An innovative and reliable water leak detection service supported by data-intensive remote sensing processing

R. Martins[1], A. Azevedo[2], A.B. Fortunato[3], E. Alves[4], A. Oliveira[5], A. Carvalho[6]

LNEC – National Laboratory for Civil Engineering/Hydraulics and Environment Department, Lisbon, Portugal
EDIA - Empresa de Desenvolvimento e de Infraestruturas do Alqueva, Beja, Portugal

[1]rjmartins@lnec.pt, [2]aazevedo@lnec.pt, [3]afortunato@lnec.pt, [4]ealves@lnec.pt, [5]aoliveira@lnec.pt, [6]acarvalho@edia.pt

Abstract. The WADI project (Water-tightness Airborne Detection Implementation), integrated within the H2020 initiative, is developing an airborne water leak detection surveillance service, based on manned and unmanned aerial vehicles. This service aims to provide water utilities with adequate information on leaks in large water distribution infrastructures outside urban areas. Given the high cost associated with water infrastructure networks repairs, a reliability layer is necessary to improve the trustworthiness of the WADI leak identification procedure, based on complementary technologies for leak detection. Herein, a methodology based on the combined use of Sentinel remote sensing data and a water leak pathways model is presented, based on data-intensive computing. The resulting water leak detection reliability service, provided to the users through a web interface, targets prompt and cost-effective infrastructure repairs with the required degree of confidence on the detected leaks. The web platform allows for both data analysis and visualization of Sentinel images and relevant leak indicators at the sites selected by the user. The user can also provide aerial imagery inputs, to be processed together with Sentinel remote sensing data at the satellite acquisition dates identified by the user. The platform provides information about the detected leaks location and time evolution, and will be linked in the future with the outputs from water pathway models.

Keywords: remote sensing, water leak service, data-intensive computing, hpc.

1   Introduction

The shortage of freshwater is a growing concern [1]. This shortage has a variety of causes, such as the increase of droughts due to climate change, the population growth, the pollution of water bodies, and the production of crops for biofuel. This problem calls for practices that improve water preservation. In this scope, water losses associated with large distribution networks, frequently used for irrigation purposes, constitute a major concern, given the difficulty to pinpoint the location of leaks over very long pipes, often of difficult physical access [2]. As a result, considerable water losses occur due to long leak detection times. The costs associated with conventional detec-

tion techniques have promoted the development of methodologies for water leak detection based on remote sensing outside of urban areas [3].

Remote sensing imagery data have become an essential tool for the scientific community to study the Earth in many disciplines, such as coastline detection, coastal morphology or marine oil spills follow-up [4,5,6]. Remote sensing is a data-intensive technology due to the typical resolution and size of the images. With the increase of the computational power and finer and more frequent image gathering, available through public access such as Sentinel satellites outputs, the conditions are met to build information systems to address societal needs for the reduction of water losses. Data-Intensive Scientific Discovery (DISD) tools are one of the avenues to handle the high volumes of remote sensing data, combine them with models to build intelligent services and make them available to society.

Recently, remote sensing has been used to provide water leak detection services at different spatial and temporal scales. Satellite-based leak detection services such as UTILIS [7] have been applied in several types of water leak detection, but their efficiency is limited when the rate of the leak is small (typically about 60% for flow rates under 1 m^3/h). Given the difficulty to reach and repair large water mains, leaks should be detected at their early stages, so a technology that provides better performance for small flows is necessary. A different approach was proposed in the project WADI. In this project, an airborne water leak detection surveillance technology was developed. It consists in coupling and optimizing off-the-shelf optical remote sensing devices (multispectral and infrared cameras) and applying them in two complementary aerial platforms (manned and unmanned). The WADI technology provides the detection of leaks on the acquired multispectral or/and infrared images, a data distribution system for the delivery of the geospatial data to the water utilities, and feedback on service performance. It was validated in an operational environment, to detect small water leaks in the Societé du Canal de Provence infrastructure in France [8]. Being developed with simple and off-the-shelf hardware and fast to operationalize, this technology has the potential to become the leading choice for small water leak detections.

While the performance evaluation of this new service is still underway, its reliability can be strengthened with complementary methodologies that can be used on top of the WADI methodology to reduce false positives and negatives, thus reducing costs and improving efficiency. Herein, a reliability methodology is proposed to be used with two distinct goals: to motivate the use of the WADI service (as a preliminary detection service) and to complement it after the initial leak detection, improving its performance. This new reliability methodology combines global data with local characteristics and merges them in a single evaluation available through a user-friendly web interface. The reliability layer is built by analyzing indexes based on Sentinel images and further complemented with a terrain-following water pathways model. This approach differs from the UTILIS system, based on microwave reflectometry, that works well at any time of the day or night, by being able to travel through atmospheric interferences such as clouds, dust particles, and aerosols.

This paper is organized as follows. Section 2 provides an overview of the WADI detection procedure and highlights its strengths and main characteristics. Section 3 presents the general reliability strategy while Section 4 is dedicated to describe in detail the Sentinel remote sensing-based analysis. Section 5 describes the WADI reliability web-portal and its architecture. Finally, Section 6 summarizes the conclu-

sions and anticipates the forthcoming work, in particular highlighting how DISD and other approaches can be used in the overall WADI leak detection service.

2   Overview of the WADI core water leak procedure

One of the main challenges of Horizon 2020 (H2020) is to reduce losses in the water distribution systems [9] and to promote efficiency and resilience on today's society against climate changes. Water availability is already a concern in the context of climate change, affecting most of the continents and in many areas. There is a need to develop better methodologies for easier and faster leak detection in the water systems since the existing ones are unreliable, time-consuming and expensive.

The WADI project, integrated within the H2020 initiative, aims at developing an airborne water leak detection surveillance service to provide water utilities with adequate information on leaks in water infrastructures outside urban areas, thus enabling prompt and cost-effective repairs. WADI's innovative concept consists in coupling and optimizing off-the-shelf optical remote sensing devices (multispectral and infrared cameras) and applying them on two complementary aerial platforms (manned and unmanned) in an operational environment [8]. The WADI service offers several benefits to infrastructure stakeholders: the reduction of water and energy consumption and emissions caused by the leak, a long-term plan for water monitoring in difficult physical access conditions, and the adaptability for accurate and tailored leak detection in water transmission systems.

To provide additional confidence to the WADI service and to support its performance analysis, a reliability layer was included in this service, through the use of complementary methods and data. While aerial images have the advantage of being reliable and precise on quality and image resolution, their application requires dedicated resources and planning for a specific site. Publicly available satellite images have a lower resolution, but provide a continuous stream of data over time that can be used for preliminary identification of leaks. Satellite-based analysis can thus support the application of the WADI service over the areas where leaks are suspected to exist, to direct the airborne leak detection efforts to the most likely leak locations. In addition, satellite data can also be used afterwards to provide added confidence on the aerial detection. This is achieved by taking advantage of the evolution in time of the water and vegetation indexes, combined with water pathways models applied from the network location downhill.

The feasibility of the surveillance service developed by the WADI project was tested in real representative conditions through water leak detection campaigns in the Provence region (France) [8] and will be applied to the Alqueva infrastructure (Portugal) in the coming months.

3   Leak-detection reliability procedure: general description

The following procedure aims to take advantage of historical, high-frequency satellite images to improve the reliability of WADI's leak detection service. Each set of images (WADI's airborne images and Sentinel 2 images) has its own advantages relative

to the other. The former has a higher resolution and is typically obtained under optimal meteorological conditions. In contrast, the latter is free, frequent (every 2-3 days) and continuously available since 2016. These different advantages suggest the possibility of combining the two sources of information to optimize the WADI service.

Information about possible leaks is obtained by analyzing the satellite images in both space and time. The analysis is performed on indexes that describe the humidity of the soil, such as the Normalized Difference Water Index (NDWI), which are computed from the remote sensing images. These indexes are described in the next section.

In the first step, images are analyzed in space to detect possible leaks. The spatial variability is assessed based on the Laplacian (i.e., the sum of the second derivatives along the two horizontal axes) of the water indexes. The signature of a leak is expected to correspond to a local minimum of this Laplacian. The algorithm thus searches for the largest minima. Furthermore, the leaks are expected to produce a signature in the water index very close to their location. Hence, the distance between a local extreme in the water index and the nearest pipe or channel provides the means to eliminate false positives. The next section describes in detail the implementation of this first step. In the future, this algorithm will be improved using the topography data and a water pathway model to further eliminate false negatives.

The second step of the procedure is applied to the points identified in the first step and takes advantage of the wealth of historical Sentinel data. In general, the water index at a given location is expected to follow a seasonal pattern, associated with rainfall or irrigation procedures (Fig. 1). This seasonal pattern is first determined by averaging the yearly signals (climatology) and then removed from the signal. The difference between the initial signal and the climatology, denoted anomaly, is expected to have a smaller temporal variability than the initial signal. Hence, abrupt increases in the water index are easier to identify. These abrupt changes can potentially correspond to the initiation of the leaks.

The outcome of this procedure, which is detailed and illustrated in the following section, is thus a list of the locations most likely to have water leaks. These locations can be ordered based on the likelihood of a leak, measured by the value of the Laplacian of the water indexes. For each point in the list, the date when the potential leak is first detectable can also be identified. This information can be helpful to determine possible reasons to explain the occurrence, whether or not it actually corresponds to a leak. For instance, the date can correspond to an intervention in a pipe which accidentally caused a leak that went unnoticed.

Such a list can be used *a priori*, to help reduce the effort of searching for leaks in the field, by concentrating the efforts in areas where leaks are most likely; or it can be used *a posteriori*, integrated in the workflow of analysis of WADI's images, to provide an additional layer of information for improved reliability. By linking these detections to a water pathway model starting at the network locations, the likelihood of a leak at that location can be corrected. For instance, if the location is far from the network or located uphill from the network, the probability that it corresponds to a leak decreases. This paper focuses on the satellite methodology, supported by Sentinel 2 images, and illustrates it at the second case study of the WADI project - the Monte Novo area in the Alqueva water distribution infrastructure (Portugal).
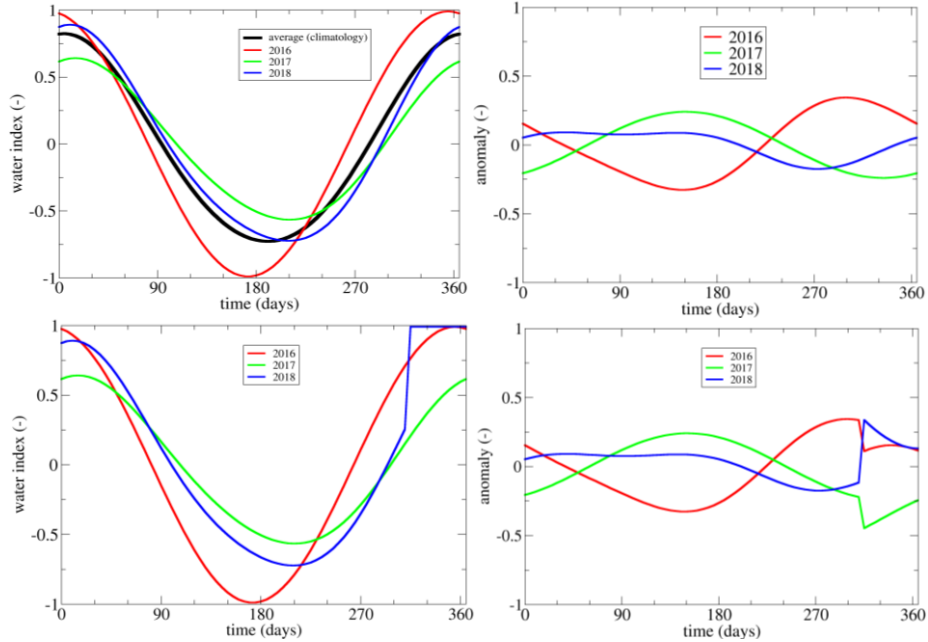
Fig. 1. Theoretical water indexes, climatology (left) and anomalies (right), for a case without (top) and with (bottom) a leak in day 315 in 2018. The leak is identifiable by the strong negative second time derivative of the 2018 anomaly.

4 Methodology for the water leak detection based on Sentinel 2 images

The methodology for processing Sentinel images in order to detect water leaks is described in detail in this section and is based on three parts. In the first part, the water indexes of all the available Sentinel 2 image sets at the area of study are processed. Normalized Difference Water Index (NDWI) and Modified Normalized Difference Water Index (MNDWI) are the two most popular water indexes available for remote sensing methodology processing.

The NDWI was proposed by McFeeters [10] to delineate open water features and enhance their presence in remotely-sensed digital imagery, by using reflected Near Infrared (NIR) radiation and visible green light to enhance the presence of such features while eliminating the presence of soil and terrestrial vegetation features:

$$NDWI = \frac{Green - NIR}{Green + NIR} \qquad (1)$$

Xu [11] determined that, while using Landsat imagery, the application of the NDWI in water regions with a built-up land background does not achieve its goal as expected because the extracted water information in those regions is often mixed with built-up land noise. This author proposed the MNDWI, a modified version of the McFeeters' NDWI, that uses Middle Infrared (MIR) instead of Near Infrared to en-

hance the water detection and even remove built-up land noise as well as vegetation and soil noise. The MNDWI is selected as main water index, since the application detects water leaks on land:

$$MNDWI = \frac{Green - MIR}{Green + MIR} \qquad (2)$$

In the second part, the climatology of the indexes is processed, as explained in the previous section. In this case, the climatology consists in a set of images averaging all the images for the same day of the year for all available years (see Fig. 2).
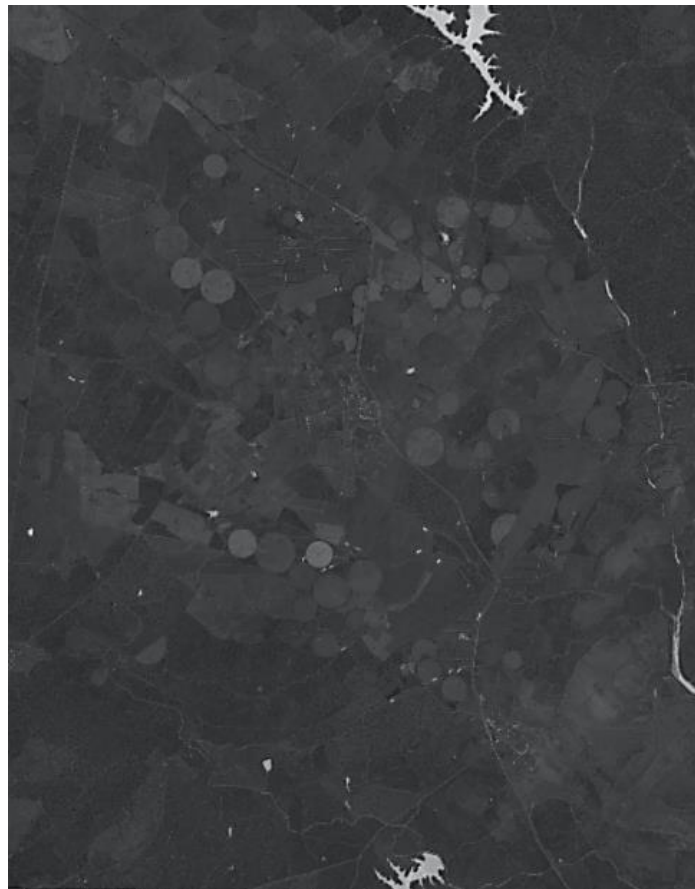

Fig. 2. An example of a climatology image of the MNDWI. The light colors correspond to water bodies

Possible leaks are determined in the third part. The user selects a Sentinel image set for a specific date, and the service starts automatically processing the water index and resamples it to the user-selected region. Because images are not available every day, the anomaly is computed as the difference between the image for the user-selected date and the climatology image for the closest day available in the climatology dataset. The most likely locations for the leaks are identified by applying the La-

placian operator (second derivative) to the anomaly, which is estimated by finite differences as:

$$Q_{i,j} = \frac{(P_{i-1,j} - 2P_{i,j} + P_{i+1,j})}{dx^2} + \frac{(P_{i,j-1} - 2P_{i,j} + P_{i,j+1})}{dy^2} \qquad (3)$$

where P is the anomaly image, Q is the second derivative image, i and j represent the pixel position and dx and dy represent the pixel size, corresponding to 10 m in the Sentinel images. The lowest values of the Laplacian image are marked as possible leaks (Fig. 3).

These locations are then confronted with the water network location and the possibility of water (from a leak) traveling from the network to the hypothetical leak location.
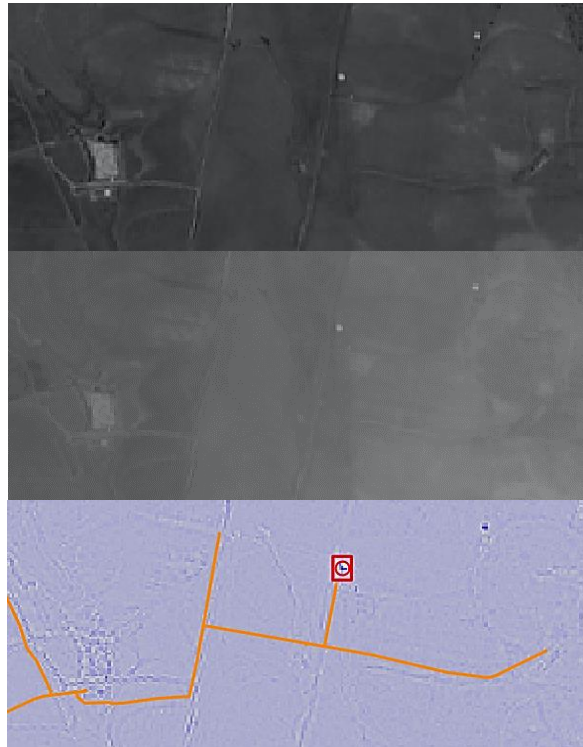

Fig. 3. Example of leak detection: MNDWI (top), Anomaly (middle), Laplacian and possible leaks (bottom) of a small area of the AoS. The network location is shown in orange.

However, there are some caveats that must be taken into account in this methodology. The first issue is how to process and show water indexes of an area of study where two or more image sets overlap each other in space for the same date. This problem was addressed by merging the water index of the various images into a single image.

A second issue that must be addressed is the different levels of image acquisition in Sentinel 2 images: Level-1C (L1C) and Level-2A (L2A). The algorithm developed

herein uses by default the L2A images, which are referenced to the Bottom of Atmosphere through the application of an atmospheric correction. When no L2A images are available in the area of study, the algorithm automatically downloads the L1C images and applies the atmospheric correction of the SNAP toolbox, developed by ESA.

5   A Web interface for the reliability layer in WADI

5.1   Functionalities of the web interface

A web-portal interface was developed to facilitate the usage of the reliability methodology. This web-portal is developed in Django, a Python web framework to create web applications with a frontend and a backend. The frontend is the presentation layer of the web application, through which the framework loads the pages for the user to interact with and requests data from them. The backend is the 'engine' running on the back of the portal to attend and process user requests.

The WADI reliability web-portal is structured in "workspaces" and "areas of study" (AoS). Each user can define several workspaces, that typically correspond to a region of interest for water leak analysis. A workspace is defined through a map where users can interact and create/edit/delete areas of study. An AoS is a small area defined by the user on that map to process images and visualize results. In each workspace, there are two panels to interact. The top panel is a menu bar where the user configures the region for leak detection analysis (create/edit/delete/select the AoS) and then proceeds to the functions 'climatology'/'leak detections'. The right panel organizes the analysis performed by a user under a workspace through a list of previously created AoS. Each AoS has some options of actions to be performed by the user. Fig. 4 illustrates the typical roadmap for a leak detection study.
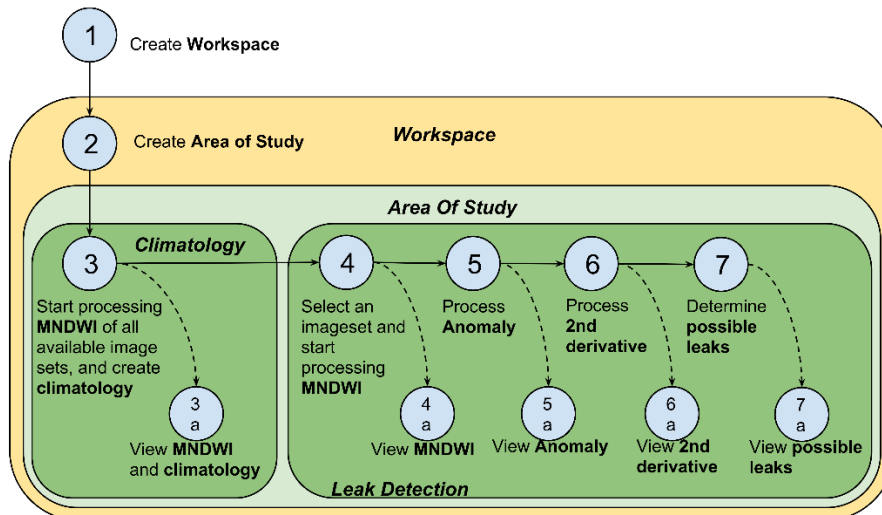

Fig. 4. Sequential steps for creating a leak detection map in an area of study

First, the user must create a workspace. Then, in the 'climatology' page, the user selects "Create Area of Study", defines the region to process on the map and names it (see Fig. 5). The new AoS will appear on the right panel.
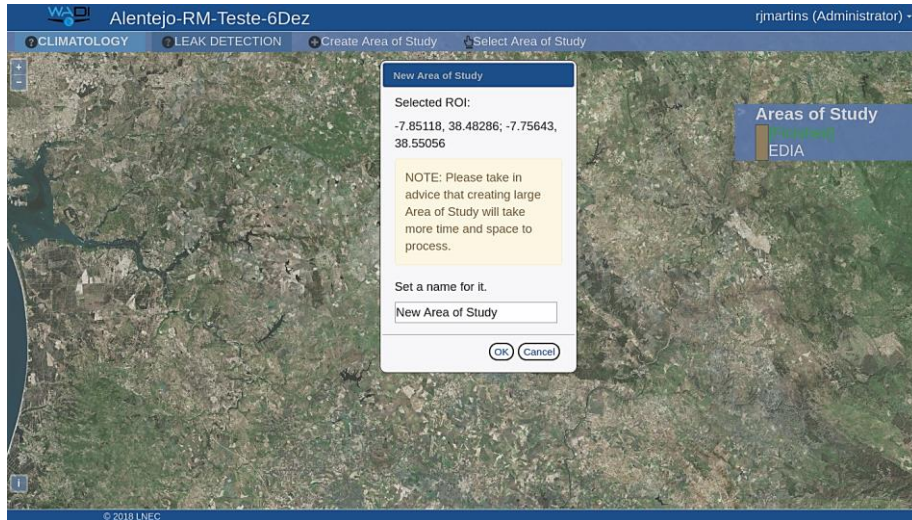

Fig. 5. Climatology overview - Create new area of study

In order to interact with the AoS and start the image processing, the user selects the AoS on the map and selects "See Climatology Status" on the right panel. A popup is shown to the user with the list of available Sentinel images for processing the climatology (Fig. 6). The user then clicks on "Calculate Climatology", and the WADI processing service starts processing the images asynchronously. Processing the climatology takes a considerable time, as all the available Sentinel images for that region are processed. The future usage of speed-up strategies like HPC is very important when processing large AoS or many images simultaneously.

On the 'climatology' page, the user can see the AoS' processed indices and climatology and check its status and availability for new Sentinel products available at the ESA Sentinel repository.

After the climatology processing, the user can start validating the data by viewing the processed climatology images and/or checking its pixel value. To check the pixel values, the user selects "Pixel Climatology" and clicks on a point inside the AoS on the map. A balloon pointing to that point is loaded with a chart representing the pixel values of the processed water index images for each day of the year.
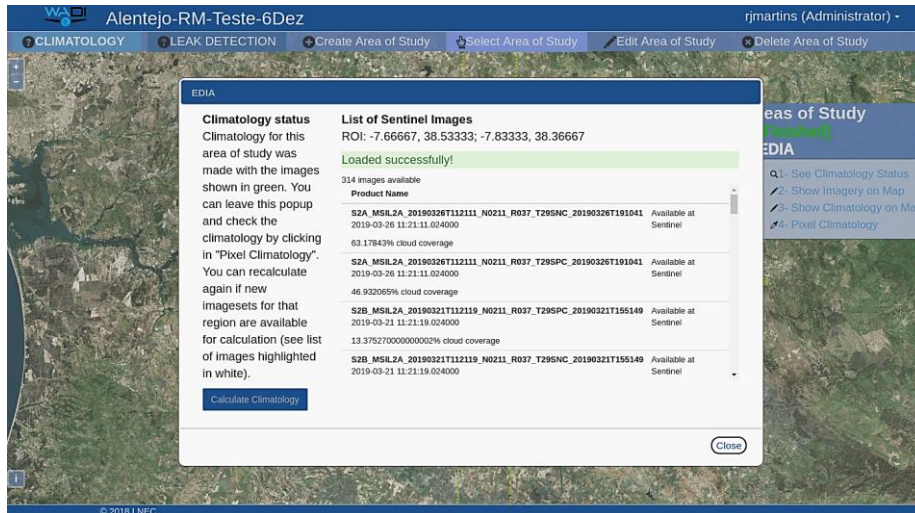
Fig. 6. Climatology overview - list sentinel images available for climatology processing

With the processed climatology, the user can start the leak detection for that AoS. In each step, the user checks the results and decides whether the processing should continue. The user selects a Sentinel imageset from the repository, and the service automatically initiates the water indexes processing. The user can then start processing the remaining steps: the anomaly, the Laplacian and the possible leaks identification (Fig. 7).
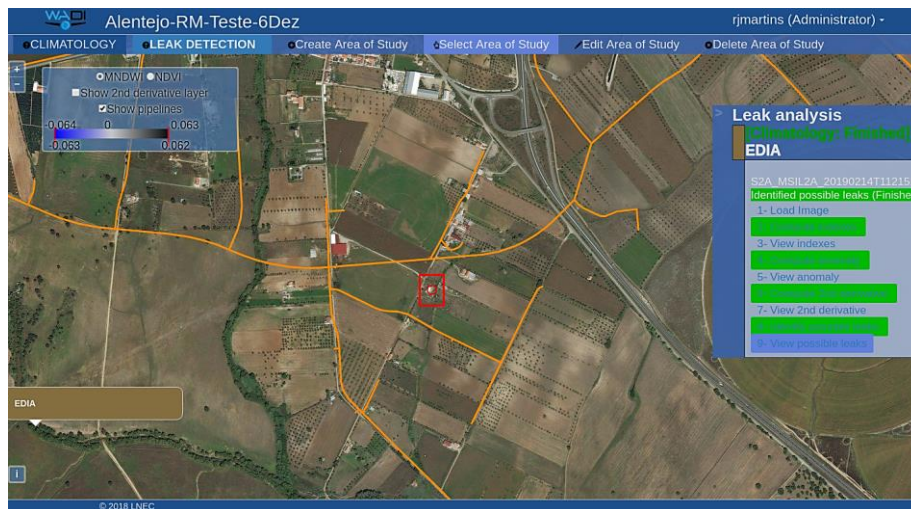

Fig. 7. Leak detection overview - View possible leaks

5.2 Architecture of the Web interface

The current web-portal architecture is described in Fig. 8. Inside the web-portal, the frontend will interact directly with the backend. It processes the frontend requests and

checks the ESA's Copernicus Open Access Hub repository for available sentinel image sets. The backend has an embedded raster layer service to load and store the processed images as rasters. The asynchronous processing service interacts with the backend to provide its processing status, to store the final products in the raster layer service, and to download image sets from the repository.
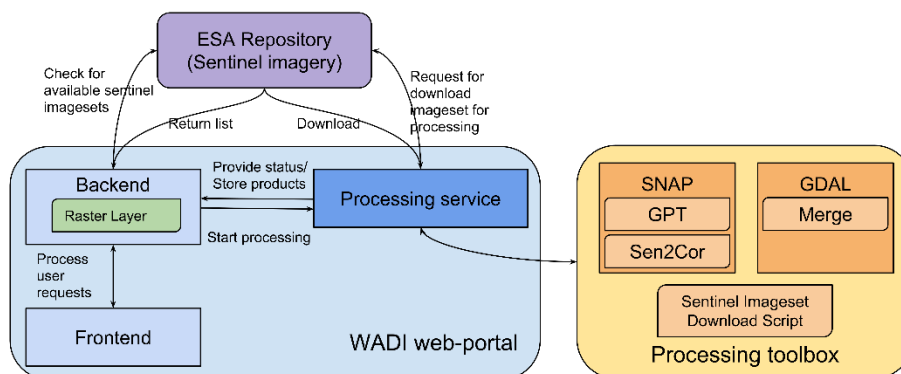

Fig. 8. WADI Architecture diagram

The processing service depends on executables from the processing toolbox to perform various tasks: to download/convert/merge image sets, to process the data, to store the processed images and to apply other steps for leak detection. The processing toolbox consists of SNAP, a toolbox for Sentinel image processing with Sen2Cor for conversion, GPT for script processing (rescaling and water index calculation), and GDAL for other tasks, such as image merging. A new approach for processing images needs to be studied and applied in HPC, since SNAP does not presently support parallel processing.

6   Conclusions and future work

The reliability strategy proposed herein brings an additional level of confidence to the WADI service, aiming to improve its quality through the usage of data-intensive algorithms applied to public satellite data. Given the low resolution in time of the core WADI service (typically a few plane passages at each location), remote sensing by satellites can work both as a predecessor of the WADI service, to motivate its usage for fine pinpointing of leaks, and also as a post-processing tool, combined with a water pathways model. Both avenues can now be explored through the web portal where this service is available, integrated in the Portuguese National Computing Infrastructure. Herein, we focused on the satellite data component of the methodology, the water pathways model being currently under implementation.

The present methodology has many avenues to evolve to a faster and more detailed service. Future work consists in adapting this methodology into the Data-Intensive Scientific Discovery (DISD) paradigm. Data-Intensive Scientific Discovery focuses on exploring the best strategies that use the available computational resources for processing large datasets, and to provide faster data outputs to analyze. DISD is already being applied for control in critical areas, such as oil industry and airport traffic.

Several DISD strategies can now be pursued to improve the performance of the methodology presented here.

High Performance Computing (HPC) is one of them, as it has an important role in the solution of data-intensive and computationally-demanding problems, such as numerical modeling or image processing. HPC takes the advantage of speeding up the computation by efficiently using the resources available on a machine or a cluster of them in cloud/grid environment. Still, many challenges in applying HPC to leak detection remain, such as finding the best strategy for the data decomposition to assure the correct load balancing, node communication and resulting accuracy, and the possible parallel processing limitations caused by hardware bottlenecks.

Future work consists in applying HPC on the reliability WADI service, which is important to provide faster detection and reduce costs and water losses. Our methodology will be tested for speedup the processing by using CPUs and/or GPUs, since the images have a high resolution, and the cost and time to process them sequentially is larger too. One possible way to do that is being studied and consists in 'slicing' equally the images to be processed by each CPU/GPU core, using Python with OpenCV and/or numpy.

Finally, the validation of the methodology in the Alqueva infrastructure for in-situ identified leaks that happened since the Sentinel 2 data became available is underway. This analysis will then be used together with the WADI flights processed images for high accuracy detection results.

7 Acknowledgements

References

1. FAO (2007). Coping with water scarcity. Challenge of the twenty-first century. UN-Water, http://www.fao.org/3/a-aq444e.pdf.
2. Carvalho, A., Oliveira, A., Alves, E. (2018). WADI – Detecção inovadora de fugas de água em grandes redes de distribuição a partir da vigilância através de aviões tripulados e drones. In: 14th Water Congress, 7-9 March 2018, Évora, Portugal.
3. Agapiou, A., Alexakis, D.D., Themistocleous, K., Hadjimitsis, D.G. (2016). Water leakage detection using remote sensing, field spectroscopy and GIS in semiarid areas of Cyprus, Urban Water Journal, 13:3, 221-231, DOI:10.1080/1573062X.2014.975726
4. Azevedo, A., Oliveira, A., Fortunato, A., Bertin, X., Bertin, And. (2009). Application of an Eulerian-Lagrangian oil spill modeling system to the Prestige accident: Trajectory analysis. Journal of Coastal Research SI. 56. 777-781.
5. Evagorou, E., Mettas, C., Agapiou, A., Themistocleous, K., Hadjimitsis, D. (2019). Bathymetric maps from multi-temporal analysis of Sentinel-2 data: the case study of Limassol, Cyprus, Adv. Geosci., 45, 397-407, https://doi.org/10.5194/adgeo-45-397-2019.
6. Fingas, M., Brown, C. (2017). A Review of Oil Spill Remote Sensing. Sensors. 18:91 doi: 10.3390/s18010091.

7. Whitepaper of UTILIS - Satellite Based Leak Detection: https://utiliscorp.com/wp-content/themes/utilis/files/new-files/other/UTILIS_whitepaper_A4_WEB.pdf
8. Chatelard, C., Krapez, J.-C., Barillot, P., Déliot, P., Yves-Michel, F., Pierro, J., Nouvel, J.-F., Helias, F., Louvet, Y., Legoff, I., Serra, G. (2018). Multispectral Approach Assessment for Detection of Losses in Water Transmission Systems by Airborne Remote Sensing. 10.29007/4xs9.
9. EC (2014). H2020-EU.3.5. - SOCIETAL CHALLENGES - Climate action, Environment, Resource Efficiency and Raw Materials, https://cordis.europa.eu/programme/rcn/664389/en
10. McFeeters, S. (1996). The Use of Normalized Difference Water Index (NDWI) in the Delineation of Open Water Features. International Journal of Remote Sensing - INT J REMOTE SENS. 17. 1425-1432. doi: 10.1080/01431169608948714.
11. Xu, H. (2006). Modification of Normalized Difference Water Index (NDWI) to Enhance Open Water Features in Remotely Sensed Imagery. International Journal of Remote Sensing. 27. 3025–3033. doi: 10.1080/01431160600589179.