

Combining machine learning and spatial data processing techniques for allocation of large-scale nature-based solutions

Caloir, Beatriz Emma Gutierrez; Abebe, Yared Abayneh; Vojinovic, Zoran; Sanchez, Arlex; Mubeen, Adam; Ruangpan, Laddaporn; Manojlovic, Natasa; Plavsic, Jasna; Djordjevic, Slobodan

DOI

[10.2166/bgs.2023.040](https://doi.org/10.2166/bgs.2023.040)

Publication date

2023

Document Version

Final published version

Published in

Blue-Green Systems

Citation (APA)

Caloir, B. E. G., Abebe, Y. A., Vojinovic, Z., Sanchez, A., Mubeen, A., Ruangpan, L., Manojlovic, N., Plavsic, J., & Djordjevic, S. (2023). Combining machine learning and spatial data processing techniques for allocation of large-scale nature-based solutions. *Blue-Green Systems*, 5(2), 186-199. <https://doi.org/10.2166/bgs.2023.040>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

Combining machine learning and spatial data processing techniques for allocation of large-scale nature-based solutions

Beatriz Emma Gutierrez Caloir ^a, Yared Abayneh Abebe ^{b,c}, Zoran Vojinovic  ^{b,d,e}, Arlex Sanchez ^p, Adam Mubeen ^{b,f}, Laddaporn Ruangpan ^{b,f,*}, Natasa Manojlovic ^g, Jasna Plavsic  ^d and Slobodan Djordjevic  ^e

^a Water Science and Engineering – Hydroinformatics Department, IHE Delft Institute for Water Education, Delft, The Netherlands

^b Water Supply, Sanitation and Environmental Engineering Department, IHE Delft Institute for Water Education, Delft, The Netherlands

^c Department of Hydraulic Engineering, Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands

^d Faculty of Civil Engineering, University of Belgrade, Belgrade, Serbia

^e College of Engineering, Mathematics and Physics, University of Exeter, Exeter EX4 4QF, UK

^f Faculty of Applied Science, Delft University of Technology, Delft, The Netherlands

^g Institute of River & Coastal Engineering, Hamburg University of Technology, Hamburg, Germany

*Corresponding author. E-mail: l.ruangpan@un-ihe.org

 BEGC, 0000-0002-5663-5377; YAA, 0000-0002-6416-6443; ZV, 0000-0002-7601-4041; AS, 0000-0003-3146-2841; AM, 0000-0003-1934-6813; LR, 0000-0002-7803-0600; NM, 0000-0002-0958-655X; JP, 0000-0001-9679-8851; SD, 0000-0003-1682-1383

ABSTRACT

The escalating impacts of climate change trigger the necessity to deal with hydro-meteorological hazards. Nature-based solutions (NBSs) seem to be a suitable response, integrating the hydrology, geomorphology, hydraulic, and ecological dynamics. While there are some methods and tools for suitability mapping of small-scale NBSs, literature concerning the spatial allocation of large-scale NBSs is still lacking. The present work aims to develop new toolboxes and enhance an existing methodology by developing spatial analysis tools within a geographic information system (GIS) environment to allocate large-scale NBSs based on a multi-criteria algorithm. The methodologies combine machine learning spatial data processing techniques and hydrodynamic modelling for allocation of large-scale NBSs. The case studies concern selected areas in the Netherlands, Serbia, and Bolivia, focusing on three large-scale NBS: rainwater harvesting, wetland restoration, and natural riverbank stabilisation. Information available from the EC H2020 RECONNECT project as well as other available data for the specific study areas was used. The research highlights the significance of incorporating machine learning, GIS, and remote sensing techniques for the suitable allocation of large-scale NBSs. The findings may offer new insights for decision-makers and other stakeholders involved in future sustainable environmental planning and climate change adaptation.

Key words: flood risk reduction, large-scale nature-based solutions, machine learning, NBS planning, spatial data processing

HIGHLIGHTS

- The paper provides enhanced methodologies for mapping NBSs using novel techniques.
- The methodology combines machine learning (ML) and spatial data processing techniques for allocation of large-scale NBSs.
- The methodology also includes outputs from a 2D hydrodynamic model.
- The methodologies have been thought to be applicable worldwide and not only in the study areas, and it is only necessary to have available data.

1. INTRODUCTION

Hydro-meteorological hazards such as floods, landslides, droughts, and heatwaves have increased globally due to human activities and climate change (Singh *et al.* 2016). These hazards produce environmental damage, life losses, and world economy stress every year (IPCC 2014). To reduce impacts of climate change, nature-based solutions (NBSs) are suitable measures, especially considering the conventional hydraulic structures are not sustainable and adaptive. NBSs are solutions and actions that mimic nature, integrating hydrology, geomorphology, hydraulic, and ecological dynamics (European Commission Directorate-General for Research & Innovation n.d.; Haring *et al.* 2021). NBSs can offer multiple benefits through multi functionality, but their effectiveness varies due to the scale of applications.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY 4.0), which permits copying, adaptation and redistribution, provided the original work is properly cited (<http://creativecommons.org/licenses/by/4.0/>).

NBSs can be categorised into three types based on the scales of analysis: fine, local, and regional scales (Somarakis *et al.* 2019). Fine-scale NBSs include gardens, small parks and vegetated roofs, and walls. They are usually implemented to mitigate noise and heat islands, support biodiversity, reduce flood risk, and save energy. Local-scale NBSs are larger parks, forests, and other green spaces that impact the local environment. These measures manage the local environment, enhance air quality, protect wildlife, and manage water. Regional-scale NBSs are applied in rural, highland and seaside regions, such as swamps and river restoration, dunes, and wetlands. Fine and local-scale NBSs are also called small-scale NBSs, while regional-scale measures are considered as large-scale NBSs. Large-scale NBSs provide more co-benefits than the small-scale ones, considering they are applied in larger areas and interacting more with catchment processes (Watkin *et al.* 2019; Ruangpan *et al.* 2020; Kumar *et al.* 2021; Vojinovic *et al.* 2021). The co-benefits include long-term solutions, job generation, carbon sequestration, and enhancing ecosystem and biodiversity.

Examples of large-scale NBSs include rainwater harvesting (WH), wetland restoration (WR), and natural riverbank stabilisation (NRS). WH involves intentionally collecting rainwater from surfaces such as roofs, ground surfaces, and watercourses, and storing it in structures or the soil profile (Malesu *et al.* 2006). This practice effectively mitigates the impacts of floods and droughts in rural areas. WR focuses on identifying degraded areas within wetlands and restoring them to their original state, thus rehabilitating hydrological functions (World Bank 2008). Wetland degradation often results from land use changes to croplands. Wetlands play a crucial role in reducing flood risks by attenuating flow peaks and mitigating droughts through water storage. NRS are measures implemented to manage erosion and protect and stabilise riverbanks (Allan & Castillo 2007). Rivers undergo constant geomorphic processes, including erosion, sedimentation, and bank instability, particularly during extreme events. NRS implementation involves bioengineering and green solutions such as riverbank revegetation, reducing the impacts of floods and landslides.

Implementation of NBS projects involves key phases such as scoping, planning, design and construction and operation and maintenance, and in some cases, decommissioning. Planning includes hazard data collection, NBS assessment and socioeconomic analysis. Locating suitable NBS sites is a crucial step at the planning phase. A systematic review by Mubeen *et al.* (2021) identified seven tools/methods for spatial NBS allocation. Most of the tools/methods focus on small-scale urban NBSs, except for some scholars (Guerrero *et al.* 2018; Mubeen *et al.* 2021) who developed a methodology using 'floodplain hydro morphological landscape units' to identify potential areas for large-scale NBSs. Following the review, Mubeen *et al.* (2021) developed suitability mapping methodologies for detention basins, retention ponds, floodplain restoration, river widening, forest buffers and afforestation. They combined criteria based on slope, soil type, distance from streams, land use type and road buffers to create general suitability maps, which were further refined with NBS-specific criteria. These methodologies were implemented using Esri ArcMap software's model builder functionality. However, there is still the need to develop mapping methodologies for other large-scale NBSs or to further improve the existing ones.

This article further contributes to NBS planning by developing suitability mapping methodologies for WH, WR, and NRS. For WH, Mubeen *et al.* (2021) developed a methodology for allocating retention ponds, which requires consideration of climatic conditions for rainwater availability. Droogers & van Kempen (2014) proposed a parametric approach that uses seven spatial datasets of which some are specific to certain WH designs. The present research does not include those datasets while considering other relevant, spatial information used as constraints. As for WR and NRS, there is no standard procedure or study for mapping suitable locations. Hence, our work provides contribution in this direction. The methods developed are tested and analysed on at least one study area where each of the three large-scale NBSs were evaluated. The study was conducted within the EC-funded H2020 RECONNECT project (<http://www.reconnect.eu/>).

2. DATA AND METHODS

As the three NBSs have different characteristics and can be applied in different settings, the datasets needed and methodologies developed are also different. The upcoming sections describe the three methodologies developed. In all cases, the ModelBuilder functionality in ArcGIS has been used to create geoprocessing models (toolboxes). ModelBuilder (ESRI 2022) is a visual programming language that allows automating and documenting spatial analysis and data management processes. The tool enables data integration, analysis, and visualisation, facilitating workflow sequencing and model sharing.

2.1. ML for GIS applications

Part of the common input data for the toolboxes concerning WR and NRS requires a land cover raster dataset in order to produce results with higher accuracies. Several land cover products are available, such as [European Spatial Agency \(ESA\) WorldCover \(2020\)](#), and [ESRI \(2020\) Global Land Cover Map \(ESRI\)](#), which offer a spatial resolution of 10 m. Alternatively, Copernicus Corine Land Cover (CLC), and Copernicus Global Land Cover (GLC) provide a spatial resolution of 100 m. However, it has been observed that some of these products exhibit limitations in accurately identifying certain land cover types, particularly differentiating between croplands and grasslands. This issue arises due to the large scale at which these products are developed, often covering vast areas worldwide. On the other hand, while some products can correctly identify croplands, they may contain incomplete sections due to their coarse spatial resolution. [Table 1](#) presents a summary of the key features of the four existing land cover products mentioned above, all of which are freely accessible online.

Table 1 | Summary of open access land cover data features based on product information

Description	Spatial resolution (m)	Global cover	Source imagery	Availability	Accuracy (%)
Copernicus Corine Land Cover	100	No	Landsat 5 Landsat 7 Spot 4/5 IRS P6 LISS III RapidEye Sentinel2 Landsat 8	1990, 2000, 2006, 2012, 2018	85.0
Copernicus Global Land Cover	100	Yes	Prova-V Sentinel 2	2015, 2016, 2017, 2018, 2019	75.43
ESRI 2020 Global Land Cover Map	10	Yes	Sentinel 2	2020	86.0
ESA WorldCover (2020)	10	Yes	Sentinel 2 and Sentinel 1	2020	74.4

Note: The data accuracy has been increasing, according to the improvements in technology and data mining used to create the different products through the years. For example, CLC 2000 has a reliability of 74.8%, evaluated in randomly selected locations. Similar accuracy variations occur for products from GLC (see more at <https://land.copernicus.eu/pan-european/corine-land-cover>).

Source: Own elaboration based on technical reports ([European Space Agency 2020](#); [ESRI Inc. 2021](#); [Tsendbazar et al. 2021](#)).

To obtain a land cover dataset with higher accuracy specifically tailored for the study areas, one of the machine learning (ML) techniques, a Random Forest (RF) classifier was employed in a supervised classification workflow implemented in Google Earth Engine (GEE), a cloud-based platform. The RF classifier was chosen due to its ability to handle complex classification problems and provide reliable results ([Rodriguez-Galiano et al. 2012](#); [Collins et al. 2020](#)).

The classification process primarily utilised the Sentinel Multispectral image collection, which offers a spatial resolution of 10 m. In some cases, Landsat imagery was used for time periods predating the availability of Sentinel images. The selection of images was based on their availability within the desired time range.

The output of the classification process is a land cover raster with enhanced accuracy compared to existing products developed for larger geographic extents, such as worldwide or continental areas. The accuracy assessment of the RF classification was conducted using the Kappa index. The Kappa index is a statistical measure that evaluates the performance of a classifier by comparing the observed and expected accuracy of the raster classification process ([Landis & Koch 1977](#); [Congalton & Green 2008](#)). It is calculated using Equation (1).

$$\kappa = \frac{(P_o - P_e)}{(1 - P_e)} \quad (1)$$

where κ represents the Kappa index, P_o denotes the proportion of observed agreement between the classifier and the reference data, and P_e indicates the proportion of expected agreement due to chance.

By computing the Kappa index, the accuracy of the RF classifier can be quantified, providing an objective measure of its performance in classifying the land cover in the study areas. [Landis & Koch \(1977\)](#) provide the

categorised interpretation of the Kappa coefficient, which indicates the level of agreement between the classification results and the reference data. Kappa values greater than 80% and less than 40% indicate strong and low agreement, respectively, between the classification and the reference data, while values between 40 and 80% represent a moderate level of agreement (Congalton & Green 2008). This classification scheme enables the assessment of the accuracy of the RF classifier based on the calculated Kappa index.

2.2. Mapping suitable sites for WH

The methodology is based on an open pond WH scheme. Considering the use of WH for diverting and collecting stormwater to prevent flooding and using stored water in dry periods, the present research uses six main input datasets – total annual precipitation, coefficient of variation of the seasonal monthly precipitation, aridity index, runoff coefficient, slope, and population density. According to Droogers & van Kempen (2014), the total annual precipitation provides vital information on the amount of water available in a basin throughout the year. The variation in the monthly precipitation is used to identify the peaks and extreme events during the rainy season. The aridity index shows which areas need WH most while the runoff coefficient indicates the areas that generate higher surface runoff. Slope, which is extracted from a digital elevation model (DEM), shows the influence of topography on the surface runoff flow. As WH is implemented to benefit communities, population density identifies the spots with higher concentration of settlements in a basin.

The input data are processed in raster format and have the same spatial resolution and projection. However, as the data have different units of measure, they must be normalised, from 0 to 100. In this process the highest value of the raster is represented by 100 and the lowest by 0. Once map algebra is applied, the result is a dimensionless raster where the highest values are the most suitable to locate WR than the lowest values.

In order to link the different rasters, a multi-criteria analysis was used with the Analytical Hierarchy Process (AHP). This method allowed complex structures to be broken down into their components, ordering them in a hierarchical structure, where numerical values are obtained for preference judgments and, finally, synthesised to determine which component has the highest value (priority) (Cleveland & Ayres 2004). The AHP uses a matrix of judgments, comparing pairs of alternatives, prioritising one over another, and assigning a value according to its importance. This matrix allows obtaining a set of weights to score every option (Hopkins 2001). The pair-wise comparison matrix includes the six input datasets. According to the method of AHP, the consistent ratio index (Equation (2)) must be below 10%, which means the pair-wise comparison is consistent.

$$CR = \frac{CI}{RI} \quad (2)$$

where CR is consistent ratio index, RI is the random index linked to the size of the matrix, and CI is a term that depends on the number of analysed variables and a lambda factor whose maximum value is related to the own maximum value of the framework (Xiong *et al.* 2019).

The weight values obtained with the AHP are used as coefficients for an algebra expression that enables combining the raster input data which are previously normalised. Those weights could be changed by considering the opinions of experts and stakeholders, and field survey to build a validated pair-wise matrix. The result of the map algebra expression is the sum of the input variables multiplied by its coefficient (Equation (3)). To coherently perform spatial analysis, we converted the resolution of the input data to match that of the slope data, which is 12.5 m, as it is the finest resolution of all. The output data is a raster file showing the potential areas for WH, and its values range from 0 to 100%.

$$\begin{aligned} \text{Raster result} = & \text{total annual precipitation} * a + \text{coefficient of variation monthly precipitation} * b \\ & + \text{runoff coefficient} * c + \text{aridity} * d + \text{slope} * e + \text{population density} * f \end{aligned} \quad (3)$$

where a , b , c , d , e , and f are coefficients determinate by the found weights.

Not all areas that satisfy the above input conditions can be used to implement WH schemes. The places where WH cannot be applied should be removed at the end. The constraints considered in this study are: transport lines (in this case, roads), water bodies (in this case, rivers) and buildings. Roads and rivers are usually available in vectors and should be converted to raster, whereas buildings are extracted from land use data in raster format. The raster cells that contain buildings, buffered roads and rivers should be assigned null values to avoid using them as potential suitable areas.

Based on meteorological and remote sensing data, the present research applies various methods and datasets to analyse and generate data. The coefficient of variation of seasonal monthly precipitation was calculated based on data statistics from rain gauge stations within or near the study area. This analysis considered the seasonal variation during the wet period. The data were then interpolated in ArcGIS to produce a raster. Annual precipitation, runoff coefficient and aridity data can be generated for any region worldwide using information from meteorological services or satellite data sources such as CHIRPS, TRMM, and PERSIANN. GEE platform was employed to obtain raster data for annual precipitation, monthly seasonal variation coefficient (Banerjee *et al.* 2020), runoff coefficient and aridity index maps.

A DEM was used to delineate the catchment area of the study region and determine the slope. In this study, DEM Alos Palsar (2023) was used due to its higher spatial resolution (12.5 m) and the ability to delineate a basin and river network with greater precision. It is important to note that while this DEM covers most regions in South America and is freely accessible online, it is not available worldwide. For other parts of the world, other DEMs with different resolutions can be used. Population density data are usually accessed from a country's statistics office.

The facilities/buildings layer was derived from the land cover data provided by ESA. River and road networks from OpenStreetMap were buffered and converted to raster format. These buffers represent constraint areas where WH cannot be implemented and are assigned null pixel values. The buffer distance for roads follows the standard set by the American Association of State Highway and Transportation Officials (AASHTO 2020), which designates a 50-m width on each side of the road centreline as the 'right-of-way' area. Similarly, the same criteria of a 50-m buffer on each side of the river or water body centreline were adopted.

2.3. Mapping suitable sites for WR

The international convention Ramsar, established in 1971 with the mission of preserving and promoting the sustainable global development of wetlands, provides online free access to a comprehensive database on wetlands worldwide (RAMSAR 2010). This database serves as a valuable resource for the study area, and the following methodology involves identifying areas within Ramsar wetlands where natural land cover has been converted into croplands. Both vector and raster data have been used for this purpose.

The required vector data for this analysis are the Ramsar Polygons (RAMSAR 2010), which define the boundaries of the wetland. In addition, vector data from OpenStreetMap were employed to identify railways and roads, which serve as constraints in the analysis. The buildings layer used in the analysis was derived from the ESA WorldCover (2020) product, chosen for its high accuracy in recognising buildings. This product incorporates samples from various locations worldwide, enhancing its ability to identify buildings accurately. The land cover is the result of the RF classification computed in GEE, discussed in Section 2.1. The RF classifier's accuracy relies on the sampling process, which may be limited in small, specific areas.

The toolbox workflow begins by clipping the land cover dataset, buildings layer roads and railways using the wetland polygon as a mask. This retains only the relevant portions within the wetland area for analysis. The land cover classification output is then processed to extract water bodies and crop layers, which are converted into vector format for easier integration and analysis. A 50-m buffer is applied around road and railway features based on standard guidelines (AASHTO 2020), representing the designated right-of-way area. These layers are then intersected with the crop and building layers to identify intersections that may impact WR suitability.

To refine the analysis, the resulting intersection layer is further intersected with the study area polygon, excluding the regions outside the wetland study boundary. This step ensures that only areas within the wetland study area are considered for WR measures, excluding regions beyond the study boundary. The toolbox generates a vector file indicating suitable areas for WR.

2.4. Mapping suitable sites for NRS

Some of the datasets required to map suitable sites for NRS are land cover information derived from the RF classification, slope data from a DEM, soil type, and erosion index. A hydrodynamic model was developed to generate the maximum depth and velocity of a river. Velocity thresholds are defined to distinguish between permissible and potentially dangerous velocities, which are dependent on the riverbank (Bureau of Reclamation 2015). For example, the threshold of permissible velocity for sandy loam soil is 0.53 m/s.

The suitability mapping process began with resampling and clipping the erosion index raster. At the same time, land cover classification rasters were used to extract four Boolean rasters (contain values true or false). Layer classified as crops and global erosion index were intersected, identifying the weakest zones. Values higher than a permissible velocity of 0.53 m/s are considered erosive velocities in the riverbanks. This value was used to create a new maximum velocity raster, which was intersected with the previous erosion index-crops raster. This is the first intermediate raster output. The normalised classification raster was combined to remove all the features not located close to the river borders (layer of sediments, water-riverbed). Separately, the slope data was intersected with the maximum depth raster, a second intermediate raster output. Finally, the two intermediate raster outputs were intersected to provide areas prone to destabilisation and require NRS. NRS are typically applied along a river, near settlements and croplands, to protect them from floods. Implementing these measures along the entire river would disrupt natural processes and lead to undesirable impacts and ecosystem disservices (Somarakis *et al.* 2019).

3. STUDY AREAS AND DATA SOURCING

In this study, we applied the suitability mapping methodologies to four areas. However, the methodologies can be applied to any case study given data availability, as the toolboxes are generic. Below, we described the four study areas.

3.1. Study area for WH

The arid region of Santiago de Machaca, Bolivia, was selected for the WH analysis due to challenging weather conditions and data availability. Droughts during long dry seasons and riverine floods during short wet seasons cause economic damage to potato and barley crops in this area. Figure 1(a) shows the study area, which is part of the La Paz Department in Bolivia.

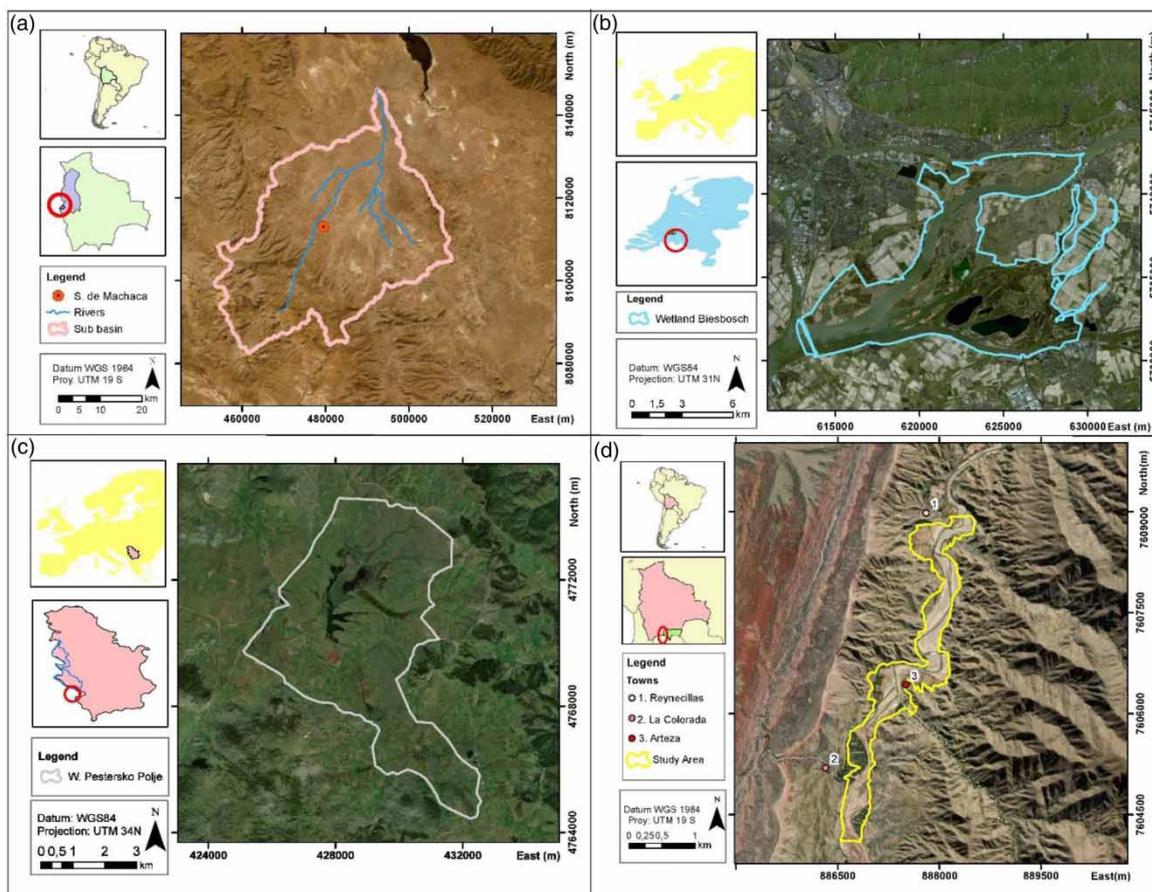


Figure 1 | Location of the study areas: (a) Santiago de Machaca – Bolivia, (b) Ramsar Wetland Biesbosch, the Netherlands, (c) Ramsar Wetland Peštersko Polje, Serbia, and (d) River San Juan del Oro Tarija, Bolivia.

Datasets required to develop the toolbox were obtained from the latest water balance report by *MMAYA et al.* (2018), including total annual precipitation, total annual evapotranspiration, and runoff coefficient. The aridity index is a sub-product calculated by dividing total annual precipitation by total annual potential evapotranspiration. All rasters have a spatial resolution of 30 m except for precipitation, which has a coarse spatial resolution of 5,000 m. To ensure uniformity, the spatial resolutions of all input rasters and sub-products were set to 30 m.

Precipitation data from rain gauge stations in the study area were used to calculate the coefficient of variation of the seasonal monthly precipitation. The wet season was analysed separately to identify sudden heavy rain events. The coefficient of variation values were interpolated to create a raster dataset. Population data from the GeoBolivia website were interpolated and converted into raster format. Alos Palsar DEM was another source of information applied to obtain sub-products such as slope, hillside, and the delineation of the study basin. The last raster source of information was the land cover *ESA 2020/ESRI* products, which was used to extract the buildings layer and identify constraining locations.

3.2. Study area for WR

To test the WR toolbox, which identifies areas with a land use change inside a wetland, two RAMSAR polygons within the RECONNECT project were chosen. *Figure 1(b)* shows the first study area, a Ramsar wetland called National Park Biesbosch, located in Dordrecht, the Netherlands. The area of the wetland is about 103 km². The input data and general methodology depend on the availability of information and the proper characteristics of the place. *Figure 1(c)* displays the second study area, a Ramsar wetland called Peštersko Polje, located in Serbia. The wetland has an area of about 34 km², and it belongs to the Drina River basin.

3.3. Study area for NRS

The study area selected to test the NRS toolbox is located along the San Juan del Oro River, near Arteza town in Tarija, Bolivia. The region primarily relies on agriculture, and the study area is susceptible to unstable riverbanks and flooding, impacting nearby agricultural areas and towns. Riverine flood has resulted in significant agricultural and livestock losses, negatively affecting local and regional economies. The study area was selected because it follows up a previous study (*MMAYA & BID 2021*) and the potential expansion of research regions in South America within the RECONNECT project.

Figure 1(d) illustrates the study area, including nearby towns along the river. The region exhibits scattered vegetation cover and steep slopes, making it susceptible to destabilisation. The available information for this research includes a topography survey of the Arteza region and the proposed location of a protection dike according to the final report. Hydro-meteorological time series data were collected from National Meteorology and Hydrology Service of Bolivia, while data on towns, country borders, and rivers were obtained from (*Geobolivia 2020*).

4. RESULTS AND DISCUSSION

4.1. Applying WH toolbox

Following the methodology of AHP, a 6 × 6 matrix was built, using the raster variables: total annual precipitation, coefficient of variation of the seasonal monthly precipitation, aridity index, slope, population density, and runoff coefficient. The matrix was computed and its statistical indexes of lambda, CI, and CR were 6.32, 0.06, and 5.8%, respectively. A CR value below 10% means the pair-wise comparison is consistent. The values of the weights obtained in the pair-wise matrix for every raster variable are 0.31 for total annual precipitation, 0.41 for variation coefficient, 0.12 for runoff coefficient, 0.08 for aridity, 0.09 for slope, and 0.21 for population density. These weights were introduced in the toolbox through a map algebra expression.

The output results consist of a continuous raster in which four categories have been set: null, low, medium, and high potentiality. *Figure 2* shows the resulting map for potential areas for WH. It can be inferred that the sites offering medium and high potentiality are close to the higher elevations (or in between) upstream of the study basin. These zones are characterised by orographic precipitation due to the presence of mountain chains.

The result depends on the weights assigned to each input data obtained from the AHP. However, the weights, and hence the result, could have been different if the relative importance of a variable in the pair-wise comparison matrix were different.

Regarding the thresholds to define the four categories mentioned, they were set based on a statistical analysis from ArcGIS graphics. These statistics were extracted from the raster of the AHP, a boxplot was obtained, which the quartiles helped to define the thresholds.

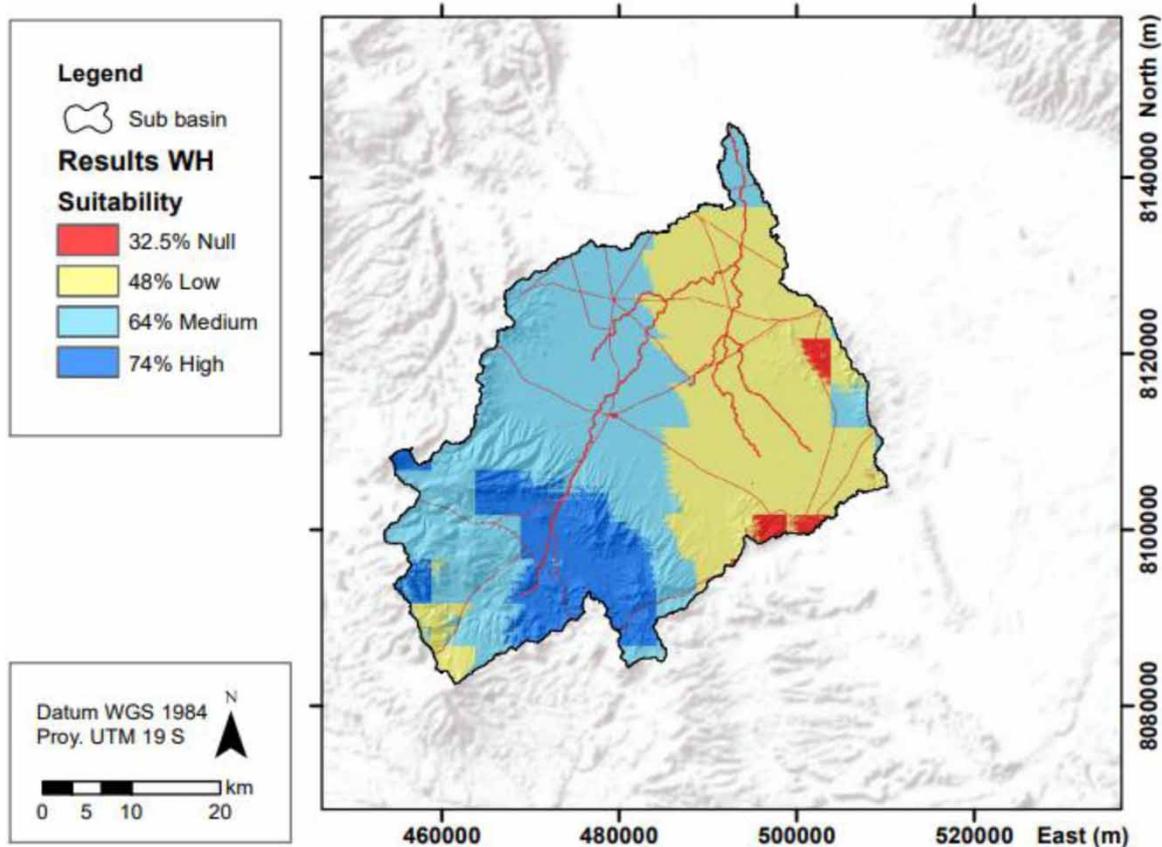


Figure 2 | Suitable location of potential areas for WH.

4.2. Applying WR toolbox

4.2.1. Results of the study area of Ramsar wetland Biesbosch

The land cover computed in GEE has high accuracy, which is expressed by a Kappa training value of 99.96% and a Kappa validation value of 87.92%. The area used for croplands obtained for 2015 is 6.86 km², and for 2021 is 7.79 km² as shown in [Figure 3](#). This means in 2021 there was an increment in cropland areas of 11.9% compared to 2015.

4.2.2. Results of the study area of Ramsar wetland Peštersko Polje

About the results of the study area of Ramsar wetland Peštersko Polje, after running the toolbox, the suitable areas for WR are about 24.6 km². It means that 71.95% of the wetland is used for other purposes instead of natural areas. To visually verify results obtained, [Figure 4](#) shows a comparison of a randomly selected spot inside the wetland. The three circles show the comparison of (a) arable area for crops, seen in the base map of ArcGIS, (b) suitable areas obtained for WR, and (c) [ESA 2020](#) land cover. The ArcGIS base map clearly shows land prepared for cultivation, surrounded by cropland.

The area obtained from the toolbox and ESA classification correctly identifies the brown arable land as cropland. However, the surrounding area is also cropland, but it is green with grown crops. From this it follows that ESA classification is not correct for the surrounding area because it classifies it as grassland.

4.3. Applying NRS toolbox

The results obtained from the application of the NRS toolbox (highlighted in yellow colour in [Figure 5](#)) indicate several areas that are prone to destabilisation. There the vulnerable area includes the river bank near the Arteza town entrance, the upstream riverbanks and the bank along the San Juan del Oro river. Based on these results, it can be seen that the NRS results along the San Juan del Oro river are at the same location that reported riverbank destabilisation and proposed dike locations to mitigate floods (highlighted in red colour in [Figure 5](#) as documented by [MMAYA & BID \(2021\)](#)).

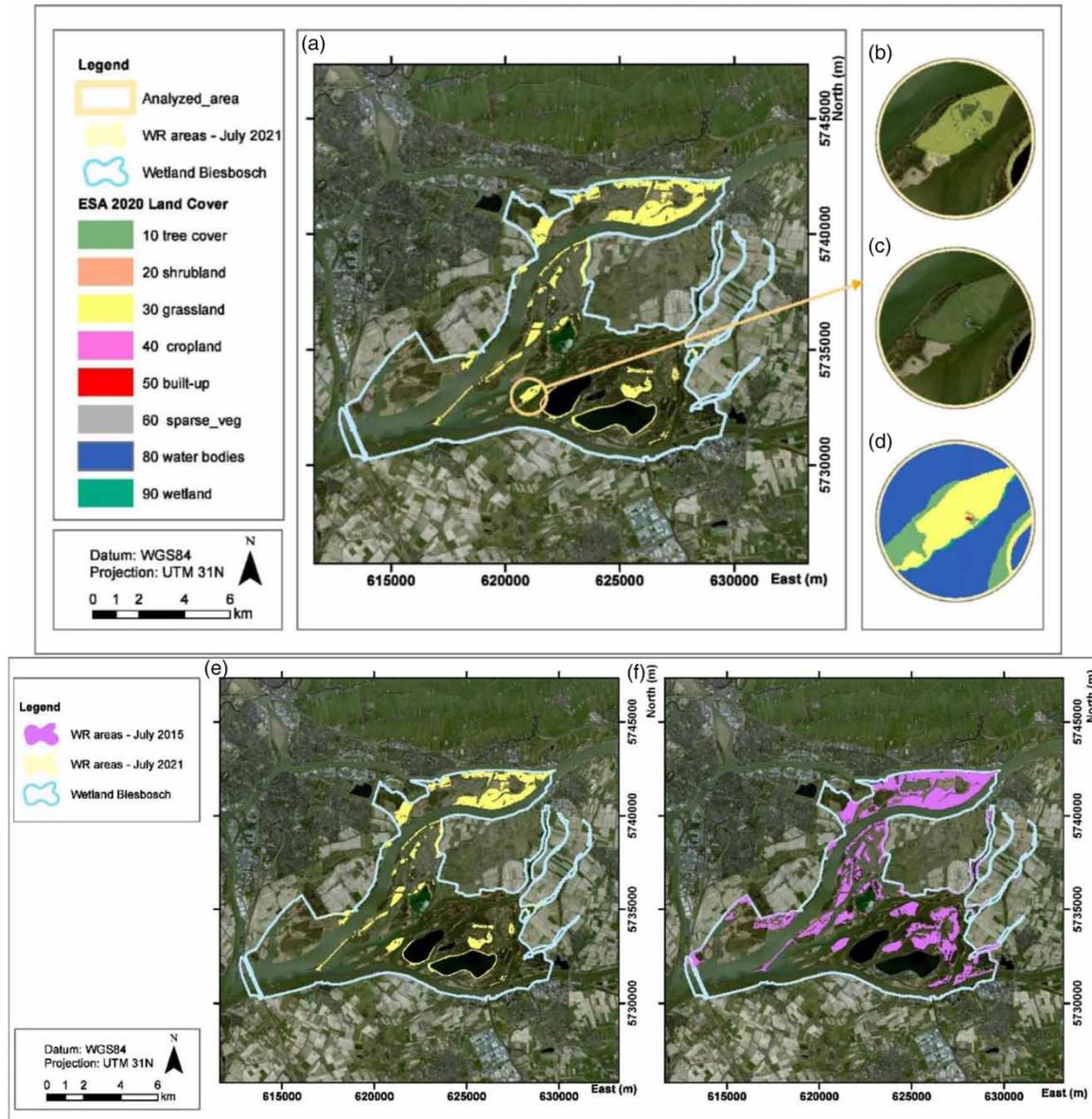


Figure 3 | Results of suitable areas obtained for WR inside the polygon of the Biesbosch, using RF classification. (a) Results for WR areas for July 2021. The next three maps make a comparison of a selected specific cropland area: (b) shows suitable areas for WR obtained with RF (c) same spot, cropland seen with a base map of ArcGIS (d) same spot, ESA 2020 land cover. Finally, maps (e) and (f) show the results of suitable areas for WRE for two different dates (e) July 2015 and (f) July 2021.

The toolbox conducted spatial analysis considering soil type, land cover, erosion, maximum flow velocity, and maximum flow depth, while [MMAYA & BID \(2021\)](#) based their results on a hydrodynamic model. Consider as part of validation the small overlap of 30 m between the dike and the areas identified by the toolbox. It can be inferred that the toolbox successfully identified the required areas.

The study site has a soil type of sandy loam. The results obtained using the NRS toolbox set the input condition for the maximum velocity raster with values higher than 0.53 m/s, higher values produce erosion in riverbanks. These results can be improved using sensitivity analysis for variables such as velocity, depth, and slope. A simple uncertainty analysis was conducted by examining the boxplot velocity values for the first, median, and third quartile. [Figure 6](#) includes the results for these new thresholds of maximum velocities. It can be seen that higher velocity values reduce the locations where natural bank stabilisation is required. Those locations are more concentrated on the riverbank along the close curves of the meanders.

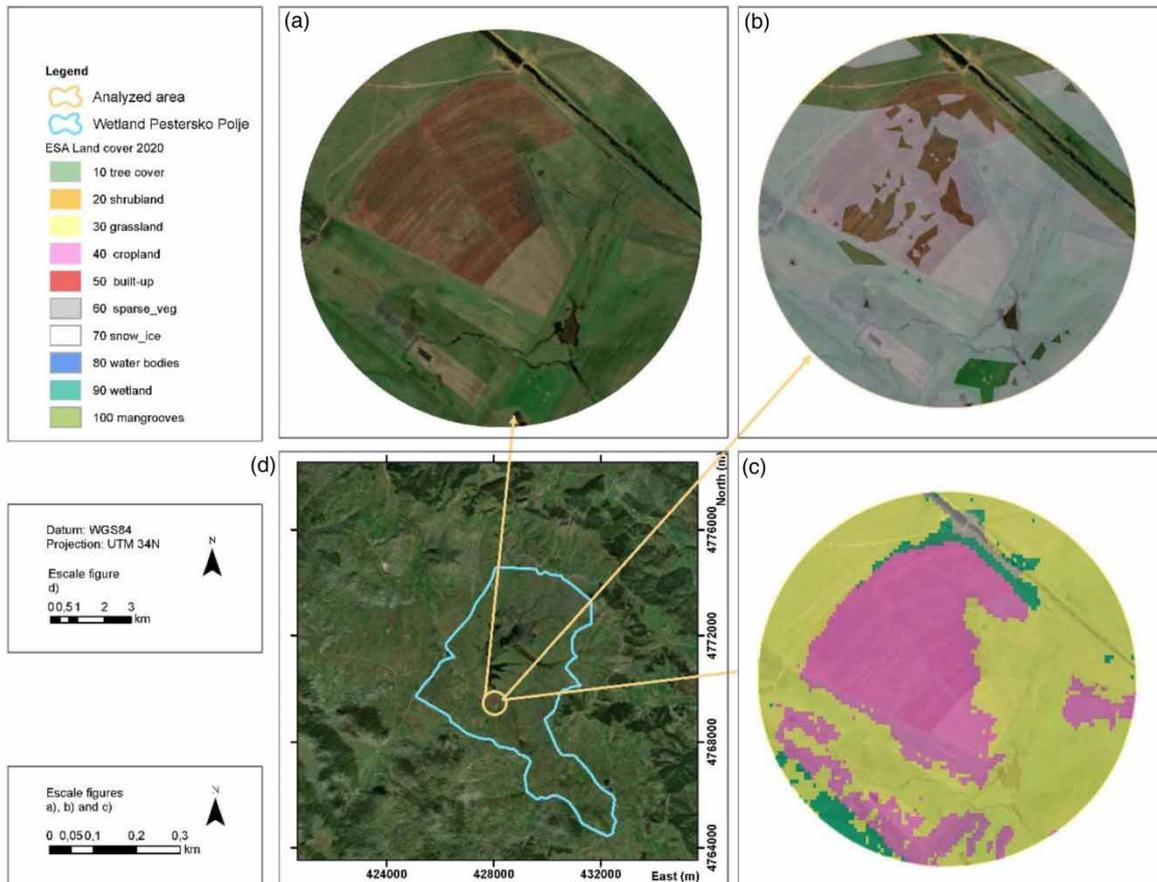


Figure 4 | Verifying results for a specific cropland area by comparison (a) cropland seen with a base map of ArcGIS; (b) suitable areas for WR; (c) ESA 2020 land cover; and (d) study area of Peštersko Polje.

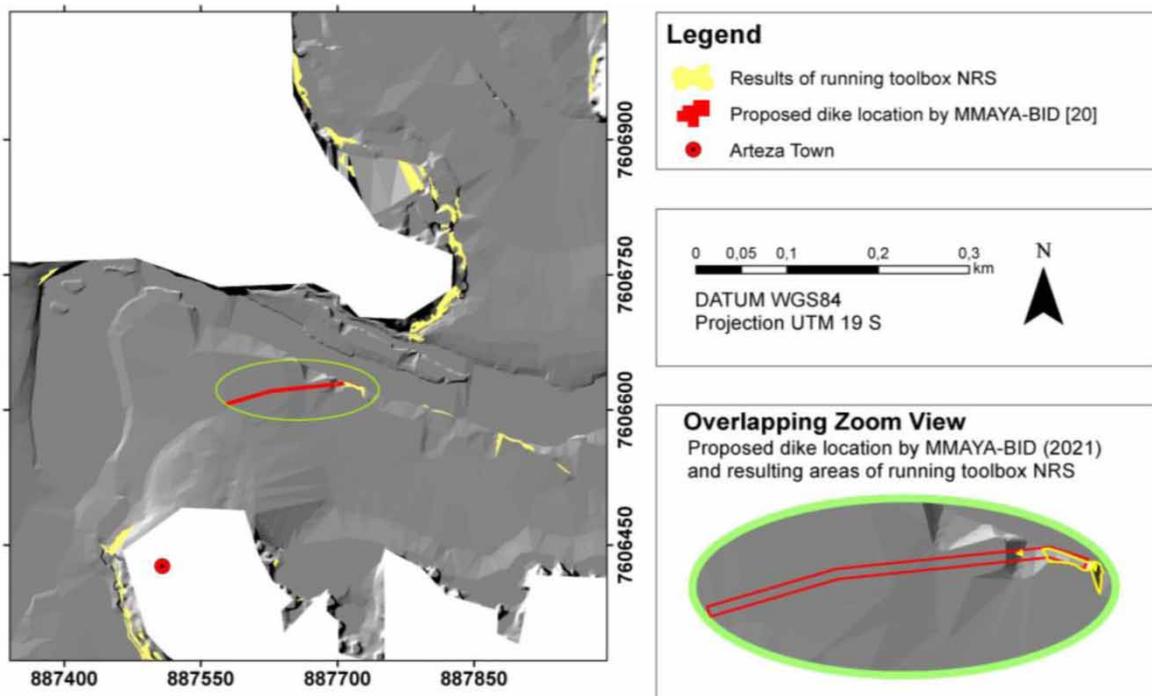


Figure 5 | Comparison of results for NRS with a previous result report (MMAYA *et al.* 2018). *Source:* Own elaboration based on results of this research and results of the dike location of proposed civil works in Arteza Town (MMAYA & BID 2021). The grey background shows the surface terrain.

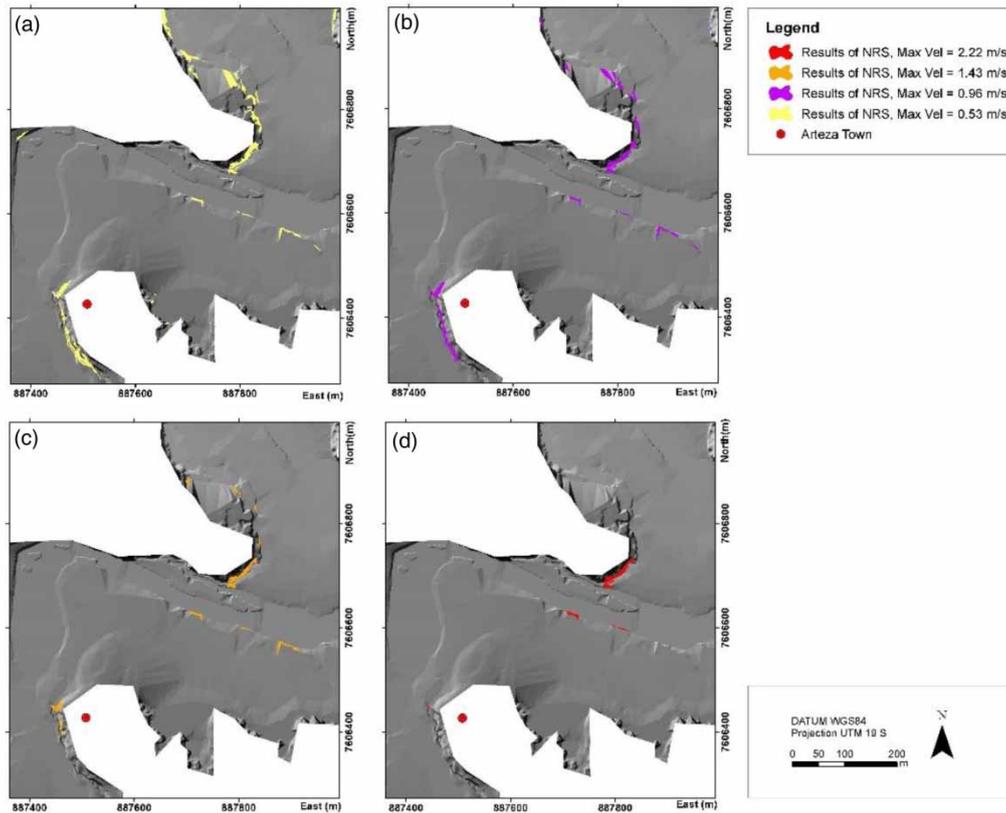


Figure 6 | Comparison of results of areas that require NRS using different values of maximum velocity: (a) 0.53 m/s, (b) 0.96 m/s, (c) 1.43 m/s, and (d) 2.22 m/s. The grey background shows the surface terrain.

5. LIMITATIONS AND INSIGHTS INTO FUTURE WORK

The main challenge with WH is how to display the final results. Raster results are tied to user satisfaction, and the criteria for this can vary among stakeholders. Some sensitivity analysis was conducted, and the results show that there are some sensitive in changing the classification of the thresholds. The obtained results are very similar results to a classification based on natural breaks. It is important to note that the underlying raster data remain constant; the only variation lies in how the information is presented to the user. Furthermore, this is a limitation to define better the thresholds to classify and display the results. Future works can incorporate diverse precipitation data, including modelled precipitation, coming from rainfall-runoff models, which use the temperature data from climate change models. This enables scenario creation and comparison against the baseline, which uses existing precipitation data from sources like measured rain gauge data, available modelled data like CHIRPS, TRMM, or PERSIANN. Another future improvement consists of using field validation data and insights from stakeholders and experts from the study area, to assign the weights of importance and compute the pairwise matrix. Finally, is necessary to refine the raster result presentation by exploring alternative classification criteria for better visualisation.

The main limitation in WR is the field validation and measurement. Therefore, the field inspection/validation is necessary not only for the results of the toolbox developed, but for any source of land cover data. This criterion is particularly relevant for Pestersko Polje, where a visit like the one made to Biesbosch National Park was not possible. Therefore, a field inspection is necessary to validate not only the results of the toolbox developed, but also any land cover data. However, we can enhance the toolbox by incorporating additional methodologies to identify various wetland degradation causes and pinpoint their locations.

The primary constraint for NRS is acquiring a high-resolution DEM from a topography survey, which serves as input for the hydrodynamic model. Nevertheless, further validation is essential through comprehensive field data collection to confirm the accuracy of flood-prone areas along the riverbank. In the coming years, these tools will be indispensable for decision-makers in managing hydro-meteorological hazards. Future iterations of the toolbox may integrate seamlessly with the hydrodynamic model, streamlining the process.

6. CONCLUSIONS

The main objective of the present research work was to enhance the existing planning toolbox by adding spatial allocation methodologies for additional large-scale NBSs to manage hydro-meteorological risk reduction. The work combines ML and spatial data processing techniques to allocate large-scale NBSs. The RF ML technique was applied to obtain high accuracy of land cover raster for WR and NRS. This approach ensures robust results for the land cover classification as defined by the statistical index Kappa, which is then used for both training and validation purposes.

To achieve this, three separate toolboxes were developed within ArcGIS to facilitate the mapping of suitable locations for NRS, WR, and WH. These toolboxes were then applied in four case study areas. Furthermore, these measures were selected based on a targeted search into the database of large-scale NBSs belonging to the EC H2020 RECONNECT project, considering the preferences/needs of the stakeholders (Ruangpan *et al.* 2021), as well as the need to expand the previously developed toolbox.

WH is an improvement of the existing toolbox for Retention Ponds (Mubeen *et al.* 2021). The added methodology is based on multi-criteria analysis applying the Analytical Hierarchical Process (AHP). The pair-wise matrix computed in the process determines the weights of importance of every raster input data, meeting consistency conditions.

The WR methodology primarily relies on land cover input data. Land use change from natural areas to crops is the primary cause of wetland degradation (Yazar & Ali 2016). To identify croplands, the online remote sensing products for land cover were applied with some limitations related to spatial resolution, study area coverage (European Environment Agency 2006), accuracy, and errors in defining croplands and grasslands (European Space Agency 2020; ESRI Inc. 2021). A land cover dataset with higher accuracy was obtained using the RF classifier in the cloud platform GEE. The case studies focused on Ramsar wetlands in Biesbosch and Peštersko Polje, where areas suitable for restoration were identified. To validate the results, the toolbox outputs were compared with high-definition satellite base maps in ArcGIS. However, further verification is still needed by cross-referencing with field data.

The NRS methodology is based on multi-criteria analysis. It considers the combination of maximum velocity and maximum depth with the geomorphologic features of the study area. In this case, the HEC RAS 2D hydrodynamic model was built and simulated to obtain output rasters with peak velocities and peak depths. They were part of the inputs used in the toolbox for mapping zones for NRS. The toolbox identified destabilisation-prone areas as output results. A comparison with an existing report provided locations for a future protection dike. Decision-makers can utilise these results to determine specific areas for implementation of NRS. This approach may contribute to the protection of ecosystems, settlements, crops, and other vulnerable areas without compromising river behaviour.

ACKNOWLEDGEMENTS

The research leading to these results has been conducted within the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 776866 for the research project RECONNECT. The study reflects only the authors' views, and the European Union is not liable for any use that may be made of the information contained herein.

Regarding Bolivian case studies, the dataset used belongs to the Ministry of Environment and Water (Ministerio de Medio Ambiente y Agua MMAyA) based on the National Water Balance report (MMAYA *et al.* 2018), and from a technical study of preparation for pre-investment (MMAYA & BID 2021).

DATA AVAILABILITY STATEMENT

All relevant data are available from an online repository or repositories: (<https://developers.google.com/earth-engine/datasets>; <https://search.asf.alaska.edu/#/>; <https://geo.gob.bo/>; <https://www.openstreetmap.org/#map=7/52.154/5.295>; <https://rsis.ramsar.org/>; <https://www.arcgis.com/apps/mapviewer/index.html?layers=cfc7609de5-f478eb7666240902d4d3d>; <https://land.copernicus.eu/en/products/corine-land-cover>; <https://worldcover2020.esa.int/viewer>).

CONFLICT OF INTEREST

The authors declare there is no conflict.

REFERENCES

- AASHTO. 2020 *Guide Specifications for Highway Construction*, 10th edn. American Association of State Highway and Transportation Officials, Washington, DC.
- Allan, J. D. & Castillo, M. M. 2007 *Stream Ecology: Structure and Function of Running Waters*. Second edition. Springer, Cham.
- Alos Palsar. 2023 *Alaska Satellite Facility – Distributed Active Archive Center*. ALOS PALSAR – Radiometric Terrain Correction. Available from: <https://asf.alaska.edu/data-sets/derived-data-sets/alos-palsar-rtc/alos-palsar-radiometric-terrain-correction/>.
- Banerjee, A., Chen, R., Meadows, E. M., Singh, R. B., Mal, S. & Sengupta, D. 2020 *An analysis of long-term rainfall trends and variability in the Uttarakhand Himalaya using Google Earth Engine*. *Remote Sensing* **12** (4). <https://doi.org/10.3390/rs12040709>.
- Bureau of Reclamation 2015 *Bank Stabilization Design Guidelines U.S.* Technical Service Center, Denver, Colorado, Sedimentation and River Hydraulics Group, 86-68240 SRH-2015-25, p. 331.
- Cleveland, C. J. & Ayres, R. U. 2004 *Encyclopedia of Energy*. Elsevier Academic Press, Boston, MA.
- Collins, L., McCarthy, G., Mellor, A., Newell, G. & Smith, L. 2020 *Training data requirements for fire severity mapping using Landsat imagery and random forest*. *Remote Sensing of Environment* **245**, 111839. <https://doi.org/10.1016/j.rse.2020.111839>.
- Congalton, R. & Green, K. 2008 *Assessing the Accuracy of Remotely Sensed Data: Principles and Practices*, 2nd edn. Taylor & Francis Group. <https://doi.org/10.1201/9781420055139>.
- Droogers, P. & van Kempen, C. 2014 *Mapping the Potential for Rainwater Harvesting Under Various Scenarios*. Report. (Vol. 127, p. 17). FutureWater.
- ESRI 2020 *Global Land Cover Map*. Available from: <https://www.arcgis.com/home/item.html?id=cfc7609de5f478eb7666240902d4d3d>
- ESRI 2022 *ModelBuilder*. Available from: <https://pro.arcgis.com/en/pro-app/latest/help/analysis/geoprocessing/modelbuilder/what-is-modelbuilder.htm>.
- ESRI Inc. 2021 *LC 2020 (Mature Support) [Global]*. Available from: <https://www.arcgis.com/home/item.html?id=d6642f8a4f6d4685a24ae2dc0c73d4ac>.
- European Commission, Directorate-General for Research and Innovation. n.d. *Towards an EU Research and Innovation Policy Agenda for Nature-Based Solutions & Re-Naturing Cities*. Final Report of the Horizon 2020. Expert Group on 'Nature-Based Solutions and Re-Naturing Cities'.
- European Environment Agency 2006 *The Thematic Accuracy of Corine Land Cover 2000 Assessment Using LUCAS (Land Use/Cover Area Frame Statistical Survey)*. p. 90. 7/2006. Available from: https://www.eea.europa.eu/publications/technical_report_2006_7.
- European Space Agency 2020 *WorldCover_PVR_v1.0 Version: 1.0 Product Validation Report (D12-PVR)*. p. 38. <https://doi.org/10.5281/zenodo.5571936>.
- Geobolivia 2020. *Geobolivia*. Available from: <https://geo.gob.bo/portal/>.
- Guerrero, P., Haase, D. & Albert, C. 2018 *Locating spatial opportunities for nature-based solutions: A river landscape application*. *Water* **10** (12). <https://doi.org/10.3390/w10121869>.
- Haring, C., Schielen, R., Guy, J., Burgess-Gamble, L., 2021 Chapter 15: Introduction to fluvial systems: Key themes, objectives, gaps, and future directions. In: *International Guidelines on Natural and Nature-Based Features for Flood Risk Management* (Bridges, T. S., King, J. K., Simm, J. D., Beck, M. W., Collins, G., Lodder, Q. & Vicksburg, M. S. eds). U.S. Army Engineer Research and Development Center, Vicksburg, MS, pp. 772–779.
- Hopkins, L. D. 2001 Multi-attribute decision making in urban studies. In: *International Encyclopedia of the Social & Behavioral Sciences* (Smelser, N. J. & Baltes, P. B., eds). Pergamon, pp. 10157–10160. <https://doi.org/10.1016/B0-08-043076-7/04437-5>.
- IPCC 2014 *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. IPCC, Geneva, p. 151.
- Kumar, P., Debele, S. E., Sahani, J., Rawat, N., Marti-Cardona, B., Alfieri, S. M., Basu, B., Basu, A. S., Bowyer, P., Charizopoulos, N., Gallotti, G., Jaakko, J., Leo, L. S., Loupis, M., Menenti, M., Mickovski, S. B., Mun, S. J., Gonzalez-Ollauri, A., Pfeiffer, J., Pilla, F., Pröll, J., Rutzinger, M., Santo, M. A., Sannigrahi, S., Spyrou, C., Tuomenvirta, H. & Zieher, T. 2021 *Nature-based solutions efficiency evaluation against natural hazards*. *Modelling Methods, Advantages and Limitations* **784** (147058), 1–27. <https://doi.org/10.1016/j.scitotenv.2021.147058>.
- Landis, J. R. & Koch, G. G. 1977 *The measurement of observer agreement for categorical data*. *Biometrics* **33** (1), 159–174. JSTOR. <https://doi.org/10.2307/2529310>.
- Malesu, M., Khaka, E., Mati, B., Oduor, A., de Bock, T., Nyabenge, M. & Oduor, V. 2006 *Mapping the Potentials for Rainwater Harvesting Technologies in Africa: A GIS Overview on Development Domains for the Continent and Nine Selected Countries*. Technical Manual No. 7. WorldAgroforestry Centre (ICRAF), Netherlands Ministry of Foreign Affairs, Nairobi, Kenya.
- MMAYA & BID 2021 *Pre-investment Technical Design Study 'Construction Works for the Reduction of Risks and Adaptation to Climate Change – Defensive District I (Tojo) Yunchará Municipality'*. Environment and water ministry (MMAYA), p. 157.
- MMAYA, Stockholm Environment Institute (SEI), Servicio Nacional de Meteorología e Hidrología SENAMHI, Hydrology & Hydraulic Institute of San Andres Major University IHH-UMSA, & Hydraulic Laboratory of San Simón Mayor University LH-UMSS 2018 *Bolivian Surface Water Balance (Diffusion Document)*. p. 60.

- Mubeen, A., Ruangpan, L., Vojinovic, Z., Sanchez Torrez, A. & Plavšić, J. 2021 **Planning and suitability assessment of large-scale nature-based solutions for flood-risk reduction**. *Water Resources Management* **35** (10), 3063–3081. <https://doi.org/10.1007/s11269-021-02848-w>.
- RAMSAR 2010 *Water-related Guidance: An Integrated Framework for the Convention's Water-Related Guidance*. Ramsar Handbooks for the Wise use of Wetlands, 4th edn. Vol. 8. Ramsar Convention Secretariat, Gland, Switzerland.
- Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M. & Rigol-Sanchez, J. P. 2012 **An assessment of the effectiveness of a random forest classifier for land-cover classification**. *ISPRS Journal of Photogrammetry and Remote Sensing* **67**, 93–104. <https://doi.org/10.1016/j.isprsjprs.2011.11.002>.
- Ruangpan, L., Vojinovic, Z., Di Sabatino, S., Leo, L. S., Capobianco, V., Oen, A. M. P., McClain, M. E. & Lopez-Gunn, E. 2020 **Nature-based solutions for hydro-meteorological risk reduction: A state-of-the-art review of the research area**. *Natural Hazards and Earth System Sciences* **20** (1), 243–270. <https://doi.org/10.5194/nhess-20-243-2020>.
- Ruangpan, L., Vojinovic, Z., Plavšić, J., Doong, D.-J., Bahlmann, T., Alves, A., Tseng, L.-H., Randelović, A., Todorović, A., Kocic, Z., Beljinac, V., Wu, M.-H., Lo, W.-C., Perez-Lapeña, B. & Franca, M. J. 2021 **Incorporating stakeholders' preferences into a multi-criteria framework for planning large-scale nature-based solutions**. *Ambio* **50** (8), 1514–1531. <https://doi.org/10.1007/s13280-020-01419-4>.
- Singh, R., Arya, D. S., Taxak, A. & Vojinovic, Z. 2016 **Potential Impact of Climate Change on Rainfall Intensity-Duration-Frequency Curves in Roorkee, India**, *Water Resources Management*. pp. 1–14. 2016/7/26. <https://doi.org/10.1007/s11269-016-1441-4>.
- Somarakis, G., Stagakis, S. & Chrysoulakis, N. 2019 *Thinknature Nature-Based Solutions Handbook*. ThinkNature Project Funded by the EU Horizon 2020 Research and Innovation Programme.
- Tsendbazar, N.-E., Tarko, A., Li, L., Herold, M., Lesiv, M., Fritz, S. & Maus, V. 2021 *Copernicus Global Land Service: Land Cover 100m: Version 3 Globe 2015–2019: Validation Report (Dataset v3.0, doc issue 1.10)*. Zenodo. <https://doi.org/10.5281/zenodo.4723975>.
- Vojinovic, Z., Alves, A., Gómez, J. P., Weesakul, S., Keerakamolchai, W., Meesuk, V. & Sanchez, A. 2021 **Effectiveness of small- and large-scale nature-based solutions for flood mitigation: The case of Ayutthaya, Thailand**. *Science of The Total Environment* **789**, 147725. <https://doi.org/10.1016/j.scitotenv.2021.147725>.
- Watkin, L., Ruangpan, L., Vojinovic, Z., Weesakul, S. & Sanchez, A. 2019 **A framework for assessing benefits of implemented nature-Based solutions**. *Sustainability* **11** (23), 6788. MDPI. <https://doi.org/10.3390/su11236788>.
- World Bank 2008 *Biodiversity, Climate Change, and Adaptation: Nature-Based Solutions from the World Bank Portfolio (English)*. (Vol. 46726; p. 112). World Bank Group. Available from: <http://documents.worldbank.org/curated/en/149141468320661795/Biodiversity-climate-change-and-adaptation-nature-based-solutions-from-the-World-Bank-portfolio>.
- Xiong, J., Li, J., Cheng, W., Wang, N. & Guo, L. 2019 **A GIS-based support vector machine model for flash flood vulnerability assessment and mapping in China**. *ISPRS International Journal of Geo-Information* **8** (7). <https://doi.org/10.3390/ijgi8070297>.
- Yazar, A., Ali, A., 2016 Water harvesting in dry environments. In: *Innovations in Dryland Agriculture* (Farooq, M. & Siddique, K. H. M., eds). Springer International Publishing, pp. 49–98. https://doi.org/10.1007/978-3-319-47928-6_3.

First received 11 August 2023; accepted in revised form 7 November 2023. Available online 15 November 2023