Towards automation of river water surface detection

Stefano Conversi[1], Daniela Carrion[1], Francesco Gioia[1], Alessandra Norcini[2], Monica Riva[1]

[1] Dipartimento di Ingegneria Civile e Ambientale, Politecnico di Milano, Piazza Leonardo Da Vinci 32, 20133 Milano, Italy - (stefano.conversi, daniela.carrion, monica.riva)@polimi.it, francesco1.gioia@mail.polimi.it
[2] Struttura Natura e Biodiversità, Unità Organizzativa Parchi, Biodiversità e Sistema delle conoscenze, Direzione Generale Territorio e Sistemi Verdi, Regione Lombardia, Piazza Città di Lombardia 1, 20124 Milano, Italy – alessandra_norcini@regione.lombardia.it

Keywords: Remote Sensing, GeoAI, Sensor fusion, BMax Otsu, Drought monitoring, Google Earth Engine

Abstract

European rivers are increasingly impacted by frequent and lasting dry periods, with consequences on jeopardized ecosystems and local economies. Tools for monitoring the evolution of such impacts may be profitable exploited by public administration to assess environmental conditions and draw safeguard policies. This work presents the evolution of a methodology which integrates optical and radar imagery, by Copernicus Sentinel constellations, to map river water surfaces. Despite the base methodology being developed as a man-supervised classification, with necessity for the user to manually define training polygons, the proposed advancements will allow the system to automate training sample extraction. The process is based on the realization of binary masks, originated by processing optical and radar imagery with a BMax Otsu algorithm for image segmentation. The masks are then furtherly refined to obtain a reliable set of classified pixels, from which the training samples are extracted. A sensitivity analysis is performed for assessing the optimal amount of pixels to be considered, with respect to the total area of interest. Furthermore, the performances of several Machine Learning supervised classification algorithms are compared, leading to the selection of the best algorithm to be considered for future developments of the methodology.

1. Introduction

1.1 Drought Events and Impacts on European Large Rivers

It is well known that climate change impacts are increasingly affecting European territory, often in the shape of extreme natural events. Among those, in recent years, heat waves due to global warming contributed to the growth of intensity of dry periods (Dai, A., 2011). Changing the temporal reference, in future climates the severity and frequency of drought events is expected to increase, due to the processes related to global warming (Mukherjee et al., 2018). Particularly, the Mediterranean areas are expected to face extraordinary hot summers and increasingly frequent drought events, which may clearly affect the population (European Environment Agency, 2021). Multiple examples could be presented, such as Loire River (France) and Ebro River (Spain) cases, for which the average water discharge was measured to be under average values for 11 months in the year (Copernicus, 2022). As mentioned, such phenomena eventually lead to huge consequences on several river ecosystem-related activities, as presented by Nikolova (2022) in the case of Danube. Impacts were reported in terms of economical losses, as an example due to limitations that vessels experienced because of water level decrease, but also of ecological damages. Bulgarian WWF, reports Nikolova (2022), stated the depletion of 90% of the records of autochthonous species' eggs. A further confirmation of the delineated crisis comes from the observations of Copernicus Climate Change Service on European rivers discharge data (Copernicus, 2022). The presented analyses clearly show the effects of an exceptional drought event in 2022 (that actually lasted until the first months of 2023), which had consistent impacts on large rivers and their ecosystems, as summarized in Figure 1. Another key example, in this case related to medium width rivers, is the one of Po River (the longest Italian river), for which in 2022 the worst water scarcity of the past two centuries was recorded (Montanari et al., 2023). Experts agreed on the exceptionality of the phenomenon, stating nevertheless the repeatability of such events in near future (Bonaldo et al., 2022).

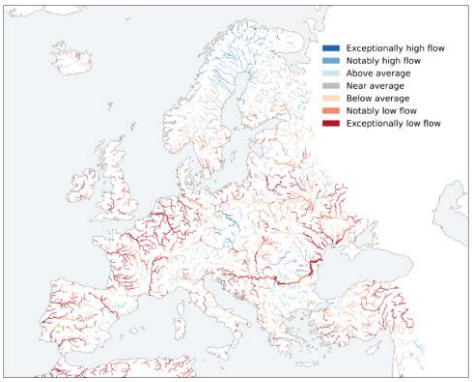


Figure 1. Monthly average river discharge anomalies across all 12 months in 2022 – August, from (Copernicus, 2022).

The increments of severity and frequency of such events highlight the necessity for local authorities to be capable of monitoring the state of rivers and, if needed, promptly enacting countermeasures to face the related issues on their territory. This paper will present the last findings of a procedure intended support to public administration in river drought monitoring, which was, in the first phases of the work, triggered by *Regione Lombardia*, one of the the main regional authorities of Northern Italy. Specifically, the authors partnered with the public authority in the drafting of a procedure devoted to integrating optical and radar satellite imagery, aimed at combining two sensors for mapping water, overcoming the typical issues of the single data sources (Conversi et al., 2023), which will be

discussed in the following section. The current work embeds the efforts made by the team towards the automation of the procedure, aimed at making it more user friendly and, most of all, profitably employable in *Regione Lombardia*'s services for environmental monitoring. In addition, a process of selection of the best performing Machine Learning algorithm is proposed, considering the application within the advanced version of the methodology.

1.2 Reference Methodology: Integration of Optical and Radar Imagery to Enhance River Drought Monitoring

Before introducing the latest evolutions of the work, it may be of interest to summarize the principal characteristics of the first procedure proposed in (Conversi et al., 2023), schematized in Figure 2.

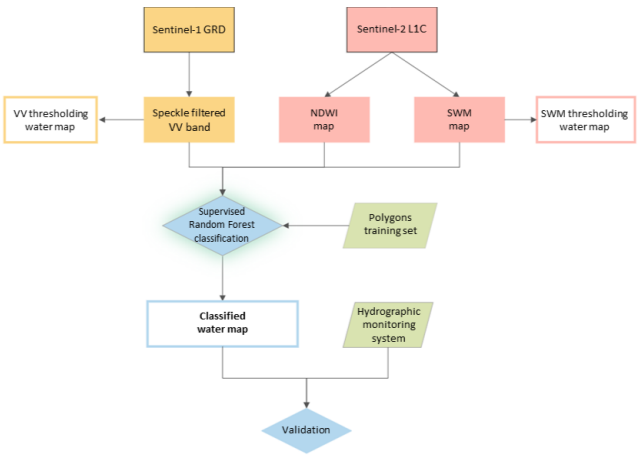


Figure 2. Workflow of optical and radar imagery integration for mapping river water surface, proposed in (Conversi et al., 2023).

The main goal of the work is providing the public administration with an efficient tool for monitoring drought impact on rivers. It was pursued by obtaining a system capable of delivering a water coverage map from satellite imagery referred to a short interval of time, under the assumption of considering the variation of water surface itself as a good proxy of impacts experienced by rivers in dry periods. As summarized by the previous figure, two satellite data sources were identified, one for optical imagery (ESA's Copernicus Sentinel-2 constellation) and the other for radar imagery (ESA's Copernicus Sentinel-1 constellation).

As it is well known, image segmentation techniques allow to extract objects or regions from images, distinguishing between pixels belonging to the interest class and pixels classifiable as background (Huang, 2021). Such techniques may profitably be used for identifying (and extracting) water surface pixels from both radar and optical imagery. Nevertheless, the two families of sensors bring with them intrinsic limitations that may reduce the quality of classification results. Optical imagery suffers the presence of clouds (or whichever obstacle prevents the full vision of the object), leading to loss of information on the territory and, in the case of interest, to the underestimation of water surface. On the other hand, imagery sensed by Synthetic Aperture Radar (SAR) can be prone to misclassification when used for water pixels detection, due to several conditions, ranging from soil moisture to background noise (Landuyt et al., 2019). Anyway, the work proved the possibility of overcoming these errors by integrating the two data sources into a unique classification process. The base process and the modifications oriented at its automation presented in this paper, were built in Google Earth Engine environment, the cloud-based service for global geospatial analyses powered and owned by Google (Gorelick et al., 2017). Within this environment, codes can be written in JavaScript language, and run, obtaining outcomes in forms of values, graphs or maps, which can be displayed in an inbuilt map visualizer. The tool consents to upload local data and manage them within the coded geoprocessing as well, allowing then to export and download the results in various formats. Furthermore, it also offers the possibility to explore a *Data catalog* (GEE, 2024), that contains lots of opensource geospatial resources, among which satellite imagery, such as the Sentinel-1 and Sentinel-2 collections used in this study.

Input imagery was selected to ensure complete coverage of the area of interest, with mosaicking if necessary images coming from multiple close dates, a reliable assumption considering that the drought is usually a slow phenomenon. The interval of time between images was minimized by the code, depending on data quality and availability. Once the datasets were retrieved from the catalogue and preprocessed, their integration was performed using a Random Forest (RF) Machine Learning algorithm for supervised classification, deemed as particularly suitable for the case because of its capability of classifying pixels characterized by unknown frequency and distribution (De Fioravante et al., 2021). The RF algorithm was then fed with two images representing spectral indices (Normalized Difference Water Index and Sentinel Water Mask), (JRC, 2011; Milczarek et al., 2017), elaborated from orthorectified Sentinel-2 level 1C, and an image derived from Sentinel-1 (L1 GRD, VV band, processed with speckle filtering). In addition, a set of polygons representing "water" pixels and "background" pixels was drawn by the author by photointerpretation, in order to train the algorithm with reference data. The procedure was then applied to a real case study (Po River, in between Piacenza and Spessa Po sections) and a multitemporal analysis conducted over selected dates in a period of 8 years. The outcomes of the procedure were constituted by a map of water surface detected over the area of interest, for each of the considered time intervals, complemented with an estimate of the extent in km^2. The results showed a satisfactory level of accuracy in classification and, most of all, a good level of agreement with hydrometric data recorded on the field, proving the capability of the system in matching the real trend of river behaviour (Conversi et al., 2023).

2. Towards Automation of Ground Truth Extraction

On the basis of the overall quality of the methodology results, it was of interest to the whole team to understand how the procedure itself could be enhanced to make a step towards the actual usage within the context of public authorities' policies. It is evident that one of the main limitations to this process is the necessity for the user of providing to the Machine Learning (ML) algorithm training polygons of reference, which have to be drawn for each iteration of the procedure and must be specified for the different cases that may be encountered. As an example, the polygons representative of water pixels for a certain time window may not be correctly usable for the analysis on the same area, but in a different period, thus forcing the user to re-identify and draw a new set of reliable polygons. This part of the procedure, based on photointerpretation and analysis of the whole set of imagery (optical indices and SAR) obviously requires a certain level of understanding of the topic

and, most importantly, the mastery of the used instruments. Public administration operators may not be experts in satellite imagery analyses nor in GIS or GEE use. Thus, this led to the recent development of the methodology itself, which is oriented to the complete automation of the steps.

2.1 Otsu and Bmax Otsu Algorithms for the Identification of a Threshold From Bimodal Histograms

The code was integrated with an automatic thresholding method, named after its author, the Otsu method (Otsu, 1979). This methodology allows to define an optimal threshold value for distinguishing background from foreground objects. In mathematical terms, Otsu algorithm can be defined as an iterative algorithm that evaluates the variance for different thresholds k, selecting the k* that maximize a parameter λ, corresponding to the ratio between class variance (to be maximized) and intra-class variance (to be minimized). In this application the concept of class has to be interpreted as subset of values representing grey levels, each of which is associated to a certain probability of occurrence (Otsu, 1979). Aiming to state the potential of the method, its simplicity has to be stressed (it does not require many parameters to be run), it does not suffer from instability due to local properties of the pixels' histogram distribution (because it is based on integration and not on differentiation), (Otsu, 1979). Furthermore, it can be exploited not only on incident radiation, but also on computed indices, as it will be shown in the next paragraph. Otsu methods found through the years lots of different applications, in diverse fields, that only share the necessity of image segmentation to draw conclusions, such in the case of medical diagnoses (Bindu & Prasad, 2012) and monitoring of infrastructural damages (Akagic et al., 2018). The reduction of computational time granted by the Otsu algorithm in real life applications, jointly with its flexibility fostered the development of a family of methods aimed at improving the original one (Goh et al., 2018). Among these, a particular role is played by Bmax Otsu algorithm (Cao et al., 2019), which is oriented to mitigate one of the main issues of the main algorithm, i.e. its dependency on the bimodality of processed images. In facts, in real case studies and procedures, such as the ones proposed in this work, it is rare to be able to operate on images presenting a clearly bimodal histogram of values. BMax Otsu was originally developed for Sentinel-1 imagery analysis and then it has been proved to be generally employable for images characterized by multiple classes or complex backgrounds (Markert et al., 2020). It is based on a sequence of processing steps meant to provide, disregarding the peculiarities of the analyzed image, a bimodal histogram to a classic Otsu. To initialize the method, a first threshold should be provided, so to estimate each class variance; then the image is subjected to a process of chessboard segmentation, which divides it in cells of user-defined dimensions (Cao et al., 2019). A bimodality test is then performed: for each cell the maximum normalized Between-Class Variance (BVC) is evaluated and only the cells exceeding a certain level are selected (Markert et al., 2020), as exemplified in Figure 3. Usually, a grid cell is considered to be bimodal when the BVC (or Bmax) reaches values of 0.65-0.75 (Cao et al., 2019). The so-sampled cells, along with their histogram values distribution, are used to derive the final threshold by the mean of a classic Otsu procedure.

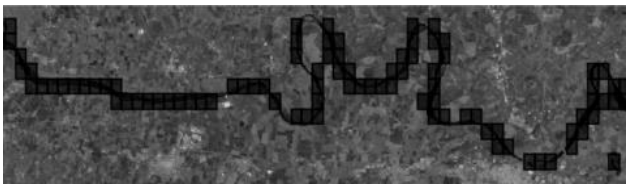


Figure 3. Chessboard segmentation obtained through BMax Otsu algorithm implementation, bimodal cells are highlighted.

2.2 Otsu-based Extraction of Training Samples

Figure 4 depicts the workflow of the training samples extraction.

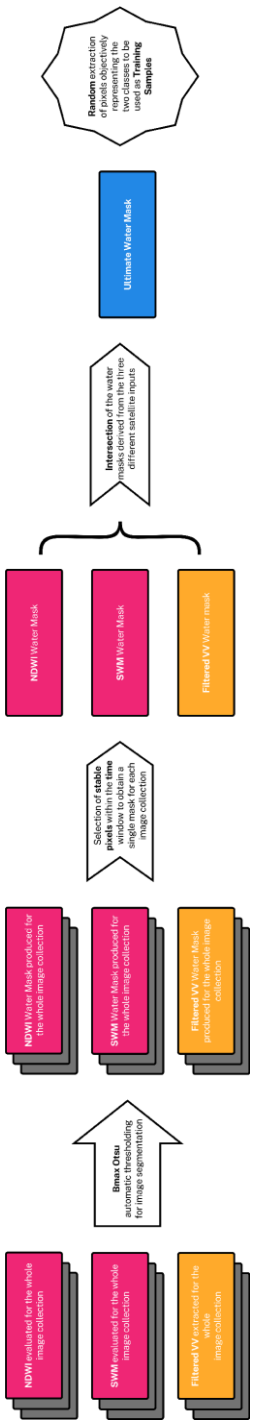


Figure 4. Scheme of automated training samples extraction.

As shown on the previous scheme, the implementation of Bmax Otsu algorithm plays a crucial role in the identification of training samples necessary to the correct deployment of the classification procedure. As mentioned, the main concept behind its use is related to the possibility of removing the contribution of a "human" user in the selection of areas (previously identified as training polygons) that are representative of the two class of pixels (namely "water" and "background"). Aiming to automate the process it is relevant that the code can produce autonomously a mask of pixels for each of the classes, containing only pixels that are reliably belonging to the single class of interest. Clearly, removing the role of human eye and expertise, it would be difficult to obtain masks containing only pure pixels for the two classes, but with a sort of step by step procedure for masks production, reasonably good results can be achieved
. First, the Bmax Otsu algorithm was applied separately to the three imagery inputs: NDWI, SWM, SAR filtered VV band. The initial estimates of the thresholds of the two indices were identified in common values used for differentiating water from non-water surfaces, as suggested by (McFeeters, 1996) and (Milczarek et al., 2017). For SAR, instead, a calibration was performed, and the most common value was selected, also in agreement with literature values for studies devoted to water detection through radar sensors (Carreño & De Mata, 2019). The grid size necessary for the application of the method, as introduced in the previous section, has been defined based on the river width. It was considered as a linearly increasing function (1), with the width calibrated so to maximize the similarity with a bimodal histogram distribution.

$$\text{Grid size} = \text{River width [m]} \cdot 0.008/250, \quad (1)$$

The calibration tests were performed with the goal of obtaining a selection of cells (on which the Otsu algorithm would be run) that contain the same quantity of water and non-water pixels, at least as a good proxy. As it is shown by Figure 5, in the cells that are extracted, thus characterized by high bimodality, half of the area is covered by the river and the remaining part is made of background pixels. It is obvious that this formulation should be intended as an approximation and was built under the hypothesis of not considering distortional effects of the conformal map on the distances, as one is going far from the equator.

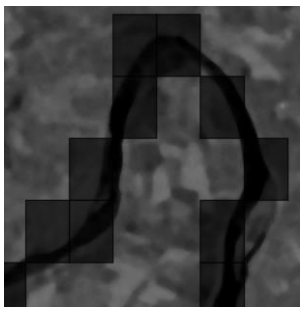


Figure 5. Detail of bimodal cells extracted through BMax Otsu algorithm implementation.

By applying Bmax Otsu algorithm, the code is then capable of individuating for each of the images a specific threshold value that splits the histogram of values associated with water and values associated with background. Applying a segmentation based on the thresholds, three rough water masks are produced over the image collections; once again, it is not possible to completely trust these results, but they can be considered a good proxy of the presence of water in the area of interest. In order to refine the obtained masks, a procedure was built, capable of going through the images found for the reference time and extracting from them only the pixels that remain constantly categorized as water (or background) for the whole period. The further and last refinement of the mask is then performed by extracting from the three smoothed masks only the pixels that are actually classified as water (and background) in all three of them (NDWI, SWM, VV). The ultimate water mask can then be addressed as the intersection among the three mentioned smoothed masks. This discretized procedure allows to increase the refinement in subsequent steps, so to obtain at the end a mask of water and background pixels that is as much as possible representative of the real situation on the ground in the area of interest, all over the reference period. It is clear that the amount of pixels in these masks is strongly dependent on the time reference, which should be significant with respect to the phenomenon that is to be observed. The machine trained on points randomly extracted from the mask will then be used for mapping the water surface in the zone of interest, allowing to estimate the water coverage in the areas surrounding the river.
In Figure 6 a portion of the area of interest is depicted: in yellow it is possible to identify the areas composed by the pure water pixels obtained from the ultimate mask (in yellow) and the background ones (in white), from which the corresponding training points are extracted, respectively in light blue for water and in red for background.

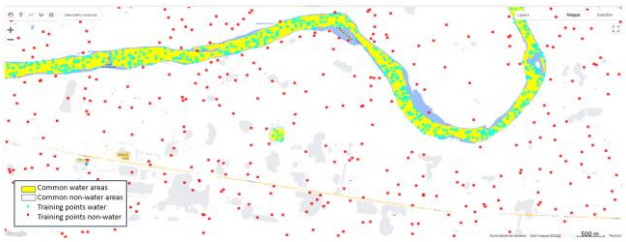


Figure 6. Detail of the ultimate water wask with superimposed extracted training points.

It is relevant to state that the number of training points to be extracted from the mask has a relevant impact on the overall quality of the water mapping results. Thus, a sensitivity analysis was performed, in order to define the optimal solution in terms of training sample size. At the same time, it was also carried out a second analysis, for comparing different Machine Learning algorithms, so to select the classification algorithm with the best performance. In this case, a fixed amount of training pixels analysis was considered, in accordance with the results of sensitivity analysis.

3. Application of the Automatic Training Sample Extraction and Machine Learning Classification Algorithm Selection

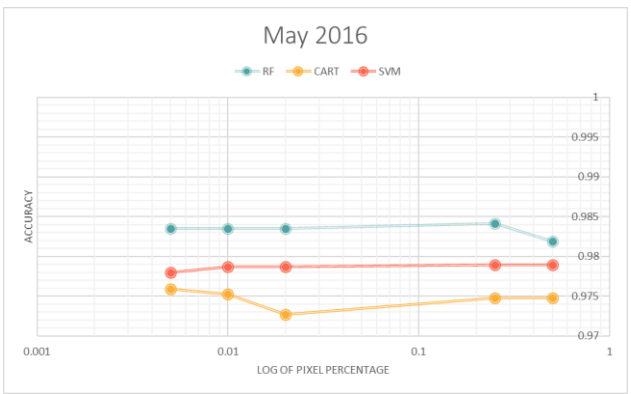
A sensitivity analysis and a classification method comparison were performed. The sensitivity analysis was devoted to the study of accuracies variation when changing the number of training sample points. Then, the results obtained with different ML algorithms were compared, testing them both on the integration of optical and radar imagery and on the single data sources considered separately. Doing so, it was possible to select the best performing algorithm and draw useful conclusions upon the data fusion concept, which is the core of the study.
In both cases the classification accuracy was estimated with the confusion matrix, built from each iteration of the procedure. As

it is well known, in case of a binary classification such the one addressed in this work, the confusion matrix will be structured as a 2x2 matrix, containing values of True Positive (TP), False positive (FP), False Negative (FN) and True Negative (TN). Obviously, such a framework requires a preliminary knowledge on the conditions on the ground, so to make the system able to recognize whether if the value prediction for each cell is correct or wrong. In the sensitivity analysis, devoted to training sample size selection, the case study of the stretch of river delimited by Spessa Po and Piacenza sections was considered. In fact, for this area, reference data were available with a good distribution known water/non water pixels location, originally estimated by photointerpretation. The intercomparison of the classification algorithms was instead performed in a different study area, without prior ground truth knowledge, considering as reference map the one obtained with supervised Random Forest classification with photointerpreted polygon training samples.
For both the aforementioned analyses the accuracy estimation was carried out exploiting three different Machine Learning classification algorithms, so to derive useful information on the methodology applicability:

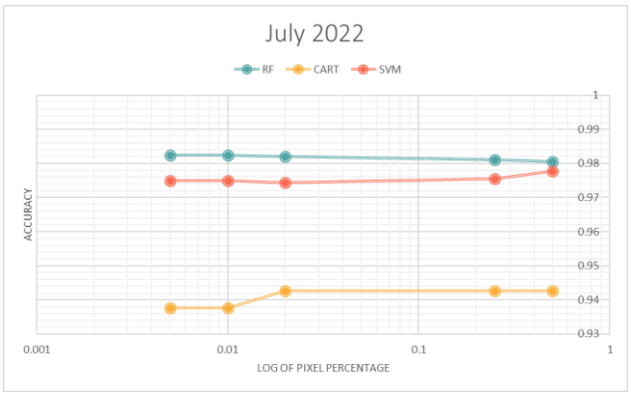
Random Forest; as it is well-known, it is a Machine Learning algorithm, originally introduced by Breiman (2001), which bases its analyses on several tree-like structures, named decision trees. A decision tree is built by recursively splitting the feature space of the training set until a predetermined stopping condition is met, having the goal of finding a set of decision rules that naturally partition the feature space, providing a robust hierarchical classification. It breaks down a dataset into smaller subsets while, at the same time, the decision tree is incrementally built. The final result is a tree with decision nodes and leaf nodes, where decision nodes have two or more branches and leaf nodes will define a final decision. Basing on bagging algorithm, in RF random samples sets are trained independently (collection of decision trees), with each one of the trees giving an independent answer to a specific task. The majority of the results will then define the output classification of the single pixel (Schonlau & Zou, 2020).
Classification And Regression Tree algorithm (CART); it is a different kind of decision tree algorithm, with a difference in the way in which the decision nodes are split. In CART algorithm, the threshold value is selected based on the best homogeneity for the sub-nodes, using the Greedy Splitting (Bittencourt & Clarke, 2003). In Greedy Splitting, the decision nodes are split based on numerical procedure where different split values are tried using a cost function, the value that minimizes the cost is then chosen as threshold.
Support Vector Machine (SVM); it is a supervised, non-parametric, statistical learning algorithm. SVM is particularly used in remote sensing because it can perform good with small training sets, often having a better accuracy with respect to other methods (Mountrakis et al., 2011). The algorithm relies on hyperplanes to separate the dataset into a discrete number of classes, hyperplanes are used in order to minimize misclassifications, optimizing the segmentation. It is a learning algorithm, meaning that there is an iterative process in finding the optimal threshold. In SVM pixels are represented as pattern vectors, where for each considered band, numerical measurements are performed (Mountrakis et al., 2011).

3.1 Sensitivity Analysis: Estimation of Training Samples Number

As anticipated in the previous section, the estimation of an optimal number of training samples depends on several factors, related to the specific parameters characterizing the processed imagery (number of bands, image complexity, etc.). Generally, an increment in the number of training pixels increases the accuracy, but in case of large datasets it can increase as well the opportunity of including spectral inhomogeneities, which can hamper the results quality (A. T. Joyce, 1978).
For each of the introduced supervised classification algorithms a set of simulations was performed, for different time intervals, considering variable amount of pixels (intended as a percentage of the total area of interest): 0.005%, 0.01%, 0.02%, 0.25% and 0.5%. The large variability of the proposed values was determined in order to be able to observe a diverse set of results, so to seize the optimal response that mitigates the impacts of the different factors. The validation was carried out considering a set of 5000 points distributed over the area of interest, representing both water and non-water pixels, classified by photointerpretation. Validation accuracies and True Positive Rate (TPR), (Ting, 2011), were evaluated for each of the Machine Learning algorithms over three different periods of time, aiming to guarantee to the analysis a good variability of conditions (e.g. a date interval during 2022 drought was considered). The plots shown in Figure 7 depict behaviours in terms of accuracy variation that are proper of the single algorithms. In facts, Random Forest seems to loose accuracy when the number of training samples is increased, while Support Vector Machine represents the opposite trend (increasing the amount of samples, the accuracy reaches higher values). On the other hand, Classification And Regression Tree algorithm results in a sort of mixed trend, with an initial drop in accuracies, recovered as long as the number of training samples increases. It should be noted that the overall accuracies report quite high values in all the simulations, even though the analyses of TPR showed CART has a tendency to the overestimation of water pixels, with larger amounts of false positives than the other algorithms.



(a)



(b)

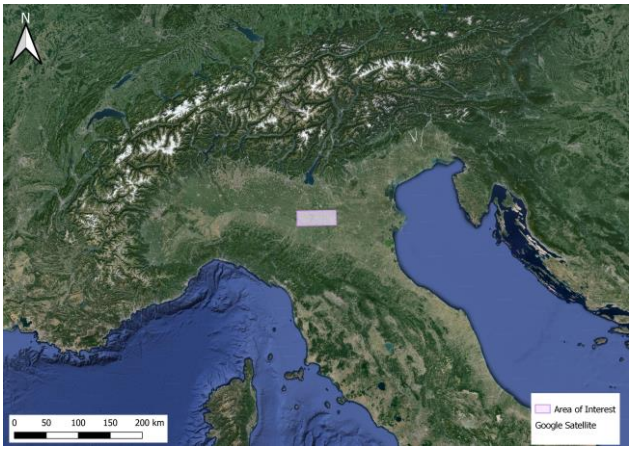
(c)

Figure 7. Sensitivity analysis plots. Variation of validation accuracy for varying amount of traing samples. Aiming to enhance the readability of the different trends, it was decided not to uniform y-axis minima. Results shown for RF, CART and SVM algorithms in (a) May 2016, (b) July 2022, (c) February 2023.

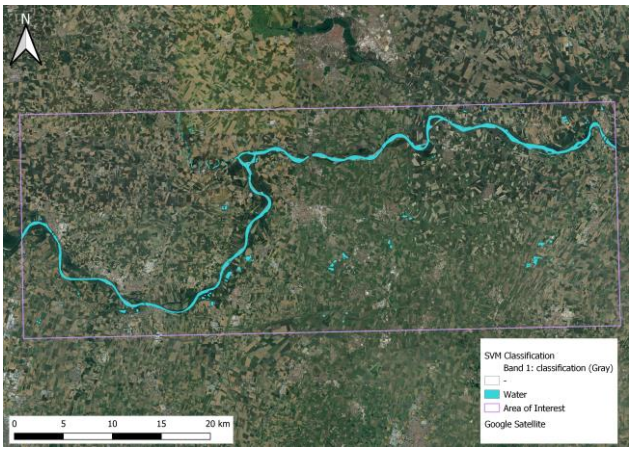
All in all, aiming to find a balance among the illustrated trends and the unavoidable increase of computational time needed for large set of training samples, it was decided to set the amount to 0.15% of the area of interest. Indeed, it can be noticed that in between 0.1% and 0.25% simulations, all of the trends reach an optimal threshold value after which not significant changes are appreciable.

3.2 Selection of Machine Learning Algorithm With the Best Performance

In order to test the applicability of the procedure to different territories, the simulations oriented to the selection of the ML classification algorithm with the best performance were conducted on a different area of interest with respect to the test case used in the sensitivity analysis. This was done also to ensure as much as possible the independency of this process from the already discussed sensitivity analysis.
Po River was still taken as a reference, but a different stretch was selected. Once again, it was decided to focus the analysis on a portion of the river in between two instrumented sections, from which it would be possible to derive hydrometric records for further checking the results (Conversi et al., 2023). The two ends of the chosen river stretch are Borgoforte (Mantova province) and Casalmaggiore (Cremona province), two municipalities belonging to *Regione Lombardia*'s area of relevance (AIPO, 2024); the contextualization of the area and a water surface map produced over the area through the presented methodology is proposed respectively in Figures 8a and 8b.



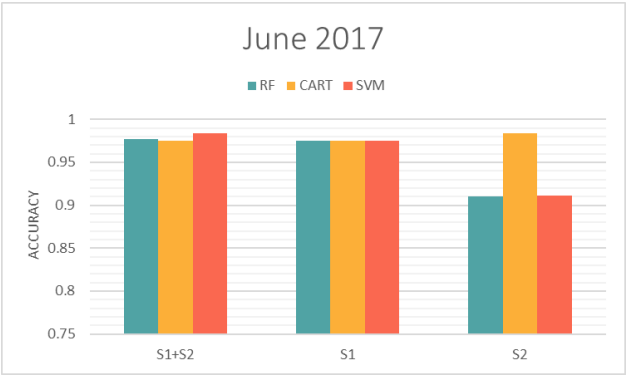
(a)



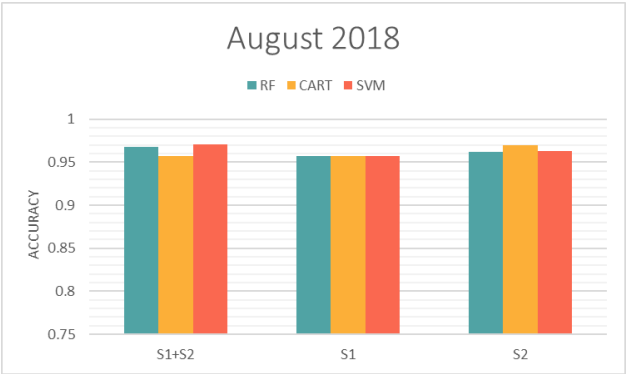
(b)

Figure 8. Contextualization of the selected area of interest, surrounding Po River stretch in between Borgoforte (Mantova province) and Casalmaggiore (Cremona province), (a) and output of the water surface mapping procedure in the region (b). Input imagery acquired between in September 2021 (1^{st} – 6^{th}).

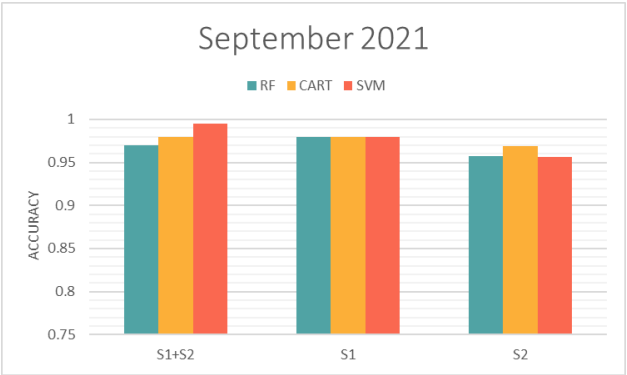
Aiming to establish a standardized procedure, a fixed number of randomly extracted training samples was set, and it was identified in the amount of 20877 pixels, corresponding to the 0.15% of the area of interest, as resulted from the sensitivity analysis. Addressing the temporal scale, five different time spans were explored, so to furtherly explore the capability of the methodology in different weather conditions; considering the years for which all of the needed imagery was available, date intervals from summer, autumn and late winter were chosen. In addition to this, within the process of ML algorithm selection, the options of considering as an input images coming from single sensors only (optical or SAR) were considered separately.
Figure 9 represents the results in terms of accuracy for the considered time periods, comparing the performances of Random Forest, Classification and Regression Tree Algorithm and Support Vector Machine applied in the case of sensor fusion (S1+S2), SAR imagery (S1) and optical imagery (S2).



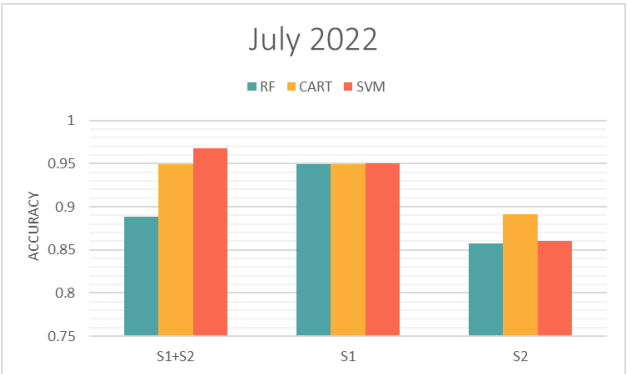
(a)



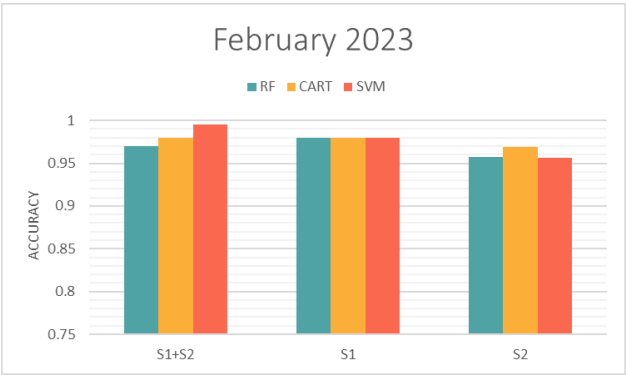
(b)



(c)



(d)



(e)

Figure 9. Comparison of accuracies among RF, CART and SVM classification algorithms. Single sensors and sensor fusion performances are compared in (a) June 2017, (b) August 2018, (c) September 2021, (d) July 2022, (e) February 2023.

Comparing the outcomes, it is recognizable that the employment of satellite data fusion is indeed a fruitful solution, as expected. In few cases actually the results achieved by the use of a single data source are comparable with or outperforming the use of integrated imagery. However, even if a stable and univocal behaviour cannot be identified, the optical and SAR combination always provides comparable or better results with respect to the optical or SAR alone. In general, the best performance is obtained with optical and SAR combined, processed with the Support Vector Machine algorithm.

4. Conclusions

The present paper proposes an innovation of a previously proposed procedure (Conversi et al., 2023), built in Google Earth Engine, devoted to the integration of optical and SAR imagery to detect river water surfaces. In particular, it fosters the possibility of automating the process, so to make the whole system user friendly and compatible with its employment in public administration procedures. Indeed, the work was triggered by the necessity expressed by a regional authority of a tool for monitoring the impacts of drought period on medium-width rivers.

The procedure is focused on the supervised classification of satellite-sensed images (NDWI and SWM indices derived from Sentinel-2 imagery and VV band extracted from Sentinel-1); the user is then required to draw by photointerpretation some polygons representing ground truth to train the algorithm. A step towards the complete automation procedure is taken by the mean of introducing into the code a technique for the automatic extraction of training samples from each of the processed images. This procedure is built on the base of the Bmax Otsu algorithm for automatic thresholding, that allows the system to reach autonomously binary masks representing water and non-water classes for each considered band. The methodology will then process all the masks to obtain an Ultimate Mask that is representative of the situation on the territory and that can be objectively trusted as a reference source of information for water presence. Training samples are then extracted from it randomly.

A sensitivity analysis was performed on the Po River to identify the percentage of the area of interest to be considered as training set and the value of 0.15% was selected. The study proposes also the outcomes of a further analysis, based on a different stretch of the same river, whose goal is to define which is the best performing supervised classification algorithm among

Random Forest, Classification And Regression Tree, and Support Vector Machine. The latter appears to be the best solution; the same analysis shows also that the combination of satellite data sources allows the process to reach better accuracies than the single sensor approaches.
The results of the study are promising, with particular regard to the possibility of optimizing the process and reach the complete automation of the classification method. Nevertheless, it should be reminded that at the moment the code is not yet totally accessible by a non-expert user. Indeed, apart from the obvious requirement of being able to manage and run a JavaScript code, there are still few parameters that must be set and inserted by hand, such as the estimate of average river width, necessary for different steps, among which the tuning of Bmax Otsu algorithm.
Concluding, the experience of this work proves that there are good possibilities of reaching the automation of such procedures, without sacrificing the classification accuracy. Even though some issues still need to be faced and solved, the authors will take advantage of data, information and results achieved in the current study to explore the feasibility of converting the methodology in a ready-to-use tool (e.g. in the shape of a WebApp) for supporting public administration's policies in environmental monitoring and climate risk mitigation.

Acknowledgements

This work is developed within the *Innovazione* project (2022) by Regione Lombardia in partnership with Politecnico di Milano. The research was triggered and constantly supported by Struttura Natura e Biodiversità, belonging to Regione Lombardia Direzione Generale Territorio e Sistemi Verdi.
The Google Earth Engine source code of the illustrated methodology can be retrieved at: https://code.earthengine.google.com/375c50fb05615b83445b441d4f4c573d?noload=true.
The maps proposed in the work contain modified Copernicus Sentinel data [2021 – 2022].

References

AIPo, 2024. Geoportale. *Agenzia Interregionale per il Fiume Po*. Last visited April 28 2024, on: http://geoportale.agenziapo.it/web/index.php/it/?option=com_ajpografd3.

Akagić, A., Buza, E., Omanovic, S., Karabegović, A., 2018. Pavement crack detection using Otsu thresholding for image segmentation. *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, 1092-1097.

Bindu, H. & Prasad, K.S., 2012. An Efficient Medical Image Segmentation Using Conventional OTSU Method. *International Journal of Advanced Science and Technology*. 38.

Bittencourt, H.R., & Clarke, R.T., 2003. Use of Classification and Regression Trees (CART) to Classify Remotely-Sensed Digital Images, in *International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 3751–3753. https://doi.org/10.1109/igarss.2003.1295258.

Bonaldo, D., Bellafiore, D., Ferrarin, C., Ferretti, R., Ricchi, A., Sangelantoni, L., Vitelletti, M.L., 2022. The summer 2022 drought: a taste of future climate for the Po valley (Italy)?. *Reg Environ Change* 23, 1 (2023). https://doi.org/10.1007/s10113-022-02004-z.

Breiman, L., 2001. Random Forests, Kluwer Academic Publisher, *Machine Learning*, 45, 5-32.

Cao, H., Zhang, H., Wang, C., Zhang, B., 2019. Operational flood detection using Sentinel-1 SAR data over large areas, *Water (Switzerland)*, vol. 11, no. 4, Apr. 2019. doi: 10.3390/w11040786.

Carreño Conde, F. & De Mata Muñoz, M., 2019. Flood Monitoring Based on the Study of Sentinel-1 SAR Images: The Ebro River Case Study. *Water* 2019, 11, 2454. https://doi.org/10.3390/w11122454.

Conversi, S., Carrion, D., Norcini, A., Riva, M., 2023. Integrating optical and radar imagery to enhance river drought monitoring. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.* XLVIII-1/W2-2023. 1363–1371. https://doi.org/10.5194/isprs-archives-XLVIII-1-W2-2023-1363-2023.

Copernicus, 2022. River discharge. *European State of the Climate 2022*. Last visited May 05 2024, on: https://climate.copernicus.eu/esotc/2022/river-discharge.

Dai, A., 2011. Drought under global warming: a review. *Wiley Interdisciplinary Reviews: Climate Change*, *2*(1), John Wiley & Sons, Inc., 45-65. https://doi.org/10.1002/wcc.81.

De Fioravante, P., Luti, T., Cavalli, A., Giuliani, C., Dichicco, P., Marchetti, M., Chirici, G., Congedo, L., Munafò, M., 2021. Multispectral Sentinel-2 and SAR Sentinel-1 Integration for Automatic Land Cover Classification. *Land* 2021, 10, 611. https://doi.org/10.3390/land10060611.

European Environment Agency, 2021. Europe's changing climate hazards – an index-based interactive EEA report. Web report. Published 17 Nov 2021. Last modified 17 Feb 2023. Last visited May 05 2024, on: https://www.eea.europa.eu/publications/europes-changing-climate-hazards-1.

GEE, 2024. Sentinel Collections. *Earth Engine Data Catalog*. Last visited April 30 2024, on: https://developers.google.com/earth-engine/datasets/catalog/sentinel.

Goh, T.Y., Basah, S.N., Yazid, H., Safar, M.J., Saad, F.S., 2018. Performance analysis of image thresholding: Otsu technique. *Measurement*, *114*, 298-307.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote SensingRemote Sensing of Environment*.

Huang, C., Li, X., Wen, Y., 2021. An Otsu image segmentation based on fruitfully optimization algorithm. *Alexandria Engineering Journal*, Volume 60, Issue 1, Pages 183-188. ISSN 1110-0168. https://doi.org/10.1016/j.aej.2020.06.054.

Joyce, A.T., 1977. Procedures for gathering ground truth information for a supervised approach to a computer-implemented land cover classification of LANDSAT-acquired multispectral scanner data.

JRC, 2011. NDWI: Normalized Difference Water Index. Product fact sheet: NDWI – Europe. DESERT Action – LMNH Unit. European Commission.

Landuyt, L., Van Wesemael, A., Schumann, G.J.P., Hostache, R., Verhoest, N.E.C., Van Coillie, F.M.B., 2019. Flood Mapping Based on Synthetic Aperture Radar: An Assessment of Established Approaches. *IEEE Transactions on Geoscience and Remote SensingRemote Sensing*, vol. 57, no. 2, pp. 722-739, Feb. 2019. https://doi.org/10.1109/tgrs.2018.2860054.

Markert, K.N., Markert, A.M., Mayer, T., Nauman, C., Haag, A., Poortinga, A., Bhandari, B., Thwal, N.S., Kunlamai, T., Chishtie, F., Kwant, M., Phongsapan, K., Clinton, N., Towashiraporn, P., Saah, D., 2020. Comparing Sentinel-1 Surface Water Mapping Algorithms and Radiometric Terrain Correction Processing in Southeast Asia Utilizing Google Earth Engine. *Remote Sens.* 2020, 12, 2469. https://doi.org/10.3390/rs12152469.

McFeeters, S.K., 1996. The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features', *Int J Remote Sens*, vol. 17, no. 7, pp. 1425–1432. https://doi.org/10.1080/01431169608948714.

Milczarek M., Robak, A., Gadawska, A., 2017. Sentinel Water Mask (SWM) – New index for water detection on Sentinel-2 images. 7th Advanced Training Course on Land Remote SensingRemote Sensing, Szent István University, Gödöllő, Hungary.

Montanari, A., Nguyen, H., Rubinetti, S., Ceola, S., Galelli, S., Rubino, A., Zanchettin, D., 2023. Why the 2022 Po River drought is the worst in the past two centuries. *Sci. Adv.9*, eadg8304. https://doi.org/10.1126/sciadv.adg8304.

Mountrakis, G., Im, J., Ogole, C., 2011. Support vector machines in remote sensing: A review, *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 3. pp. 247–259. https://doi.org/10.1016/j.isprsjprs.2010.11.001.

Mukherjee, S., Mishra, A., Trenberth, K.E., 2018. Climate Change and Drought: a Perspective on Drought Indices. *Curr Clim Change* Rep 4, 145–163 (2018). https://doi.org/10.1007/s40641-018-0098-x.

Nikolova, M., 2022. The Danube's historically low water levels are damaging both the environment and the economy. *Emerging Europe*. Published 17 Aug 2022. Last visited May 05 2024, on: https://emerging-europe.com/news/the-danubes-historically-low-water-levels-are-damaging-both-the-environment-and-the-economy/.

Otsu, N., 1979. A Threshold Selection Method from Gray-Level Histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62-66. https://doi.org/10.1109/TSMC.1979.4310076.

Schonlau, M., & Zou, R.Y., 2020. The random forest algorithm for statistical learning, *Stata Journal*, vol. 20, no. 1, pp. 3–29. https://doi.org/doi: 10.1177/1536867X20909688.

Ting, K.M., 2011. Confusion Matrix. In: Sammut, C., Webb, G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-30164-8_157.