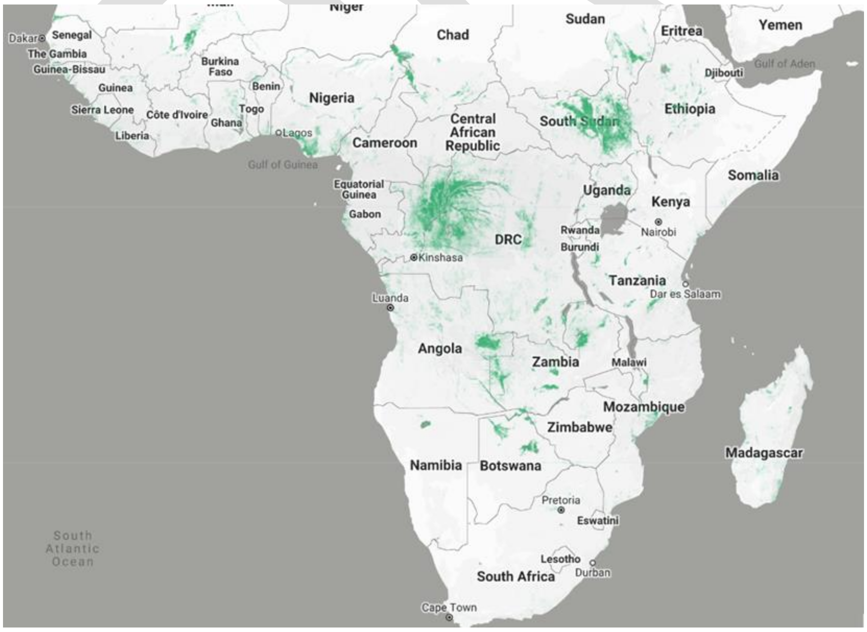

Figure 2. Vegetated wetland extent for Sub-Saharan Africa.

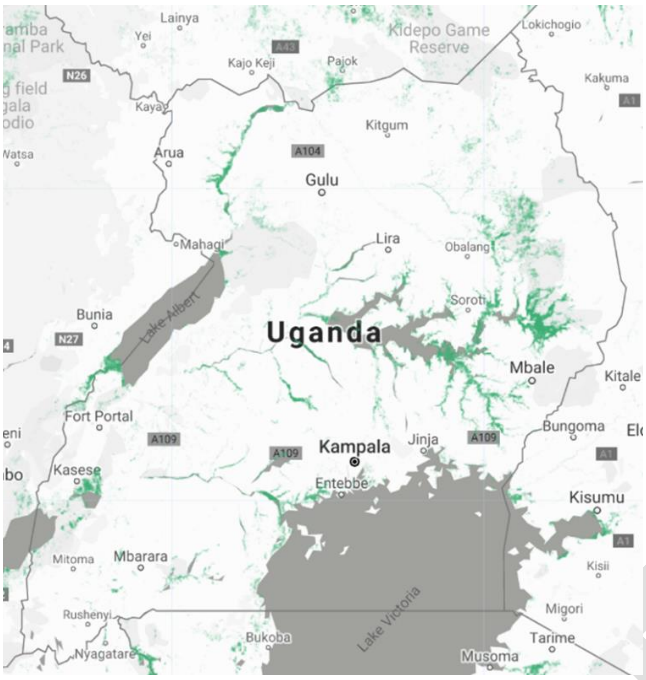

Figure 3. Vegetated wetland extent for the territory of Uganda.

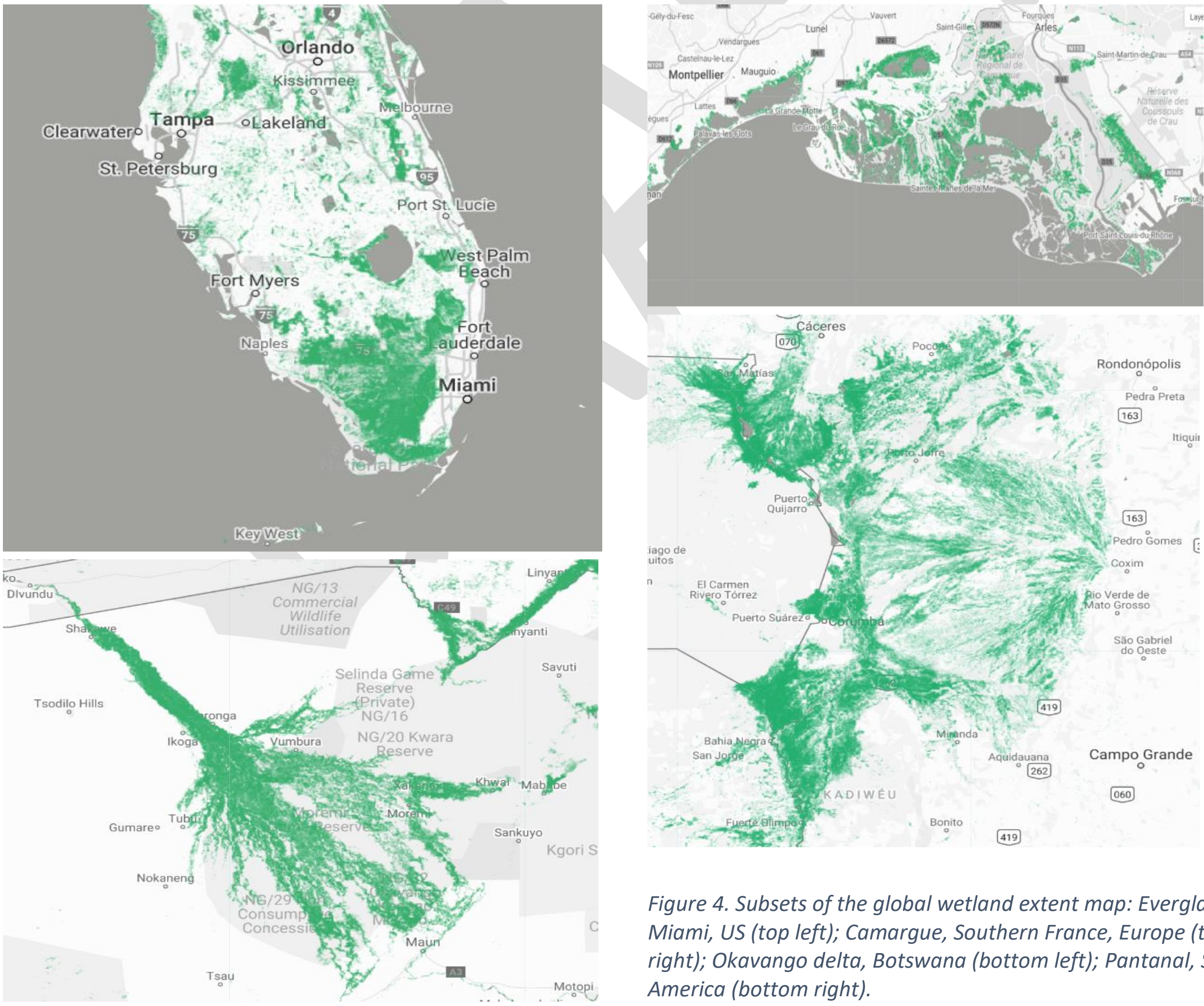

Figure 4. Subsets of the global wetland extent map: Everglades, Miami, US (top left); Camargue, Southern France, Europe (top right); Okavango delta, Botswana (bottom left); Pantanal, South America (bottom right).

Data overview, pre-processing and image reduction

For the global wetland mapping, satellite imagery from Sentinel-1 and 2 (radar and optical) together with Landsat 8 (thermal) was analyzed. This information was complemented with topographic data from satellite-derived Digital Elevation Models (DEMs).

As wetlands tend to be susceptible to high inter-annual variations, multi-annual time-series were used to even out potential annual biases and to create a more solid estimate of the wetland extent around the baseline year of 2017. All available data from 2016 to 2018 were gathered from the three different satellites: Sentinel-1 (SAR), Sentinel-2 (optical) and Landsat 8 (thermal):

- Sentinel-1 is composed of a constellation of two near-polar satellites (A and B), which share the same orbit plane with a 180° orbital phasing difference. Both satellites carry a C-band synthetic-aperture radar (SAR) instrument which collects data in all-weather conditions, day or night. Sentinel-1 time series are a major asset to wetland classification as they relate to structural characteristics of wetlands including size, density, orientation of vegetation as well as to soil moisture content. Sentinel-1's all-weather capability and 12-day repeat cycle ensures that wetland monitoring can be continuously performed even under cloudy conditions.
- Sentinel-2 comprises a constellation of two polar-orbiting satellites (A and B) placed in the same sun-synchronous orbit, phased at 180° to each other. Both satellites carry an optical instrument payload that samples 13 spectral bands providing wide swath imagery (290 km) every 5-day. The large footprint size, the spectral depth extending from the visible over the near and mid infrared spectral region along with its short revisit time allows rapid changes in ecosystems to be precisely monitored and is ideally suited to monitor dynamic habitats such as wetlands.
- The Landsat 8 satellite collects images of the entire Earth every 16 days. In this study we used data collected with the Thermal InfraRed Sensor (TIRS) which measures thermal infrared radiation in two bands. Thermal imagery is a less explored source for wetland mapping, yet an interesting additional information as land surface temperature is closely related to the surface energy balance and the wetness status. Evapotranspiration (ET) is the primary energy loss mechanism for wetlands and due to the higher moisture content, ET tends to be higher and hence temperature lower in wetlands than in their surroundings.
- Satellite derived DEMs from the Shuttle Radar Topographic Mission (SRTM)[1] (below 60°N) and 90 m Viewfinder Panoramas[2] DEM (above 60°N) were obtained and used for topographic post-processing. The formation of wetlands is dependent on water, which in turn is constrained by topography. This is because topography determines the direction, flow and storage of water. Thus DEMs can be utilized to focus wetland mapping to areas with potential wetland occurrence.

Image pre-processing

Before conducting the image analyses, essential pre-processing of the acquired data was conducted. The basic outline of the processing steps is given in this section.

In this study we used **Sentinel-2** Level-2 Top of the Atmosphere reflectance data as available on GEE. The Multi-Spectral Instrument (MSI) onboard Sentinel-2 acquires 13 spectral bands ranging from visible and near-infrared (VNIR) to shortwave infrared (SWIR) wavelengths and a spatial resolution of 10 m (four visible

[1] Jarvis, A., Reuter, H.I., Nelson, A. & Guevara, E. (2008). Hole-filled SRTM for the globe Version 4, available from the CGIAR-CSI SRTM 90m Database: http://srtm.csi.cgiar.org.
[2] http://viewfinderpanoramas.org/dem3.html

and near-infrared bands), 20 m (six red edge and shortwave infrared bands) and 60 m (three atmospheric correction bands) (cf. Figure 5)

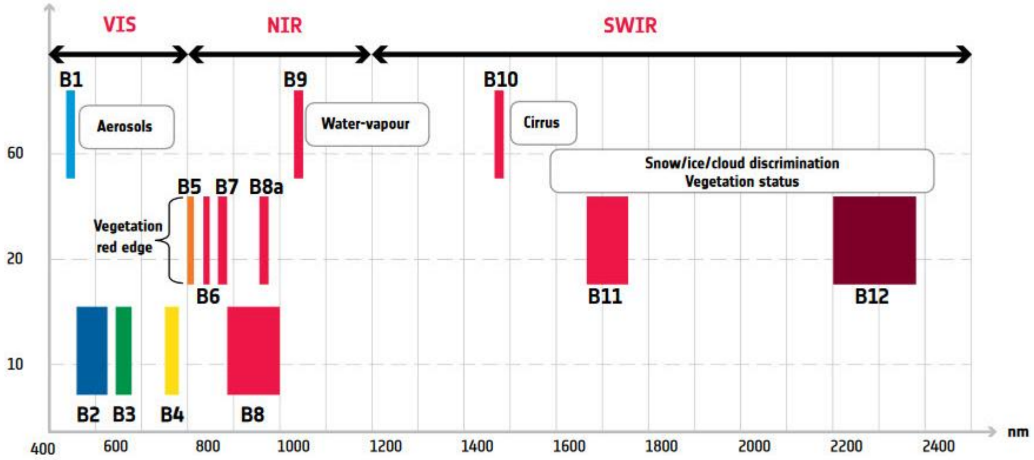

Figure 5. Spatial resolution versus wavelength of the Multi-Spectral Instrument (MSI) onboard Sentinel-2 (source: ESA).

All the Sentinel-2 bands with 10- and 20-meter spatial resolution were used in the analysis. The 20 m bands were resampled to 10 meters, and in order to remove the cloudiest images from the time series, only images with a cloud cover less than 30% (according to internal quality flags) were retained. Masking of all remaining clouds incl. cirrus clouds and cloud shadows as well as snow was done with FMASK[3].

Thereafter, various spectral indices were calculated with different combinations of the Sentinel-2 reflectance bands. Spectral indices can be useful to highlight specific properties relevant for wetland detection, such as vegetation presence and vigor as well as plant water content. Listed below are the key spectral indices used for the wetland mapping (cf. Table 1)

Table 1. Sentinel-2 spectral indices used for wetland extent prediction.

Index	Short name	Equation [Sentinel-2 bands]
Normalized Difference Vegetation Index	NDVI	("B8") - ("B4") / ("B8") + ("B4")
Visible Atmospherically Resistant Index	VARI	("B3") - b("B4") / ("B3") + b("B4") - ("B2")
Red-edge NDVI	rNDVI	("B6") - ("B4") / ("B6") + ("B4")
Normalized Difference Water Index	NDWI	("B8") - ("B11") / ("B8") + ("B11")
Modified Normalized Difference Water Index	mNDWI	("B11") - ("B3") / ("B11") + ("B3")
Normalized Multi-band Drought Index	NMDI	("B11") - ("B12") - ("B8") / ("B11") - ("B12") + ("B8")
Normalized Difference NIR - SWIR2	ND0812	("B12") - b("B8") / ("B12") + ("B8")
Normalized Difference GREEN - SWIR2	ND0312	("B12") - ("B3") / ("B12") + ("B3")
Normalized Difference SWIR1 - SWIR2	ND1112	("B12") - ("B11") / ("B12") + ("B11")

From the **Sentinel-1** dataset, the Level-1 Informetric Wide Swath (IW) and Ground Range Detected (GRD) data was used. Within GEE, the data was accessed and processed to generate a calibrated, ortho-corrected product at 10-meter spatial resolution (to match Sentinel-2). Finally, the VV and VH backscattering values

[3] Zhu, Z., Wang, S., & Woodcock, C. E. (2015). Improvement and expansion of the Fmask algorithm: Cloud, cloud shadow, and snow detection for Landsats 4–7, 8, and Sentinel 2 images. Remote Sensing of Environment, 159, 269-277.

were used for the wetland classifications, except north of 66°N where the Sentinel-1 polarization is shifted to HH/HV.

Landsat thermal imagery comes from the USGS Landsat 8 Surface Reflectance Tier 1 collection on GEE. This collection is atmospherically corrected using LaSRC (Land Surface Reflectance Code) and masked to remove cloud, shadow and snow using the CFMASK. An additional filter was applied to remove scenes with a cloud cover percentage above 50 %. Only the brightness temperature datasets from Band 10 (B10) were retained for the subsequent analysis.

Since wetland formation is highly dependent on the topographic conditions within a given region, terrain analyses on SRTM and Viewfinder Panoramas **DEMs** was applied to identify areas with potential suitability for wetland formation[4]. In the analysis the Height Above Nearest Drainage (HAND) was used, which is a hydrologically relevant terrain model. The HAND model normalizes topography according to the local relative heights found along the drainage network, and in this way, presents the topology of the relative soil gravitational potentials, or local draining potentials[5]. It has been demonstrated that the HAND model is highly correlated with the depth of the water table, providing an accurate spatial representation of soil water environments, and hence potential wetland formation[6].

Image reduction

From a remote sensing perspective, wetlands are complex phenomena with several unique challenges. For example, wetlands typically change within and between years according to varying hydrological conditions. As wetlands tend to be susceptible to high inter and intra-annual variations, temporal information is key to differentiate wetlands from other land cover types. In view of the above, the increased temporal repeat frequencies of the Copernicus Sentinel missions are ideal to capture unique salient points of vegetation phenology. On the other hand, due to this high repeat frequency huge amounts of information are accumulated. For the three satellites used in this study, Sentinel-1, Sentinel-2 and Landsat 8, the number of scenes acquired on a global level during the 2016-2018 monitoring period is approximately 580,000, 3 million and 240,000, respectively, or in total 2.8 peta byte of data. Therefore, image reduction methods that translate big data into smart smaller data are recommended to ease data handling and increase processing speed. Many studies have documented how multi-temporal metrics, such as minimum, maximum and amplitude, can improve land cover separability and provide an unbiased measure of phenological activity, irrespective of regional variations in the seasonal cycle. Hence, annual metrics are more robust for mapping at a global scale across climate regions and across the southern/northern hemisphere. For the global wetland mapping, the minimum, maximum, mean, median, standard deviation and selected percentile values (10, 15, 50, 85 and 95% percentiles), for each year over the three-year monitoring period (i.e. 2016 - 2018) as well as the amplitudes (p85-p15, p50-p15, p85-p50) were calculated from all the optical, SAR and thermal imagery as well as their associated indices.

---

[4] Donchyts, G., Winsemius, H., Schellekens, J. et al. (2016). Global 30 m Height Above the Nearest Drainage. Proceedings of the EGU General Assembly: http://global-hand.appspot.com/

[5] Nobre, A. D., Cuartas, L. A., Hodnett, M. et al. (2011). Height Above the Nearest Drainage–a hydrologically relevant new terrain model. Journal of Hydrology, 404(1-2), 13-29.

[6] Gharari, S., Hrachowitz, M., Fenicia, F. & Savenije, H. H. G. (2011). Hydrological landscape classification: investigating the performance of HAND based landscape classifications in a central European meso-scale catchment, Hydrol. Earth Syst. Sci., 15, 3275–3291, https://doi.org/10.5194/hess-15-3275-2011.

Mapping wetland extent

Predicting wetland extent using Earth Observation data relies on 5 key components: 1) stratification, 2) training the classification model, 3) machine learning classification, 4) post-processing and 5) classification model evaluation. The approach uses selected data (as described above) from Sentinel-1, Sentinel-2, and Landsat 8, and feeds the information into a random forest classifier to predict wetland probability. Then, a Digital Elevation Model is used to qualify wetland predictions and a post-processing routine converts the wetland probability map into a map of wetland extent (cf. Figure 6).

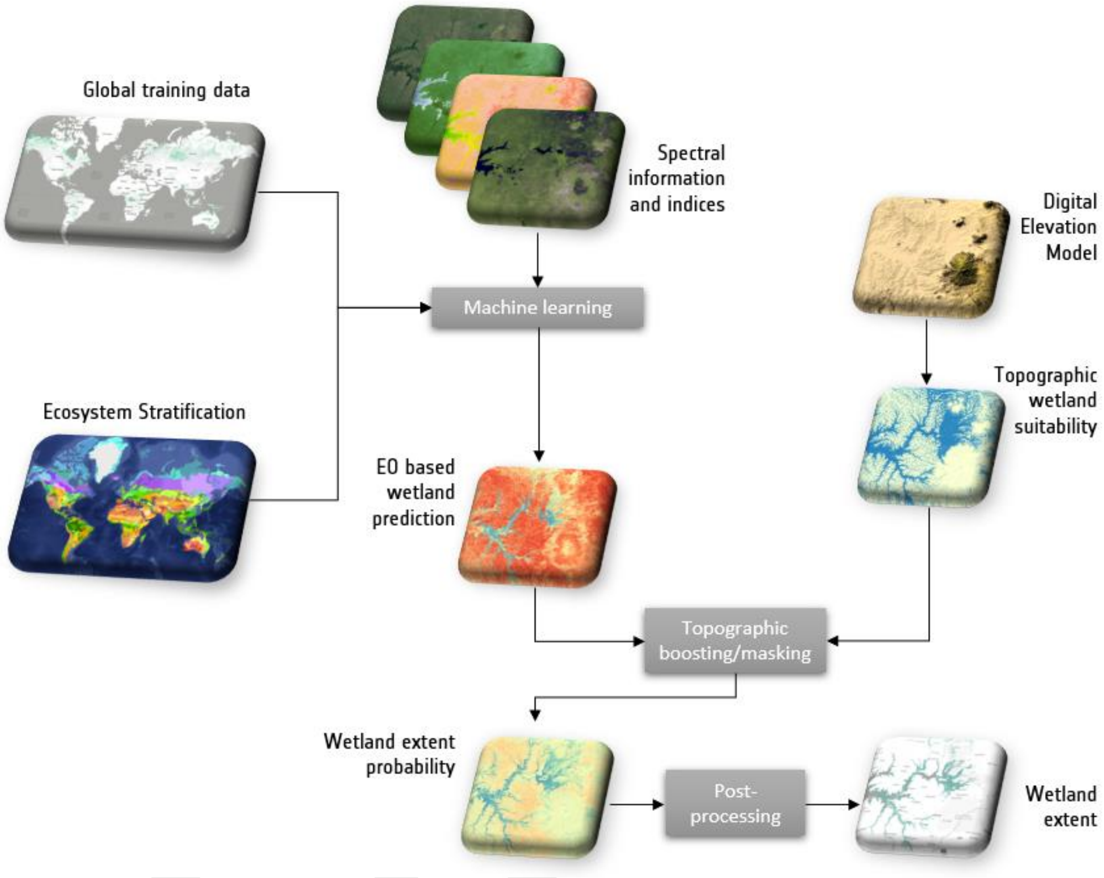

Figure 6. Workflow for mapping global wetland extent.

Stratification

Splitting the globe into meaningful regions, or strata, can help to improve the accuracy of wetland mapping. Intuitively, it makes sense that different regions have different kind of wetlands. There are also regional differences related to physical factors such as temperature, rainfall, and soil conditions, which can lead to the same wetlands, however, with different spectral and temporal satellite-measured signals. It can therefore be beneficial to stratify according to broad bio-geographical regions, so that the classification is better tuned to regional conditions. In this way, the world-wide classification is made up of multiple sub-models instead of one. A global stratification approach is a compromise between demarcation of meaningful zones and what is practically feasible. For the global wetland mapping, the RESOLVE ecoregion map was used to stratify the globe into mapping zones based on a joint realm and biome classification (Figure 7).

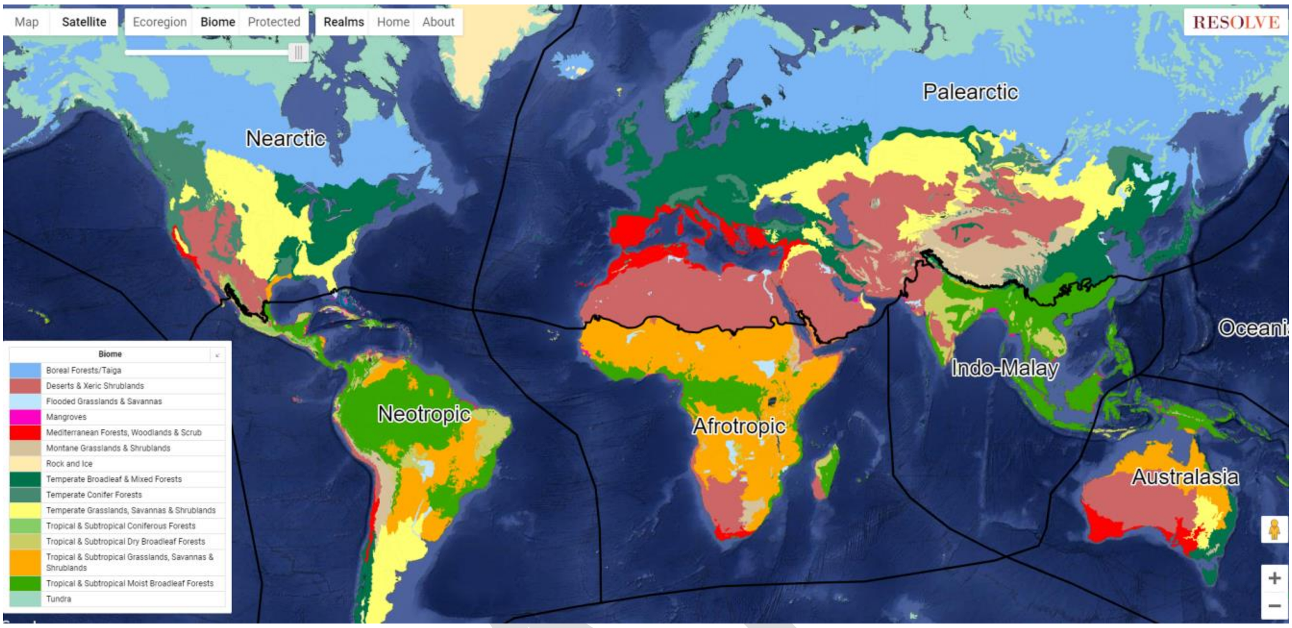

Figure 7. Mapping zones were defined by the boundaries of the 14 terrestrial biomes distributed within 7 realms in the RESOLVE ecoregions dataset (https://ecoregions2017.appspot.com/).

Training the classification model

The main requirement of a well-trained classification model is a set of training samples that represent the classes of interest. In this case, a binary (i.e. 2 classes) classification model was used which required samples of vegetated wetlands and non-wetlands locations. The training samples were compiled from existing large-area wetland/land cover maps. For the tropics and sub tropics the 2011 Global Wetlands Map[7] was used. In this map, tropical and subtropical wetlands are mapped in ~250 meters spatial resolution by combining a hydrological model and annual time series of satellite-derived estimates of soil moisture to represent water flow and surface wetness, which are then combined with geomorphological information. Despite its label, the Global Wetlands Map does not cover the region north of 40°N. Therefore, training data for the northern latitudes was supplemented with wetland/non-wetland location information from the CORINE Land Cover for 2018[8] and Copernicus Global Land Cover for 2015[9]. The two latter maps are not wetland maps per se, and it is well known that more generalized land cover maps tend to underestimate wetland extent because the adapted nomenclatures do not contain many wetland-relevant classes. This means a significant part of wetland habitats are merged within broader defined land cover classes. For the northern region, the training samples were therefore augmented with manually derived annotations of wetlands. Finally, high resolution global maps of urban[10] as well as surface water extent[11] were used for improving the sampling and hence the prediction of wetlands in the vicinity of open water and urban areas.

---

[7] https://www.cifor.org/global-wetlands/
[8] https://land.copernicus.eu/pan-european/corine-land-cover/clc2018
[9] https://lcviewer.vito.be/
[10] Pesaresi, M., Ehrilch, D., Florczyk, A. J. et al. (2015). GHS built-up grid, derived from Landsat, multitemporal (1975, 1990, 2000, 2014). European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/jrc-ghsl-ghs_built_ldsmt_globe_r2015b
[11] Pekel, J. F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. Nature, 540(7633), 418.

A systematic random sampling approach was used to the generate the training data for the model prediction. First, the entire globe was divided into 1x1 degree grids and within each grid a proportional random selection process was used to locate and demarcate areas of interest (AOI) for wetland and non-wetland classes. Each AOI was 4 km^2 and within each AOI pixel-wise sampling was used to collect training data for the wetland and non-wetland classes (cf. Figure 8).

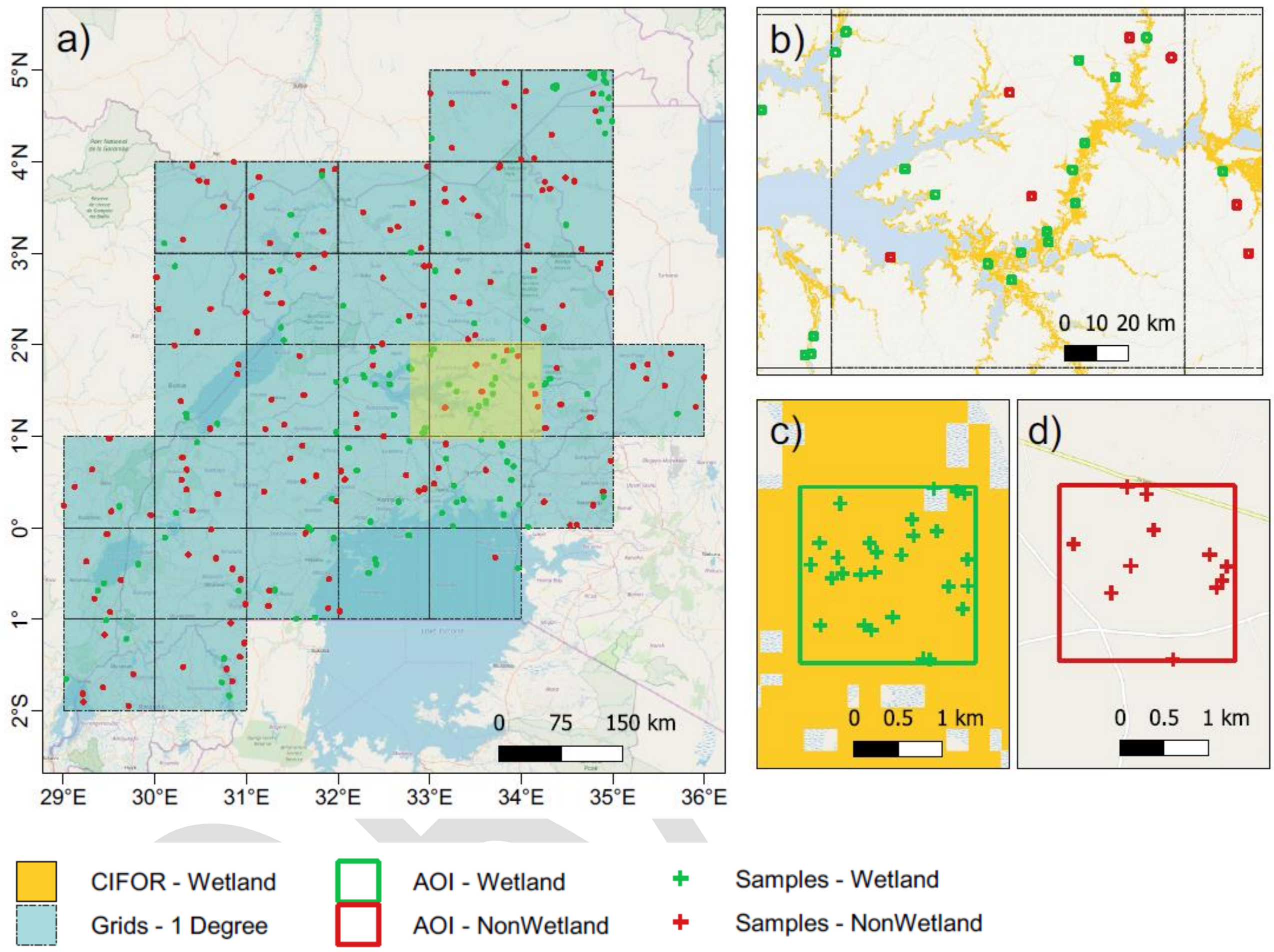

Figure 8. Overview of sampling strategy: a) Samples are collected systematically within 1x1 degree grids; b) within each 1x1 degree grid stratified sampling is used to locate areas of interest (AOIs) i.e. 2x2 km^2 squares representing both wetland and non-wetland locations and proportional to their presence; c) pixel-wise sampling for wetlands within an AOI and d) pixel-wise sampling for non-wetlands within an AOI.

The training sample had inherent noise due to the imperfect maps, the coarse spatial resolution and the method of selecting training samples. Therefore, an outlier removal algorithm was used to remove data that, with a large probability, did not belong to the training sample. To avoid excessive noise, the outlier detection was based on an estimation of the most important features. The training data was fed into the random forest classifier model, which included a parameter optimization to ensure optimal settings for prediction.

Machine learning classification

Random forest (RF) is based on the principles of decision tree classification, but instead of relying on one single tree, it creates hundreds of decision trees using random subsets of both the input variables and the training data, making each tree unique. Each tree is created by taking a random subset of samples (cf. training data). At each node of the tree, a random subset of input variables is chosen. Then the tree is split into branches based on the variable that generated the best split. This splitting continues until all the samples reside in pure leaf nodes. The variable that generates the best split is the one that minimizes the sum of the Residual Sum of Squares (RSS) error from the left and right branches:

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

Where y_i is the i^{th} value of the variable to be predicted, and $\hat{y}_i$ is the predicted value of y_i.

When given a test sample, each tree makes a prediction, and the prediction with the most votes among all trees is the one that the model chooses. By taking this approach, it transforms the decision tree approach from a 'weak learner' to a 'strong learner', resulting in a classifier which is highly robust towards training label noise, while at the same time matching or exceeding the performance of other machine learning algorithms. Another advantage of RF is the flexibility in terms of input data. Continuous variables with completely different scales are readily accepted as combined inputs. This means the classifier is well suited for fusing SAR, optical and thermal datasets as used for wetland classification.

Even after the image reduction step, still more than 90 different input features were available. Since not all of these features will effectively contribute to the prediction and too many features may potentially even lead to overfitted prediction models, removing features of low importance can improve prediction accuracy and reduce both model complexity, overfitting and computing time (for very large datasets). Therefore, feature engineering was used to further reduce the number of features used to train the machine learning model.

Feature engineering is the process of preparing the proper input dataset and improving the performance of machine learning models. Algorithms such as decision trees automatically rank the attributes in the dataset. The top few nodes in a decision tree are considered the most important features from a predictive standpoint. This feature ranking or feature importance can be used for feature extraction or dimension reduction which is the process of selecting only those attributes which "explain" the highest amount of variance in the data while ignoring the rest. From a computational point of view, such feature selection is highly desirable, and it adds to image reduction by further optimizing the datasets for easier handling and faster processing.

The feature selection approach relied first on Pearson's correlation to find highly correlated features ($\rho \pm$ 0.7-1). Then a random forest model was applied to select the most important feature among the highly corrected features. In order to preserve the multi-dimensionality of the input features this process was performed individually per input data type, i.e. Sentinel-1 (percentiles and amplitude), Sentinel-2 (percentiles and amplitude) and Landsat 8 (percentiles). In total +50% of the input features were removed during this step. In a second step, Recursive Feature Elimination (RFE) was applied to select the most useful feature among the remaining non-correlated variables. RFE begins by building a model on the entire set of features and computes an importance score for each feature. The least important feature(s) are then removed, and the model is re-built, and importance scores are computed again. In this iterative way RFE can work out the combination or subset of features that contribute most to the prediction on the target variable (or class). The optimal subset is then used to train the final model.

Through the process of RFE it was found that the model predictions across the different strata tended to stabilize around 15 input variables, i.e. the prediction accuracy was not at all or only slightly improved by using more than 15 input features. Hence, the final classification models which provided the best compromise between model overfitting, complexity, accuracy and speed were based on a maximum of 15 features out of a total of more than 90 input features. While all the models included multiple input features for prediction, the input features were not identical for the different strata, i.e. the importance of individual input features could vary between strata. A total of 55 features were used across all models, but only one feature, *mNDWI (*Modified Normalized Difference Water Index), was used by all models. On average the top 5 most important features for the wetland prediction were mNDWI (p15), ND0812 (p15), B10 (p95), mNDWI (p50) and VV (p95) according to feature importance. This supports the multi-sensor, multi-percentile-based approach to wetland classification (Figure 9).

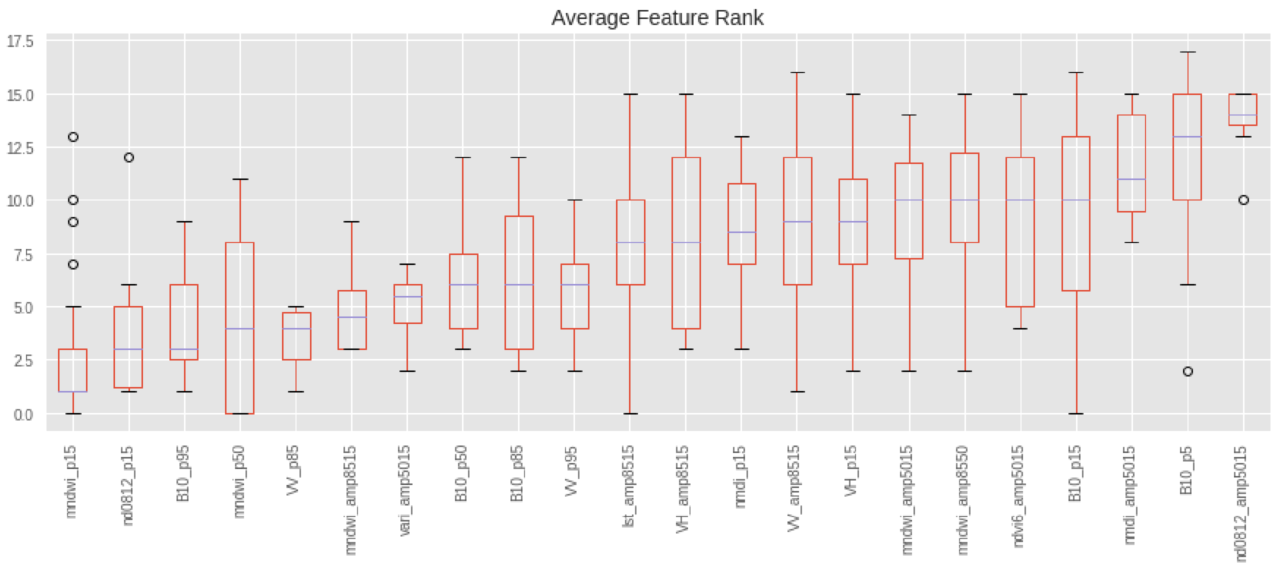

Figure 9. The average feature rank of input features across all models.

Wetland classification and post-processing

The outcome of the random forest classifier is a probability estimate (0-100%) for wetland presence in each 10x10 meter pixel. To convert this probability map into a binary wetland vs. non-wetland map a few post-processing routines were applied. As a rule of thumb, a minimum probability of 75% was required before a pixel was considered as belonging to the wetland class. However, the rule of thumb was exempted in a few cases. First, a pixel could still be classified as wetland in cases where the spectral probability was lower than 75% (but still higher than 50%) and the HAND index (scaled from 0 to 100%) was higher than 75% (*DEM boosting*). Second, wetland probabilities higher than 75% could still be disqualified if they occurred in regions with limited potential for wetland formation as identified by the HAND index (*DEM masking*). Further, post-processing included manual spectral threshold refinements and selection of best models where mapping strata overlapped. To reduce 'salt and pepper' noise and generate a cleaner map output dilation and erosion filters were applied followed by a focal filter with a 15-meter kernel to smooth edges and fill smaller gaps. In order to preserve narrow features (e.g., riverbanks), the focal filter was implemented in a way that no wetland pixels were removed. In a final step, a minimum mapping unit (MMU) filter was applied to remove solitary and smaller pixel groupings less than 25 pixels (i.e. 0.25 ha).

Model Evaluation

The performance of the classification models was initially assessed against a set of independent wetland extend maps located across Africa and derived using the GlobWetland Africa mapping approach[12]. Measures of accuracy, precision, recall and F1 scores were used. These are all measures which can be determined from the confusion matrix. The confusion matrix is a table commonly used as the basis for describing the performance of a classification model on a set of test data for which the true values are known. The confusion matrix operates with true positive and true negatives which are the observations that are correctly predicted and therefore shown in green in Table 2 below. The confusion matrix also includes false positives and false negatives which represent wrong predictions and hence displayed in red.

Table 2. Confusion matrix comparing the Global Wetland Extend Map (predicted) with independent wetland maps from GlobWetland Africa (actual).

	Predicted	
Actual	Wetland	Non-wetland
Wetland	136	76
Non-wetland	61	1307

Accuracy is the most intuitive performance measure and it is simply a ratio of correctly predicted observation to the total observations. Accuracy is a good measure of model performance if you have symmetrical datasets, but in this case, non-wetlands are much more prevalent than wetlands. It is therefore useful to also look at other parameters such as precision, recall and F-score to evaluate the model performance.

Precision is the ratio of classification model results that correctly predicted wetland locations (True Positives) to the models total predicted wetland observations, both correct (True Positives) and incorrect (False Positives). In other words, precision answers the following question: How many of the wetlands labelled by the model as wetlands are actually wetlands?

Recall, on the other hand, is the ratio of classification model results that correctly predicted wetland locations (True Positives) to all observations in the actual wetland class (Actual Positives). In other words, recall answers the following question: Of all the known wetlands, how many of those did the model correctly classify as wetlands?

F-score is the weighted average of Precision and Recall. Therefore, this score takes both false positives and false negatives into account. Intuitively, it is not as easy to understand as accuracy, but F-score is usually more useful than accuracy, especially if you have an uneven class distribution. Hence, for wetland mapping, where non-wetlands are more prevalent than wetlands, F-score might be a better measure to use as it provides a balance between Precision and Recall.

The RF machine learning models applied in this study showed an accuracy of 91% when compared to independent reference data. The individual classes of wetlands and non-wetlands were mapped with precision accuracies of 69% and 94%, respectively, while the Recall accuracies were 64% and 96% for

[12] Ludwig, C., Walli, A., Schleicher, C. et al.. (2019). A highly automated algorithm for wetland detection using multi-temporal optical satellite data. Remote Sensing of Environment, 224, 333-351.

wetlands and non-wetlands, respectively. Finally, the F-scores achieved 66% for wetlands and 95% for non-wetlands (cf.

Table 3).

Table 3. Model performance metrics for the global wetland prediction model over the African realm.

	Precision	Recall	F-score
Non-wetland	0.94	0.96	0.95
Wetland	0.69	0.64	0.66

While all the performance metrics point towards a reasonably good ability to predict wetland locations in Africa, the evaluation is limited by availability of and access to reliable and representative global reference data. Better and more qualified accuracy measures need to be developed by comparing the map to independent wetland observations in other parts of the world collected either in field or obtained from existing wetland inventories.

Known limitations and scope for improvements

Users of the global wetland map should be aware that the map represents a first line rapid assessment of the global distribution of vegetated wetlands. While it is based on a scientific sound and robust mapping approach, there will inevitable be errors in the wetland predictions both in terms of commission and omission errors. Notable commission errors are for instance high-intensive irrigated agriculture parcels being classified as wetlands because they resemble many of the inherent spectral characteristics of wetlands (i.e. high moisture and vegetation presence even in dry season). Omission errors will mainly be attributed to the large diversity of wetlands. Despite best effort to train the model across the widest range of wetlands possible, there will be types of wetlands and instances of wetland behavior that will not be adequately captured in a global model. Different wetland types also have varying level of detectability using satellite Earth Observation data. The identification, delineation and monitoring are therefore challenging from a pure Earth Observation perspective. In particular, there is frequent confusion between flooded areas (or surface waters) and wetlands, despite their major ecological difference. For example, some wetlands are only rarely and partly flooded, whereas many non-wetland habitats, such as agricultural or forested areas, can occasionally be flooded.

Further, vegetated wetlands were prioritized in the current study, why the areal estimate may lead to underestimations compared to national statistics which typically integrate metrics on surface water and coastal/marine wetlands[13]. One of the main reasons why Earth Observation-based tools often underestimate the extent of wetland habitats is that the spatio-temporal dynamics of flooded/wet areas are difficult to tackle, even with good time series. For instance, some ephemeral wetlands are rarely flooded or wet and therefore often missed by satellite datasets. In other cases, the "wet" part of a wetland

[13] Perennou, C., Guelmami, A., Paganini, M., Philipson, P., Poulin, B., Strauch, A., Tottrup, C., Truckenbrodt, J. & Geijzendorffer, I. R. (2018). Mapping Mediterranean wetlands with remote sensing: a good-looking map is not always a good map. In Advances in Ecological Research (Vol. 58, pp. 243-277). Academic Press.

may occur under a dense vegetation canopy, which is difficult to assess using Earth Observation data as used in this study, where the presence of water/moist conditions is not easily detected (cf. section on New data sources below).

For these and various other reasons, a perfect wetland map is not feasible, as all maps will always represent a generalization of the real world. Still, in this study, many of the observed deviations from the perceived real world are to a large extent also dictated by the time and resources available to develop this first global map of the distribution of vegetated wetlands. In other words, over time there is a significant scope (and readily available solutions) for taking the map to the next level. Below some of the key limitations are summarized along with some proposed solutions for improvements.

- **Stratification**: For this initial global mapping only regional stratification was possible including strata which were spanning over several countries and even continents. Ideally, and for improving national wetland inventories, the stratification should introduce national boundaries (and even state boundaries for larger countries) for better local/national prediction;
- **Training data**: Many countries do not yet have national wetland inventories and in which case global training is perfectly fine to get a first approximation of wetlands distribution, but ideally the mapping should integrate wherever possible national inventories obtained directly from countries. An iteration with countries in terms of reporting regularities could help improve the wetland map in an iterative process with the countries. In this study, the poorly trained regions are Russia, Greenland, Canada and the United States, but at least for Canada and the US it should be possible to obtain good wetland inventories for improved model calibration and training;
- **Improved DEM**: Terrain information from satellite derived DEMs is key input for mapping wetlands globally. The reference dataset is the HAND index derived from the 30-meter SRTM DEM which covers the globe from 60°N to 56°S. For the southern hemisphere this is less critical, as only Antarctica is not covered, but for the northern hemisphere large parts of Canada, Europe and Russia as well as Greenland are missing. For these areas an alternative readily available HAND model based on a 90-meter DEM was applied. Still, the quality of the HAND model north of 60°N is reduced as compared to the rest of the globe. Therefore, it is recommended to develop a new HAND model based on better elevation models in 30-meter pixel resolution, for instance from ALOS DEM or ASTER GDEM data;
- **Small islands**: Some small islands and potentially even entire small island states fall outside the acquisition plan of the Sentinels. As a result, no wetland prediction has been performed with the current approach. If needed, it would be possible to develop separate models for these missing islands using alternative input satellite data (e.g., Landsat);
- **New data sources**: For future improvements, it is also recommended to investigate L-band SAR data, which has greater vegetation penetration depths than C-band SAR data (cf. Sentinel-1). L-band SAR has successfully been used to map flooded areas under thick forest iterations. The use of L-band SAR has currently been restricted by data availability and cost, but this has recently changed with ALOS PALSAR now being distributed under a free and open license.

The global wetland extent map is the first of its kind outlining wetlands in unprecedented spatial detail. As indicated above, there is room for improvement, with several immediate solutions which could make the prediction more stable. By implementing some of the above solutions and recalibrating the model with the current global wetland inventory together with refinements submitted by individual countries, the global extent map could be enhanced substantially. This approach of continuous retraining and correcting the model will contribute to ensure a gradual increase in quality. As the global wetland extent map will evolve and improve over time it will also become relevant to reconsider a shift from a shallow learning model (cf. random forest) to a deep learning model (e.g., Convolutional Neural Networks and/or Recurrent Neural Networks). The advantage of deep learning models is the ability to more explicitly consider the temporal- and spatial aspects of wetlands when performing predictions which could further enhance the performance of the classification models.