

## **Water-tightness Airborne Detection Implementation**

# **D4.1 – Smart data processing and standards compliance - Report on the reengineering procedure undertaken for WADI data processing and analysis of the standards it complies with**

---

### **Authors:**

**Jimenez Gonzalez, D. (NTGS), Krapez J.-C (ONERA), Mendes A. (LNEC); Fortunato, A.B. (LNEC), Oliveira A. (LNEC), Alves E. (LNEC), Martins, R. (LNEC), Sanchis Muñoz J (GG), Barba Polo J (GG). De Badts E (GG)**

24th July, 2019



## Technical references

<b>Project Acronym</b>	WADI
<b>Project Title</b>	Water-tightness Airborne Detection Implementation
<b>Project Coordinator</b>	Elena Gaboardi, youris.com (YOURIS) elena.gaboardi@youris.com
<b>Project Duration</b>	October 2016 – March 2020 (42 months)
<b>Deliverable No.</b>	D4.1
<b>Dissemination level*</b>	PU
<b>Work Package</b>	WP 4 – Smart data processing and standardization
<b>Task</b>	T 4.1 – Onboard processing, specification and design T 4.2 – Definition of data processing and model T 4.4 – Data processing validation and standards compliance
<b>Lead beneficiary</b>	7 (NTGS)
<b>Contributing beneficiary/ies</b>	11 (GG) 2 (ONERA) 4(LNEC)
<b>Due date of deliverable</b>	31 May 2018
<b>Actual submission date</b>	31 July 2019

PU = Public

PP = Restricted to other programme participants (including the Commission Services)

RE = Restricted to a group specified by the consortium (including the Commission Services)

CO = Confidential, only for members of the consortium (including the Commission Services)



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 689239.

v	Date	Beneficiary	Author
1.0	13/07/2018	NTGS	Jimenez Gonzalez, D.
1.0	13/07/2018	LNEC	Mendes, A., Oliveira, A., Alves. E.
1.0	13/07/2018	GG	Sanchis Muñoz, J., Barba Polo, J., De Badts, E.
1.0	21/07/2019	YOU	Veronica Menghella, Elena Gaboardi
2.0	19/07/2019	ONERA	Krapez J.-C.
2.0	19/07/2019	LNEC	Mendes, A., Oliveira, A., Alves. E.
2.0	19/07/2019	GG	Sanchis Muñoz, J., Barba Polo, J., De Badts, E.
2.0	26/07/2019	YOU	Alice De Ferrari, Elena Gaboardi
3.0	26/07/2019	LNEC,	Mendes, A., Oliveira, A., Alves. E.
3.0	26/07/2019	GG	Sanchis Muñoz, J., Barba Polo, J., De Badts, E.
3.0	26/07/2019	YOU	Alice De Ferrari, Elena Gaboardi



## Disclaimer

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 689239.

The sole responsibility for the content of this report lies with the authors. It does not necessarily reflect the opinion of the European Union. The European Commission is not responsible for any use that may be made of the information contained therein.

## Executive Summary

The aim of this document is to define all components and processes necessary for the onboard data acquisition and processing, data pre-processing and data processing validation that will be needed to perform the software reengineering, developed later in a second document D.4.2.

The two WADI different airborne platforms of the project have different components and procedures for capturing the pipelines imagery:

- Manned (MAV): The multispectral and thermal selected cameras, gimbal and navigation are controlled by an onboard computer which also stores the images and the associated inertial and geographic data from the IMU.
- Unmanned (RPA): The multispectral and thermal selected cameras, and navigation are controlled by a field station on the ground, while RPAS stores the images and navigation data.

In any case the imagery from both platforms will need a pre-processing for obtaining orthomosaics and atmosphere correction or emissivity compensation depending on the chosen method for the parameter identification, namely here for moisture mapping. Then we can proceed to the processing separating the composite images (RED, TIR, NIR) and combine the responses from the same point at ground for obtaining the different target values of vegetation and water indexes.

This report also introduces a methodology for WADI data reliability after its post-processing, based on a combination of Sentinel images and a water pathway model. This reliability layer provides additional confidence to WADI image processing and supports the application of the WADI performance matrix.

Finally, this document presents the best practices for data file formats to make the documents and outputs of WADI accessible to others at a later time and to promote their interoperability among diverse platforms and applications. Metadata recommendations and metadata standards are also included to clearly describe data produced during the WADI project.

# Table of Content

List of Acronyms .....	vii
List of Figures .....	viii
Foreword .....	1
1 Introduction .....	2
2 Onboard data acquisition and processing .....	4
2.1 Manned aircraft platform .....	4
2.1.1 Software interface .....	5
2.1.2 Image acquisition and storage .....	5
2.1.3 Acquisition of flight parameters and synchronisation .....	5
2.2 RPA platform.....	6
2.2.1 Image acquisition and storage .....	7
2.2.2 Acquisition of flight parameters and synchronisation .....	7
2.3 Calibration and geographic information support .....	8
3 Data pre-processing .....	9
3.1 Orthomosaic of digital numbers, reflectance and brightness temperature .....	9
3.2. Considerations about pre-processing based on atmosphere correction or emissivity compensation.....	9
4 Data post-processing .....	13
4.1 Vegetation index .....	13
4.2 Temperature-VI scatterplot .....	19
4.2.1 2D scatterplot construction .....	20
4.2.2 Trapezoid boundary determination and Water Index calculation .....	24
5 Data processing validation.....	36
6 Methodology for data reliability.....	37
7 Data File Formats.....	39
7.1 Best practices for file formats .....	39
7.2 Guidelines for choosing formats .....	39
7.3 File format standards .....	40
8 Metadata.....	42
8.1 Introduction .....	42
8.2 Metadata Standards.....	42
8.3 WADI metadata standards .....	43

8.3.1	ZENODO - Research data repository.....	43
8.3.2	Dublin Core Metadata Initiative .....	44
8.3.3	INSPIRE – Infrastructure for spatial information in Europe .....	45
8.4	Metadata Editors .....	47
8.4.1	ZENODO metadata editor.....	48
8.4.2	Dublin Core metadata editor .....	49
8.4.3	INSPIRE metadata editor.....	51
9	Conclusions.....	53
10	References .....	54

## List of Acronyms

BUSARD	Banc d'essais Ultra léger pour Systèmes Aéroportés de Recherche sur les Drones.
GCP	Ground Control Point.
IMU	Inertial Measurement Unit.
IR	InfraRed.
LWIR	Long Wave InfraRed([8.0-12.0 $\mu\text{m}$ ] waveband).
SMC	Soil Moisture Content.
MAV	Manned Air Vehicle.
MSL	Mean Sea Level.
NDVI	Normalized Difference Vegetation Index.
RPA	Remotely Piloted Aircraft.
SCP	Société du Canal de Provence.
SSD	Solid-State Drive (solid-state hard drive).
SWIR	Short Wavelength InfraRed ([1.0-2.5 $\mu\text{m}$ ] waveband).
S/N	Signal to Noise ratio.
UAV	Unmanned Aerial Vehicle.
VNIR	Visible and Near InfraRed ([0.4-1.0 $\mu\text{m}$ ] waveband).
WP $n$	Work Package $n$ .
D. $n$	Deliverable $n$ .
DEM	Digital Elevation Model.

## List of Figures

Figure 1: Manned aircraft payload setup. ....	4
Figure 2: RPA payload setup.....	6
Figure 3: Map of the spectral reflectance at 0.662 $\mu\text{m}$ . ....	15
Figure 4: Map of the spectral reflectance at 0.858 $\mu\text{m}$ . ....	15
Figure 5: Vegetation index $NDVI_{\rho}$ from the spectral reflectance maps in Figure 3 and Figure 4. ....	15
Figure 6: Vegetation index $NDVI_{\rho}$ from the spectral reflectance maps in Figure 3 and Figure 4. ....	16
Figure 7: Vegetation index $OSAVI_{\rho}$ from the spectral reflectance maps in Figure 3 and Figure 4. ....	16
Figure 8: Vegetation index $OSAVI_{\rho}$ from the spectral reflectance maps in Figure 3 and Figure 4. . ....	17
Figure 9: “Raw” vegetation index $NDVI_s$ from the raw radiometric signals.....	17
Figure 10: “Raw” vegetation index $NDVI_s$ from the raw radiometric signals.....	18
Figure 11: Histogram of the $NDVI_{\rho}$ map in Figure 5. ....	18
Figure 12: Histogram of the $NDVI_s$ map in Figure 9. ....	18
Figure 13: Histogram of the $NDVI_s$ map in Figure 10. ....	19
Figure 14: Principle of soil moisture mapping from a Temperature-Vegetation Index (T-VI) plot. ....	19
Figure 15: Mosaic of TIR images. ....	20
Figure 16: 2D scatterplot when considering 1 in 10 pixels in the T and VI mosaics.....	21
Figure 17: Temperature image. ....	22
Figure 18: 2D scatterplot when considering 1 in every 10 pixels in the T and VI mosaics.....	23
Figure 19: 2D scatterplot when considering 1 in every 50 pixels in the T and VI mosaics.....	23
Figure 20: 2D scatterplot when considering 1 in every 100 pixels in the T and VI mosaics...	24
Figure 21: 2D scatterplot when considering 1 in every 50 pixels in the T and VI mosaics.....	25
Figure 22: Water index map inferred from Figure 21. ....	27
Figure 23: Close-up of the Water Index map in Figure 22 centered in the area of the artificial leak. ....	28
Figure 24: Water index profile along the black line in Figure 23. ....	28
Figure 25: 2D scatterplot when considering 1 in every 50 pixels in the T and VI mosaics.....	29

Figure 26: Water index map inferred from Figure 25. ....	30
Figure 27: Close-up of the Water Index map in Figure 26. ....	30
Figure 28: 2D scatterplot when considering the index $OSAVI_{\rho}$ . ....	31
Figure 29: Water index map inferred from Figure 28. ....	31
Figure 30: Close-up of the Water Index map in Figure 29. ....	32
Figure 31: 2D scatterplot when considering the index $NDVI_{\rho}$ . ....	32
Figure 32: Water index map inferred from Figure 31. ....	33
Figure 33: Close-up of the Water Index map in Figure 32. ....	33
Figure 34: Close-up of the apparent temperature map. ....	34
Figure 35: 2D scatterplot as in Figure 21. ....	35
Figure 36: Water index map inferred from Figure 21. ....	35
Figure 37: Example of leak detection. ....	38
Figure 38: ZENODO upload data type. ....	43
Figure 39: ZENODO Basic information example for Dataset data type. ....	44
Figure 40: Qualified Dublin Core Metadata Element Set. ....	45
Figure 41: INSPIRE in your country. ....	46
Figure 42: ZENODO metadata form example. ....	48
Figure 43: ZENODO – Dublin Core export example. ....	49
Figure 44: Dublin Core Simple Generator. ....	50
Figure 45: Dublin Core Advanced Generator. ....	50
Figure 46: Dublin Core output. ....	51
Figure 47: European Open Source Metadata Editor (EUOSME). ....	52
Figure 48: European Open Source Metadata Editor (EUOSME). ....	52

## Foreword

WADI is a H2020 funded project (Grant Agreement No. 689239) aimed to develop an airborne water leak detection surveillance service in water transportation infrastructures. The project relies on innovative concept of coupling and optimising off-the-shelf optical remote sensing devices and their application on two complementary aerial platforms - manned and unmanned - used for distinctive purposes in infrastructure performance observation, i.e.: long distance and strategic infrastructure monitoring, and difficult and/or dangerous areas observation.

The current deliverable D4.1 "*Smart data processing and standards compliance*" belongs to Work Package 4 viz. "*Smart data processing and standardization*", whose primary aim is to provide all elements necessary for software reengineering and GUI development of WP4.3.

Deliverable D4.1 specifically refers to the outcomes of Task 4.1, 4.2 and 4.4 which comprises Subtasks 4.2.1, 4.2.2 and 4.2.3 as detailed below.

- Task 4.1. *Onboard processing, specification and design*
- Task 4.2. *Definition of data processing and model*
  - Subtask 4.2.1 *Definition of pre data processing*
  - Subtask 4.2.2 *Definition of physical inverse process*
  - Subtask 4.2.3 *Validation cases for data processing*
- Task 4.4. *Data processing validation and standards compliance*
  - Subtask 4.4.1 *Data processing validation*
  - Subtask 4.4.2 *Data standards compliance*

The report's preparation has been coordinated by NTGS, with active collaboration from GG, ONERA and LNEC.

# 1 Introduction

WADI develops an airborne water leak detection surveillance service in water transportation infrastructures. The project relies on innovative concept of coupling and optimising off-the-shelf optical remote sensing devices and their application on two complementary aerial platforms - manned and unmanned - used for distinctive purposes in infrastructure performance observation, i.e.: long distance and strategic infrastructure monitoring, and difficult and/or dangerous areas observation.

In 2017, several flight campaigns were performed with ONERA's BUSARD instrumented aerial platform over a series of areas belonging to the water network infrastructure provided by SCP. Artificial leaks were present in some of these areas.

The images that were recorded with two hyperspectral cameras (VNIR + SWIR) and a thermal infrared camera (TIR) constituted a rich database. An analysis was performed in WP3 to select the spectral images with the best ability to discriminate between wet and dry areas with the objective to detect leaks on water pipelines.

Among the various combinations of spectral images that have been tested it appears that best results are obtained with the empirical T-VI triangle method ("T-VI" is for Temperature Vegetation-Index – Sandholt, 2002 [8], Krapez, 2012) [4]. A comparative analysis has been recently presented (Chatelard, 2018) [5].

This implies that three images are recorded in the same time at different locations along the tested water transportation means. A first camera that is sensitive in the VIS/NIR spectral region will provide the first two images, one on each side of the chlorophyll red edge (typically at about  $0.65\mu\text{m}$ , resp. at about  $0.80\mu\text{m}$ ). In the camera that has been chosen for the manned aerial platform the selection between the two spectral regions will be performed thanks to a rotating filter wheel. On the unmanned aerial platform the VIS/NIR camera (MicaSense Red Edge 3) actually has several focal plane detectors with different spectral filters in front of them; two of them match the aforementioned wavelengths).

The third spectral region is in the thermal infrared (i.e. between  $8\text{-}12\mu\text{m}$ ), essentially to collect an image of the temperature of the soil/canopy surface. The camera onboard of the manned aerial platform will be a cooled camera with a spectral bandwidth  $7.7\text{-}9.3\mu\text{m}$  (Noxcam 640L). The one onboard of the unmanned aerial platform is an uncooled camera with a larger spectral bandwidth  $7.5\text{-}13.5\mu\text{m}$  (Flir Vue pro R 960 V0x microbolometer).

In the following parts we will describe the procedures for:

- Onboard data acquisition and processing:
  - Sensor data recording and storage.
  - Sensor synchronisation.
  - Recording of flight parameters as GPS position, pitching/rolling, aerial platform speed, flight altitude and timing.
  - Compilation of the geo-information.
  - Calibration process.

- Real time view of images for acquisition checking.
- data pre-processing:
  - Geometrics correction (camera field of view & optical geometrics aberration).
  - Radiometric adjustments relevant to dynamics signal, weak pixel and background noise.
  - Atmospheric path correction which removes the atmospheric impact on the signal (the necessity of this correction will be discussed).
  - Registration of the images.
  - Georeferencing of the images.
- Data post-processing:
  - Vegetation index map processing.
  - Temperature-VI scatterplot processing (automatic or manual specification of dry and wet edges).
  - Water-Index map processing.

## 2 Onboard data acquisition and processing

The WADI project proposes two different aerial platforms, manned and unmanned, to bring about the monitoring task. The data acquisition is carried out at different scale depending on the platform. The manned platform (airplane) is capable of covering wide areas in one flight at reasonable ground sampling resolutions whereas the unmanned platform (RPA) covers a limited area at high resolution. Both aerial platforms have different payload weight capability leading to the use of different instrumentation equipment and data processing pipelines.

### 2.1 Manned aircraft platform

The payload of the manned aircraft is composed by a multispectral wheel camera (Pixelteq SpectroCam) and by a thermal IR camera (Noxant NoxCam). Both cameras are installed on a gyrostabilized platform to correct in real time the undesired aircraft movements. An inertial measurement unit and a GPS provide attitude and geographic position for each triggered image. The payload is controlled by an onboard computer which also stores the images and the associated inertial and geographic data from the IMU. The Figure 1 shows the manned aircraft payload setup.

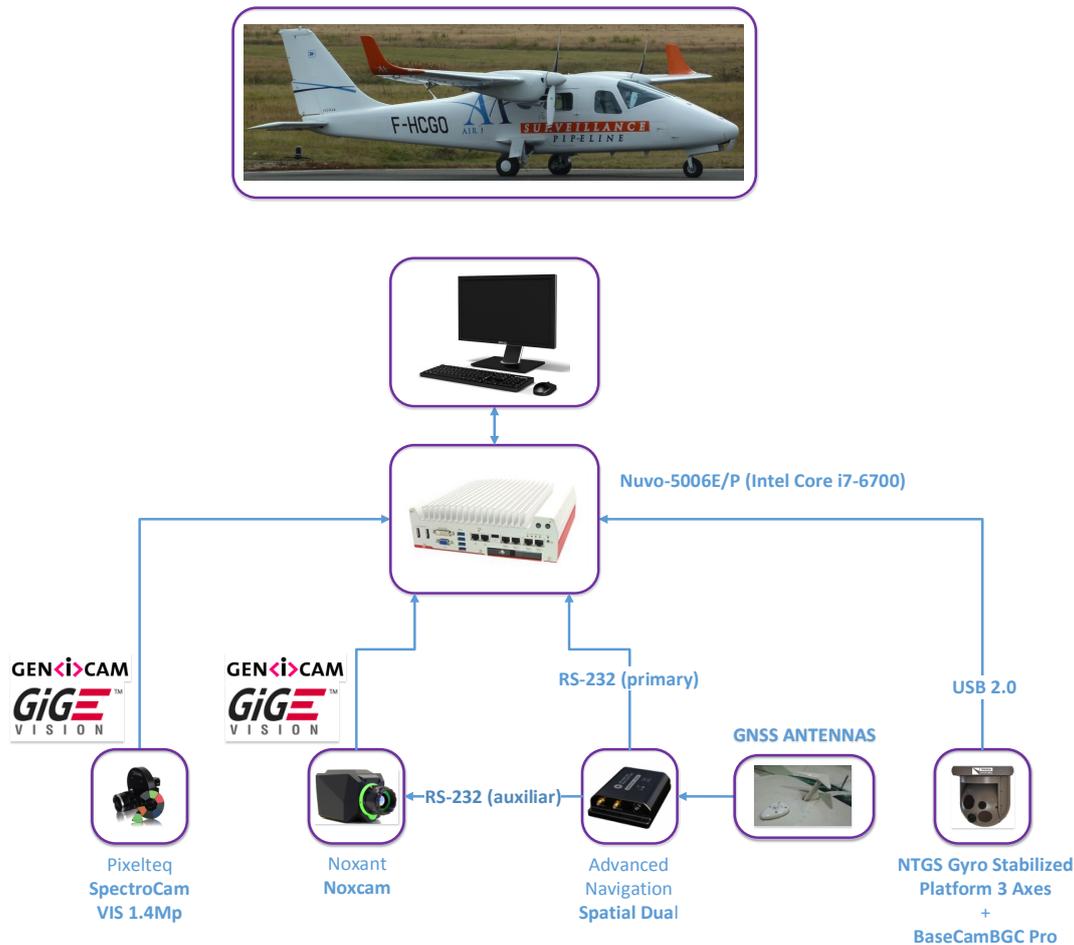


Figure 1: Manned aircraft payload setup. Both cameras are GigE Vision and Genicam compliant.

The Figure 1 shows a link between the Advanced Navigation IMU and the Noxcam. The Noxcam has an internal SSD drive which allows the recording of the images directly in the camera. The Noxcam receives attitude and geographic position data from the IMU and stores it in the associated metadata of each recorded picture.

### **2.1.1 Software interface**

The payload is controlled by a software interface called WadiFI (WADI Flight Interface) written in the frame of this project. The software interface makes use of the Genicam protocol to handle both cameras (control and image acquisition). WadiFI allows the following operations:

- Real time monitoring (both cameras).
- Cameras' configuration (exposure time, gain, etc.).
- Calculation of the minimum time between two camera shots (calculated for the most restrictive camera. In this case, the NoxCam thermal IR camera).
- Timed or maximum frame rate cameras' triggering.
- Attitude and geographic position data acquisition, monitoring and storage.
- Image storage.

Unlike the original AIRMON monitoring and acquisition onboard software interface, WadiFI is not intended to be used as a visual inspection software mainly because of the multispectral nature of the monitored data together with the need of applying an image processing pipeline which will retrieve the interest data.

### **2.1.2 Image acquisition and storage**

The images are stored in RAW uncompressed GeoTiff format (without suffering any processing step) in the onboard computer. The computer stores the images in two SSD hard drives in RAID 1 configuration. RAID 1 makes the storage fault tolerant as far as the drives are disposed as exact mirrors (two drives of the same capacity).

Unlike the case of AIRMON, it is discouraged to use external drives to store the images in real time. On the one hand, storing the images directly in an external drive makes the storing process susceptible to failures. On the other hand, at maximum ratings the system will store thermal IR pictures at 10 frames per second and spectral pictures at 20 frames per second making a massive number of accesses in external devices not intended for that purpose and making possible to have throughput problems.

### **2.1.3 Acquisition of flight parameters and synchronisation**

The data pre-processing step requires the acquisition of the following flight parameters synchronously with each camera trigger:

- Geographic coordinates (GPS position).
- Payload attitude (Euler angles roll, pitch and yaw).

- Altitude.
- Speed.
- Time tag.

Each picture has an associated set of flight parameters available in the metadata and in a spare CSV text file for convenience of the pre-processing commercial software packages and as a backup.

The flight parameters' acquisition is done at a maximum rate of 20 Hz.

Both cameras have different sensor sizes and lens focus lengths. No synchronisation is needed between both cameras.

## 2.2 RPA platform

The RPA payload is slightly different and has some limitations. The low weight capability of the RPA platform forces to choose ultra low weight cameras. No gyrostabilized gimbal is used for the same reason. To counteract these limitations the front and side overlap of the acquired images must be increased. The Figure 2 shows the RPA payload setup.

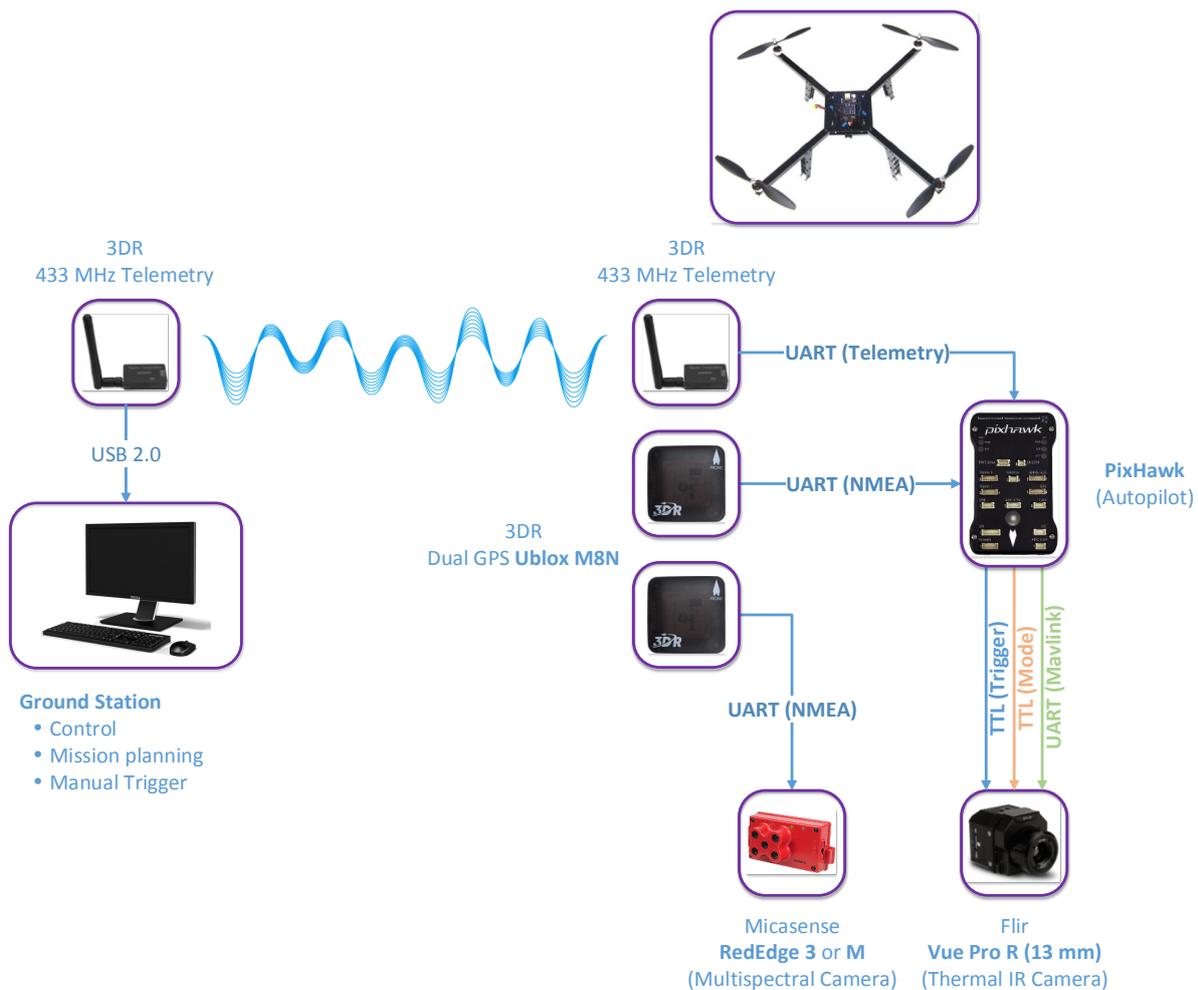


Figure 2: RPA payload setup.

Comparing with the manned platform case, the payload has been reduced to its minimum expression. The system is composed by two cameras a five band multispectral (Micasense RedEdge 3) and a thermal IR. The red and NIR bands of the multispectral camera are equivalent to those of the SpectroCam. The thermal IR camera, is a low weight and low cost equivalent to the Noxant Noxcam.

The trigger of the Micasense RedEdge 3 is automatic. It only needs a GPS and specifying the flight altitude and the required overlap. The camera triggers automatically to preserve the specified overlap at the specified altitude.

The Flir Vue Pro R can be triggered automatically in timed mode. However it allows to be triggered by the RPA autopilot. The autopilot, with a loaded planned mission and the input of GPS data, triggers the camera to meet the specified overlap at the specified altitude. The overlap and the altitude are indicated during the mission planning step.

It is possible to incorporate a video downlinks to monitor data from the ground but in this dual camera setup it is not operative. The control of the RPA is done through a telemetry link with the ground control station.

Due to the data processing pipeline used with RPA acquired data, it is not necessary the use of an IMU. Indeed, the pre-processing software package (Pix4D) does not effectively take inertial data into consideration.

### **2.2.1 Image acquisition and storage**

As mentioned, the multispectral camera acquires automatically the images based on the specified flight altitude and minimum image overlap. The GPS data are recorded in the metadata of each acquired image. The thermal IR camera makes use of the autopilot to be triggered. The autopilot also brings GPS data through the Mavlink port. The GPS data is also recorded as metadata of each picture.

In both cases, the images are stored in RAW uncompressed GeoTiff format (without suffering any processing step) in SD cards. Each camera has its own SD card where the images are recorded.

### **2.2.2 Acquisition of flight parameters and synchronisation**

The flight parameters recorded by the RPA cameras are minimal:

- Geographic coordinates (GPS position).
- Altitude.
- Time tag.

Each picture has an associated set of flight parameters available in the metadata. There is no need to record these flight parameters in a text file as far as in this case (it depends on the cameras used), the pre-processing software packages accept the metadata without problems.

The flight parameters' acquisition is done at a maximum rate of 10 Hz.

Both cameras have different sensor sizes and lens focus lengths. No synchronisation is needed between both cameras.

### **2.3 Calibration and geographic information support**

The need of specific calibration processes will depend mainly on the needs of the inversion process explained in detail in this document. The calibration, if required, can be done with a reflective reference panel with flat known spectral response. In the absence of calibration with a reference panel, it would also be necessary to compute the solar downward radiance at ground level. Additional atmospheric correction may be necessary.

Some ground control points (featured reference geographic points with known geographic coordinates) must be taken with ground topographic grade GPS in order to increase the quality of the mosaicking and registration process. The number will depend on the covered area. A minimum of 3 and a maximum of 12 ground control points is required as far as the pre-processing software packages use to accept no more than this range.

## 3 Data pre-processing

### 3.1 Orthomosaic of digital numbers, reflectance and brightness temperature

Once the images have been acquired, the first step is to obtain the orthomosaic in digital numbers, reflectance or bright temperature as appropriate. If no multispectral camera calibration has been done, the generated orthomosaic is in digital numbers. If the camera has been calibrated with a reference panel, the resulting orthomosaic is in reflectance. In the case of the thermal IR camera, if the emissivity is set to 1 (100%) the generated orthomosaic is in brightness temperature. At this step it is also possible to obtain vegetation index orthomosaics.

The generation of the orthomosaic is done by using a photogrammetric grade mosaicing software. Pix4D has been selected for pre-processing RPA acquired data, mainly for making the registration of the orthomosaiced bands automatically, while a package composed by different pieces of software is being used to pre-process the manned aerial platform data (Correlator 3D + Gefolki). The selection of the final pre-processing software is still under testing and discussion.

### 3.2. Considerations about pre-processing based on atmosphere correction or emissivity compensation.

Some elements of pre-processing are justified or not, depending on the chosen method for the parameter identification, namely here for moisture mapping. The pre-processing we are discussing now is related to atmosphere correction, emissivity compensation, and different types of instrumental calibration.

By anticipating a little on the inversion method that will be described in more details in the next part, let us mention that one of its main characteristics is that it is a *relative* method in many aspects:

- The moisture index that is provided as a final product is a *relative* index whose values are ranging from 0 for the driest areas to 1 for the most humid areas.
- It is based on how the local temperature  $T_{i,j}$  compares with the warmest areas  $T_{high}$  and the coldest areas  $T_{low}$  of the same surface type ( $i, j$  are the pixel coordinates in the composite image). According to the empirical T-VI method, the observable takes the form of a temperature *relative* contrast:  $WI = (T_{high} - T_{i,j}) / (T_{high} - T_{cold})$ . By the way, the interpretation that a temperature relative change between the limits  $T_{low}$  and  $T_{high}$  is univocally related to a moisture change, assumes that all considered areas have received the same amount of thermal energy for a given period of time, in particular the same solar radiation flux density. This constraint will be discussed in more details later.

- The temperature should be compared between areas of a same surface type, which theoretically means areas of the same cover type and the same vegetation cover fraction. Practically, the surface type is characterized by the vegetation index, which is itself a *relative* measurement of a contrast between two spectral reflectance values in the red and near infrared. The NDVI is defined as  $(\rho_{\text{NIR}} - \rho_{\text{Red}})/(\rho_{\text{NIR}} + \rho_{\text{Red}})$ . The index OSAVI is defined as  $(\rho_{\text{NIR}} - \rho_{\text{Red}})/(\rho_{\text{NIR}} + \rho_{\text{Red}} + C)$  where C is a constant (C=0.16) (Rondeaux [7], 1996). Notice that OSAVI is not exactly a relative contrast because of the presence of the constant C, it is somehow a biased relative contrast.

Since all measurements are performed by optical remote sensing, there are a series of difficulties in obtaining a precise evaluation of the temperature on the one hand and the spectral reflectance on the other hand. Regarding the temperature first, to obtain a precise value of the absolute temperature, one should first perform a correction related to the atmospheric attenuation (absorption, diffusion, ...) in the spectral band of the considered TIR camera. Then, one should perform a temperature-emissivity separation since the TIR emitted radiance not only depends on temperature but also on the soil surface emissivity. Both are actually unknown and we face an under-determined problem.

A series of multispectral methods have been proposed in the past, which are based on empirical relationships between the radiance temperatures measured at the considered wavelengths with possibly an estimation of the emissivity through the measurement of NDVI. Whatsoever, in the present case of single-band thermal radiance sensing, one could consider a first-order linear correction of the type  $T_{i,j} = a(\varepsilon_{i,j})S_{i,j} + b(\varepsilon_{i,j})$  where  $S_{i,j}$  is the recorded signal at pixel  $i, j$ , and  $a(\varepsilon_{i,j})$ ,  $b(\varepsilon_{i,j})$  are two coefficients that depend on the local emissivity  $\varepsilon_{i,j}$  and many other parameters like the sky downward TIR radiance, the atmosphere transmission, the instrument conversion factors, a.s.o.

An atmospheric propagation code like MODTRAN or MATISSE would be necessary to compute the atmospheric contributions (sky downward TIR radiance, atmospheric transmission). Hopefully, important simplifications will be introduced. We will admit that all the latter parameters are constant over the whole analysed zone, namely for all flight lines that will be combined in the "triangle" construction. For that purpose, it is important that the sky state doesn't change during these flight lines (clear sky or uniform and stable cloud cover; intermittent obscuring by clouds should be avoided). That the sky state doesn't change is also important from an energy point of view (recall that the temperature differences between two points can be related with their moisture differences only if they have received the same amount of solar radiation for the past minutes or even past hours).

The only variable parameter in the relationship between TIR signal and the true temperature is emissivity, which appears through the expressions  $a(\varepsilon_{i,j})$ ,  $b(\varepsilon_{i,j})$ . Thermal infrared emissivity varies from bare soil to vegetation. It also varies, generally to a lesser degree, between different types of soils on one side and different types of vegetation, on the other side.

However, with the present affordable data, discrimination between these surfaces could only be made through the visible and near infrared reflectance, and preferably through the inferred vegetation index because of its *relative* character. This makes that the relation between temperature and TIR signal becomes  $T_{i,j} = a(VI_{i,j})S_{i,j} + b(VI_{i,j})$ . What is now important to observe is that when computing the water index  $WI$  through the relative contrast  $WI = (T_{high} - T_{i,j}) / (T_{high} - T_{cold})$  we compare temperature levels obtained over areas of *the same vegetation index*. As a direct consequence, it is clear that when substituting in the expression of  $WI$  each temperature  $T_{i,j}$ ,  $T_{high}$  and  $T_{low}$  by the corresponding expression involving the remote sensing raw signal, i.e.  $S_{i,j}$ ,  $S_{high}$  and  $S_{low}$  the coefficients  $a$  and  $b$  cancel. Hence, at first order:  $WI = (S_{high} - S_{i,j}) / (S_{high} - S_{cold})$ , which means that we can directly use the raw IRT signal, instead of the absolute temperature.

Finally, we can do without the following operations: atmospheric correction, calibration of the thermal camera, and emissivity evaluation. Note however that it is of utmost importance that all factors that intervene in the atmosphere TIR contribution and in the camera calibration should be *constant* throughout the remote sensing recording (no cloud-cover change as mentioned before, no camera drift, no automatic gain or offset, a.s.o.). Let us add that the remote sensing recording in question entails all flights lines that contribute to a single triangle construction. If one of the factors has changed (sky cover or camera settings) a new “triangle” scatterplot should be built.

Let us now consider the vegetation index. The NDVI is most popular since it is rather immune to atmospheric variations, provided they affect equally the red channel and the NIR channel. In the course of a preliminary calibration performed on ground with a highly reflective plate of reflectance  $\rho_{ref,Red}$  and  $\rho_{ref,NIR}$  in the red and NIR bands, the recorded signal in these bands is  $S_{ref,Red} \propto \rho_{ref,Red} L_{Red}^{\downarrow}$  and  $S_{ref,NIR} \propto \rho_{ref,NIR} L_{NIR}^{\downarrow}$  where  $L_{Red}^{\downarrow}$ ,  $L_{NIR}^{\downarrow}$  is the red and NIR downward solar radiance. Then, during the flight, the measured signals over a given area are  $S_{Red} \propto \tau_{Red} \rho_{Red} L_{Red}^{\downarrow}$  and  $S_{NIR} \propto \tau_{NIR} \rho_{NIR} L_{NIR}^{\downarrow}$  where  $\tau$  is the atmospheric transmission from ground to airborne sensor and  $\rho_{Red}$ ,  $\rho_{NIR}$  is the soil reflectance in both bands.

To get the NDVI defined by  $NDVI = (\rho_{NIR} - \rho_{Red}) / (\rho_{NIR} + \rho_{Red})$  we need to compute:

$$\left( \frac{S_{NIR} \rho_{ref,NIR} / \tau_{NIR} S_{ref,NIR} - S_{Red} \rho_{ref,red} / \tau_{red} S_{ref,red}}{S_{NIR} \rho_{ref,NIR} / \tau_{NIR} S_{ref,NIR} + S_{Red} \rho_{ref,red} / \tau_{red} S_{ref,red}} \right)$$

Each value of the raw signal  $S_{NIR}$  and  $S_{Red}$  is thus multiplied by a compensation factor  $\rho_{ref,NIR} / \tau_{NIR} S_{ref,NIR}$ , resp.  $\rho_{ref,red} / \tau_{red} S_{ref,red}$ .

Since the airborne platforms will fly at low altitude, from 50m to 150m for the UAV and in the range of 1000-2000m for the airplane, it is expected that the transmission coefficients  $\tau_{NIR}$  and  $\tau_{red}$  will be close to 1. Actually it is sufficient that they are close one to the other to cancel from the NDVI expression, in a first approximation. In the general case, it would be necessary to compute  $\tau_{NIR}$  and  $\tau_{red}$  with an atmospheric propagation code.

Furthermore, in the absence of calibration with a reference plate, it would also be necessary to compute the solar downward radiance at ground level  $L_{\text{Red}}^{\downarrow}$ ,  $L_{\text{NIR}}^{\downarrow}$ . The NDVI would then be obtained from  $\left( S_{\text{NIR}} / \tau_{\text{NIR}} L_{\text{NIR}}^{\downarrow} - S_{\text{red}} / \tau_{\text{red}} L_{\text{Red}}^{\downarrow} \right) / \left( S_{\text{NIR}} / \tau_{\text{NIR}} L_{\text{NIR}}^{\downarrow} + S_{\text{red}} / \tau_{\text{red}} L_{\text{Red}}^{\downarrow} \right)$ .

On the other hand, OSAVI is a variant aimed to minimize the influence of soil brightness (Rondeaux, 1996) [7]. This index is recommended in areas with relatively sparse vegetation where soil is visible through the canopy. The originally proposed (central) wavelengths were  $0.67\mu\text{m}$  for the red band and  $0.80\mu\text{m}$  for the NIR band. To obtain the OSAVI, one simply adds the constant 0.16 to the denominator of the NDVI formula.

Originally, in the empirical T-VI method, the basic idea was that the WI scale should be defined with respect to the extreme temperatures  $T_{\text{low}}$  and  $T_{\text{high}}$ , with the interpretation that these should be related to the wettest areas, resp. the driest areas (Sandholt, 2002) [8]. However, these extreme temperature levels depend on many other factors, not only moisture, in particular on the vegetation cover fraction. In order to remove this ambiguity (at least partially), it has been suggested to apply different scales depending on the VI. The simplest approach considers a linear variation of the scale boundaries with the NDVI. Let us now introduce a new option. Assuming that the factors  $\tau_{\text{NIR}} S_{\text{ref},\text{NIR}}$  and  $\tau_{\text{red}} S_{\text{ref},\text{red}}$  (or similarly  $\tau_{\text{NIR}} L_{\text{NIR}}^{\downarrow}$  and  $\tau_{\text{red}} L_{\text{Red}}^{\downarrow}$ ) are constant for the considered flight lines (it would be sufficient again that the red+NIR radiation environment be uniform and constant throughout the remote sensing measurements).

The proposal is to consider a linear variation of the scale boundaries not with the real NDVI but with the “apparent” NDVI, which we define directly from the raw signals:  $(S_{\text{NIR}} - S_{\text{red}}) / (S_{\text{NIR}} + S_{\text{red}})$ . It is even not necessary to calibrate the signal against radiance, it is sufficient that both signals are proportional to radiance (with possibly different coefficients). Of course, since the link with the true reflectance has been broken, it is quite difficult to guess the value of the constant that should be added at the denominator to reproduce the OSAVI bare soil reflectance compensation. For that reason, we recommend to keep the standard NDVI definition with raw signal values instead of reflectance values.

This approach is even more “rustic” than the empirical T-VI method. It has the advantage of not requiring a calibration procedure or computation of the atmospheric contributions (solar radiance and atmospheric transmission). Nevertheless, this simplified method should be validated with the images obtained during the previous campaigns and one should evaluate its impact on the inferred water index.

As a conclusion on the pre-processing needs regarding atmospheric correction, we can argue that, as a first approximation, we can input raw images in the triangle T-VI procedure. No atmospheric correction is required, and a minimal instrument calibration is sufficient.

## 4 Data post-processing

At this stage we assume that mosaics of RED images, NIR images and TIR images have been separately created, one of these three composite images has been georeferenced and the other two have been registered with respect to the first one. We thus have at hand three composite images that are perfectly overlapping. Mathematical operations can then be performed on these three signals.

The perfect overlapping mentioned before is important since otherwise when combining the signals from the three images, we will actually mix the responses from neighbour points at ground. The purpose was however to combine the responses from the same point at ground. The consequence may be negligible over nearly homogeneous areas. On the contrary, it may be catastrophic in those areas where one or all signals vary rapidly in space as, for example, in the vicinity of field boundaries or along road edges.

For the following description, we will consider, as an illustration, the remote sensing results obtained in July 11th over Rians area. The intermediate results will be provided to explain the different steps in data processing.

### 4.1 Vegetation index

As explained in § 3, if a reflectance calibration has been performed to obtain an estimation of the downward solar radiance in both Red and NIR bands (in terms of numerical signal in both channels) an estimation of the NDVI is obtained by computing the relative contrast in reflectance (which explains the following subscript  $\rho$ ):

$NDVI_{\rho} = (\rho_{NIR} - \rho_{Red}) / (\rho_{NIR} + \rho_{Red})$  where  $\rho_{NIR} = S_{NIR} \rho_{ref,NIR} / \tau_{NIR} S_{ref,NIR}$  and  $\rho_{red} = S_{red} \rho_{ref,red} / \tau_{red} S_{ref,red}$ .

Since transmission is nearly the same in both bands, NDVI can be approximated by: NDVI

$NDVI_{\rho} \approx (\rho'_{NIR} - \rho'_{Red}) / (\rho'_{NIR} + \rho'_{Red})$  where appear now the “underestimated reflectances”

$\rho'_{NIR} = S_{NIR} \rho_{ref,NIR} / S_{ref,NIR}$  and  $\rho'_{red} = S_{red} \rho_{ref,red} / S_{ref,red}$ . As an option we can also compute the

OSAVI index:  $OSAVI_{\rho} = (\rho_{NIR} - \rho_{Red}) / (\rho_{NIR} + \rho_{Red} + 0.16)$ . The interest in using OSAVI instead of

NDVI (better immunity to variations of the bare soil reflectance – Rondeaux 1996) [7] must

be weighed against the requirement to have access to the true reflectance values  $\rho_{NIR}$

and  $\rho_{red}$ . Otherwise, we must fall back to the approximation

$NDVI_{\rho} \approx (\rho'_{NIR} - \rho'_{Red}) / (\rho'_{NIR} + \rho'_{Red})$ . As explained in §3, a further approximation consists in

computing the relative contrast of the raw Red and NIR signals themselves:

$NDVI_{\rho} \approx NDVI_S = (S_{NIR} - S_{red}) / (S_{NIR} + S_{red})$ .

The next illustration is with the indexes  $NDVI_{\rho}$  and  $OSAVI_{\rho}$ . As a matter of fact, during the

July 2017 campaign, the spectral images provided by a hyperspectral Hypspec camera

have been processed by a code developed at ONERA which, after correction of the

atmospheric attenuation, provides “true” reflectance images.

The reflectance images obtained at 0.662 $\mu\text{m}$  and 0.858 $\mu\text{m}$  over the Rians area are presented in Figure 3 and Figure 4 (notice that since the reflectance of vegetation is rather flat above 0.80-0.82 $\mu\text{m}$ , it may be advisable to move the NIR band to a higher wavelength and benefit for the possibility of using a wider band, thus increasing the signal-to-noise ratio). The corresponding  $NDVI_\rho$  image is in Figure 5 and the  $OSAVI_\rho$  image is in Figure 7. These images look quite similar. The essential difference is that the  $NDVI_\rho$  scale spans over [-0.79, 0.97] whereas the  $OSAVI_\rho$  scale spans over a smaller interval, namely [-0.40, 0.78]. We can already say that this simple difference cannot have a large influence on the result of the triangle method. In any case, the triangle method will be applied successively with  $NDVI_\rho$  and  $OSAVI_\rho$  to check the possible differences.

In addition, we will also consider a simulated image of the “raw” vegetation index  $NDVI_S$ . The purpose is to evaluate the impact of using the raw radiometric signal in the VI definition instead of the reflectance as it should theoretically be done. If the difference is insignificant, we could propose to implement the computation of  $NDVI_S$  in the operational post-processing code. However, the raw radiometric signals provided by the (pushbroom) hyperspectral camera on one side (former campaigns) and the raw radiometric signals provided by the (snapshot) VNIR cameras that will be used on board of the UAV and the airplane have substantial differences.

For this reason, it is difficult to simulate what could have been the  $NDVI_S$  image if, instead of the Hypsux camera, we had used the snapshot cameras. Nevertheless, one can notice that  $NDVI_\rho$  can be expressed in terms of the raw signals in a manner that is very similar to what is done for  $NDVI_S$  except the presence of a corrective factor  $\beta$ , namely  $NDVI_\rho = (S_{\text{NIR}} - S_{\text{red}}\beta)/(S_{\text{NIR}} + S_{\text{red}}\beta)$  with  $\beta = \rho_{\text{ref},\text{red}}\tau_{\text{NIR}}S_{\text{ref},\text{NIR}}/\rho_{\text{ref},\text{NIR}}\tau_{\text{red}}S_{\text{ref},\text{red}}$ . Conversely, the “raw” vegetation index  $NDVI_S$  can be expressed in terms of the true reflectance values according to:  $NDVI_S = (\rho_{\text{NIR}}\beta - \rho_{\text{red}})/(\rho_{\text{NIR}}\beta + \rho_{\text{red}})$ .

As a consequence,  $NDVI_S$  can be simulated from the knowledge of  $\rho_{\text{red}}$  and  $\rho_{\text{NIR}}$ , which have been plotted in Figure 3 and Figure 4, by assigning to  $\beta$  different values in the range of 1. A parametric analysis will then be performed to evaluate the influence on the water index, depending on the assumed value for the corrective factor  $\beta$ . Two examples have been plotted in Figure 9 and Figure 10 by taking  $\beta=0.8$ , resp.  $\beta=1.2$ . The modifications induced in the VI seem to be of small extent.

At first sight one can notice a slight translation of the VI scale downwards, resp. upwards. We can take a closer look at the corresponding VI histograms in Figure 11, Figure 12, and Figure 13. When comparing the original histogram in Figure 11 for  $\beta=1$  and the two other ones in Figure 12 for  $\beta=0.8$  and in Figure 13 for  $\beta=1.2$  we can notice a global expansion/shrinking of the main pattern when  $\beta$  is decreased/increased. We can argue that such a global transformation will have no significant impact in the next triangle processing (the reason invoked is again related to the *relative* character of the method).

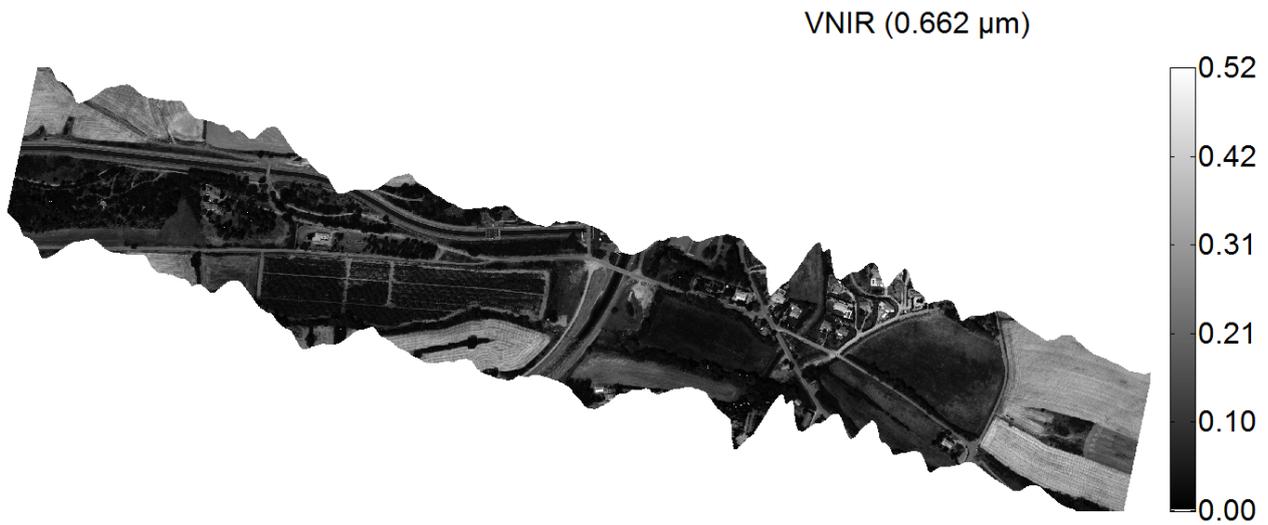


Figure 3: Map of the spectral reflectance at 0.662  $\mu\text{m}$  obtained by the pushbroom hyperspectral camera Hyspex during WADI 2017 July 11<sup>th</sup> campaign over Rians area.

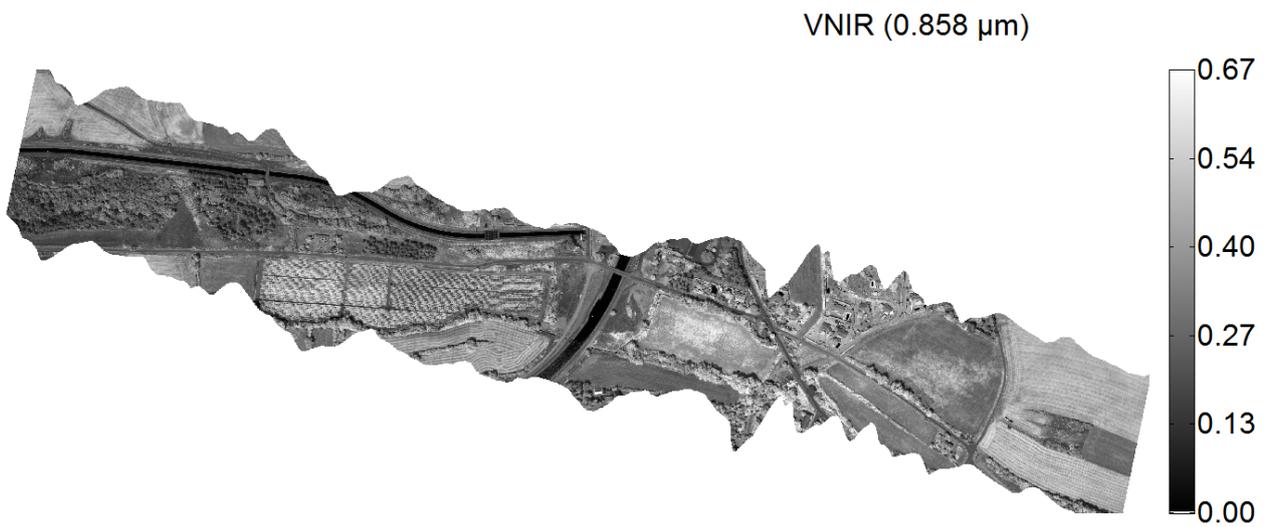


Figure 4: Map of the spectral reflectance at 0.858  $\mu\text{m}$  obtained by the pushbroom hyperspectral camera Hyspex during WADI 2017 July 11<sup>th</sup> campaign over Rians area.

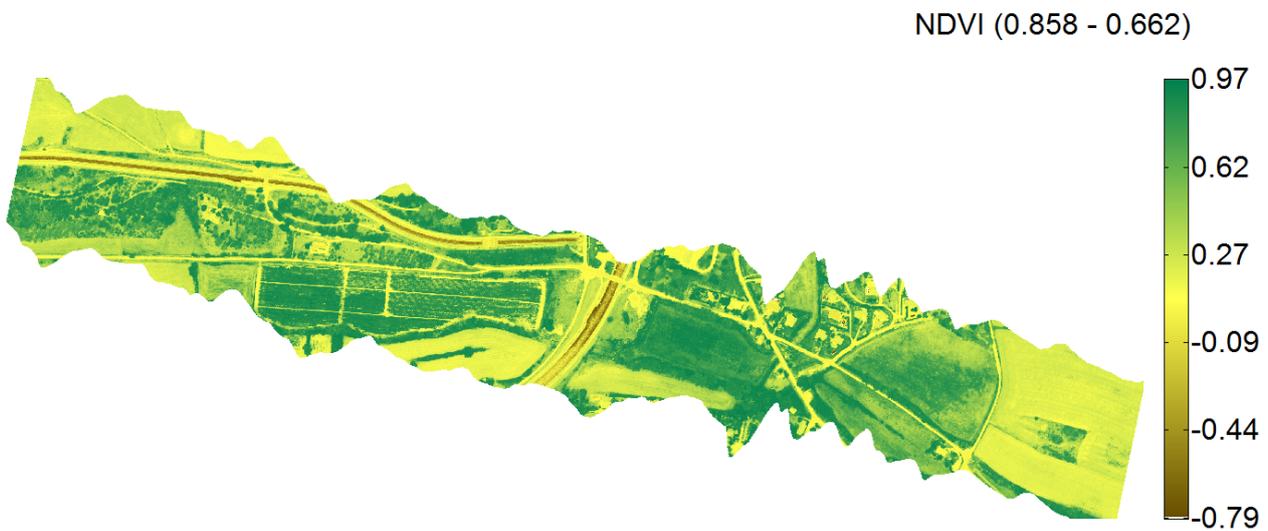


Figure 5: Vegetation index  $NDVI_{\rho}$  obtained from the spectral reflectance maps in Figure 3 and Figure 4.

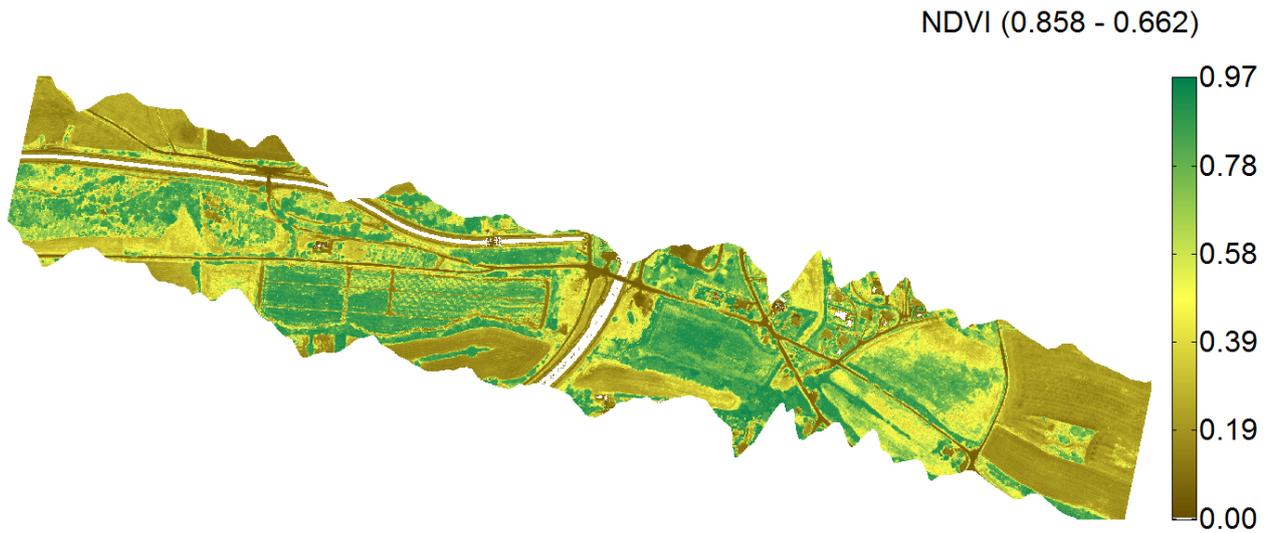


Figure 6: Vegetation index  $NDVI_{\rho}$  obtained from the spectral reflectance maps in Figure 3 and Figure 4. Here the range has been limited to  $NDVI_{\rho} > 0$ ; water surfaces (open canals, ponds, swimming pools) are now out of scale.

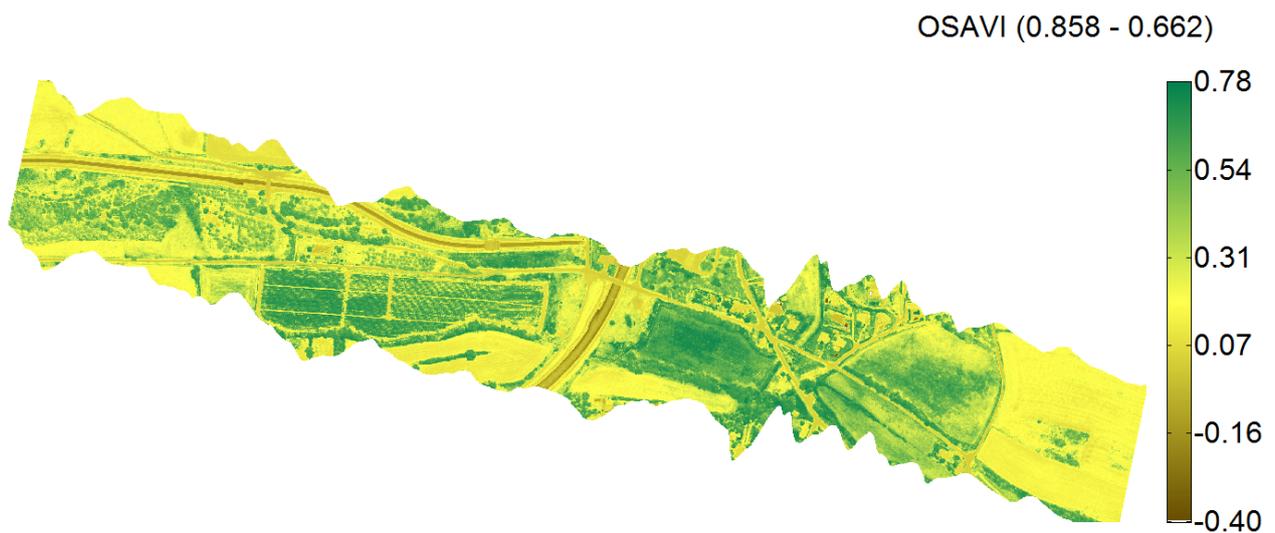


Figure 7: Vegetation index  $OSAVI_{\rho}$  obtained from the spectral reflectance maps in Figure 3 and Figure 4.

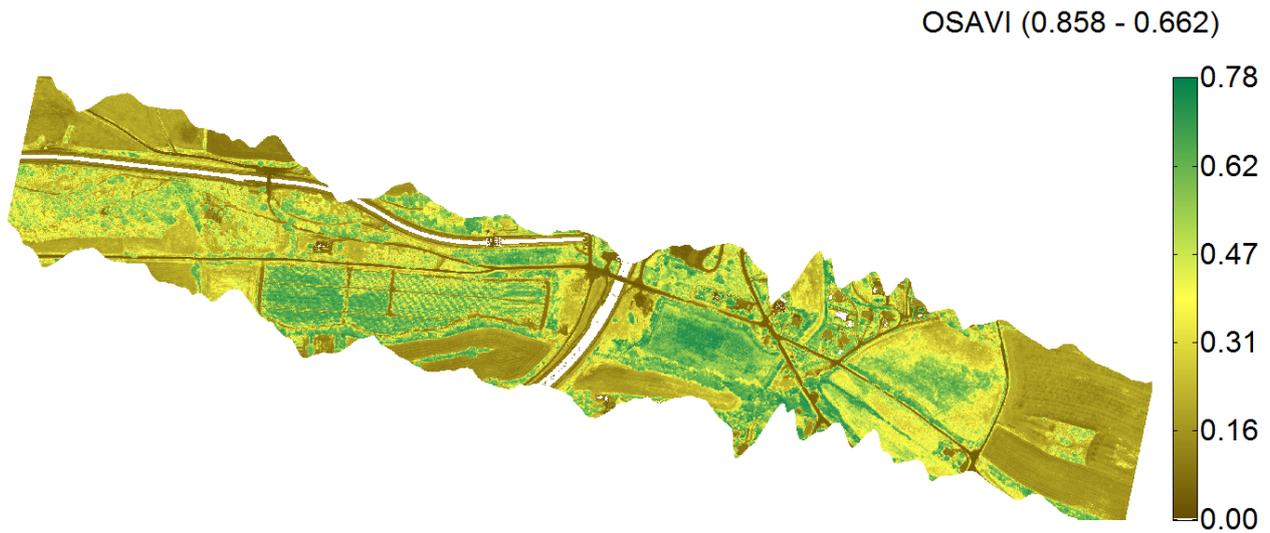


Figure 8: Vegetation index  $OSAVI_p$  obtained from the spectral reflectance maps in Figure 3 and Figure 4. Here the range has been limited to  $OSAVI_p > 0$ ; water surfaces (open canals, ponds, swimming pools) are now out of scale.

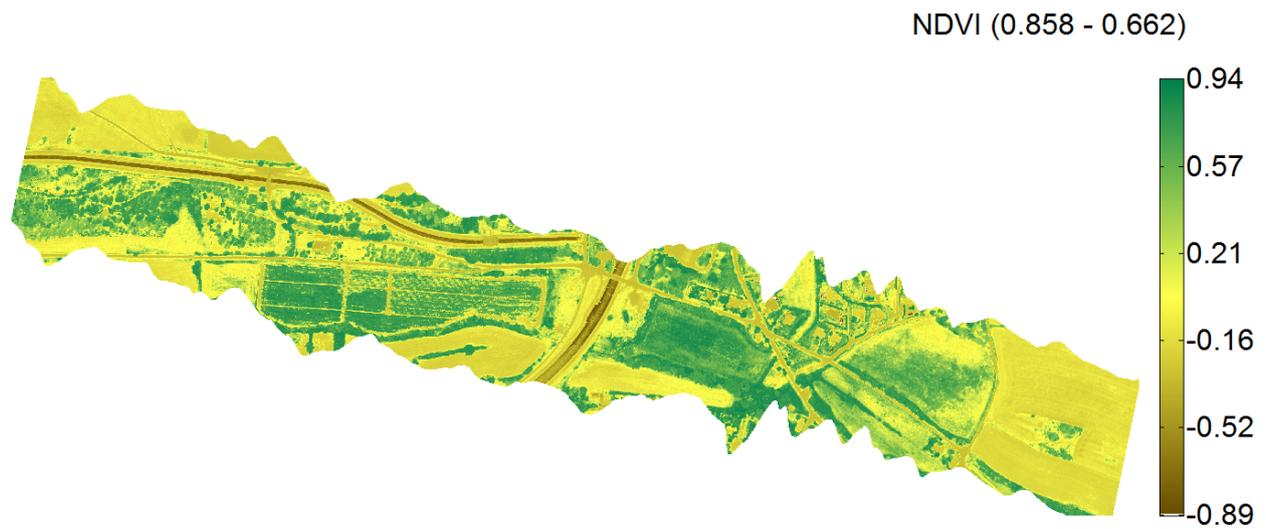


Figure 9: "Raw" vegetation index  $NDVI_s$  obtained directly from the raw radiometric signals. The corrective factor that was used to simulate  $NDVI_s$  from the reflectance maps in Figure 3 and Figure 4 was set to 0.8.

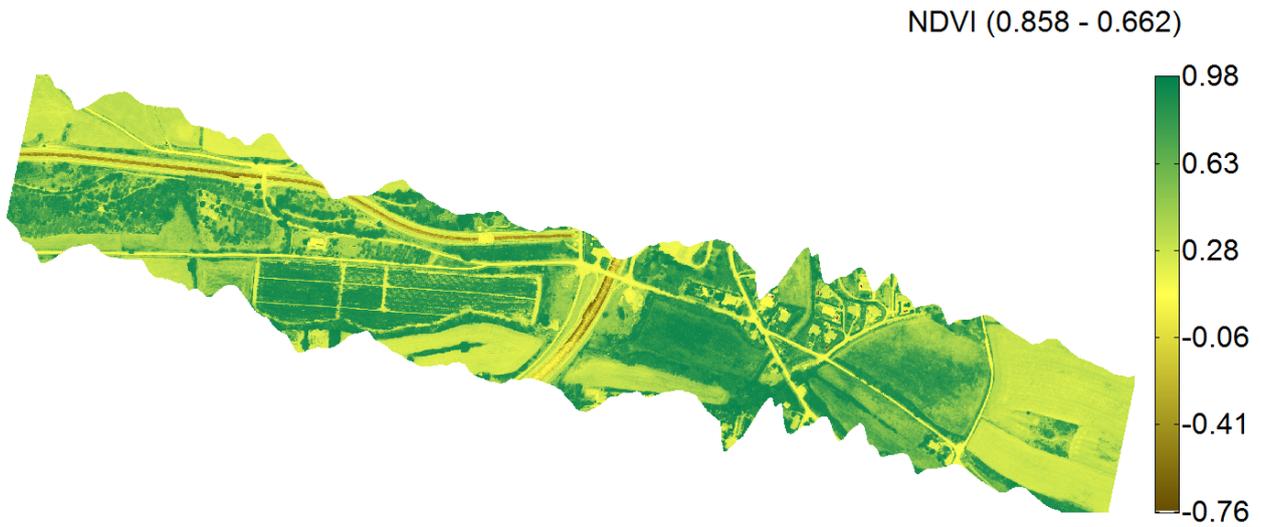


Figure 10: "Raw" vegetation index  $NDVI_S$  obtained directly from the raw radiometric signals. The corrective factor that was used to simulate  $NDVI_S$  from the reflectance maps in Figure 3 and Figure 4 was set to 1.2.

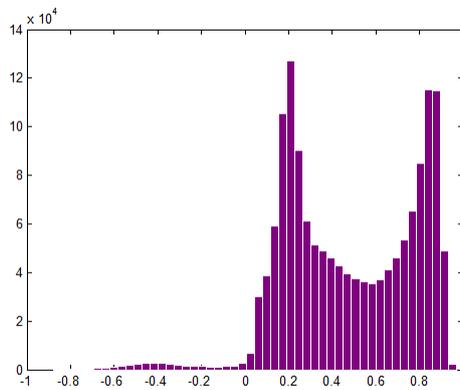


Figure 11: Histogram of the  $NDVI_\rho$  map in Figure 5 (VI obtained from the reflectance values).

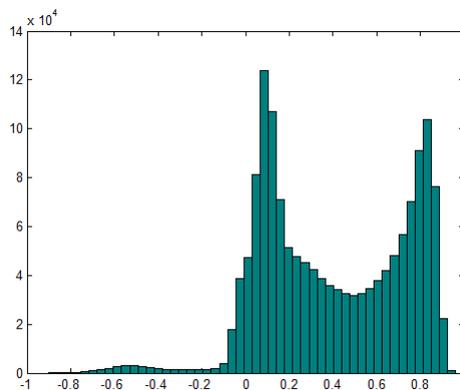


Figure 12: Histogram of the  $NDVI_S$  map in Figure 9 (simulation of a VI obtained from the raw signal values with  $\beta=0.8$ ).

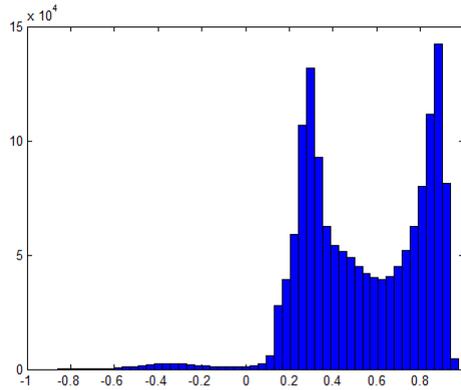


Figure 13: Histogram of the  $NDVI_S$  map in Figure 10 (simulation of a VI obtained from the raw signal values with  $\beta = 1.2$ ).

## 4.2 Temperature-VI scatterplot

The triangle/trapezoid method for moisture mapping has already been described in D2.1 and D3.1. Let us simply recall that when we plot the two-dimensional distributions of temperature and vegetation index (T-VI) corresponding to an area with well distributed vegetation cover and moisture content, we obtain a scatter of triangular or trapezoidal shape (see Figure 14). The four vertices of the trapezoidal distribution correspond to the extreme conditions of a crop (Moran 1994 [6]): 1: well-watered full cover, 2: non-transpiring full cover, 3: wet bare soil, 4: dry bare soil. Soil moisture increases from the “dry” edge on the right (which is also named the “cold” edge) to the “wet” edge on the left (which is also named the “warm” edge).

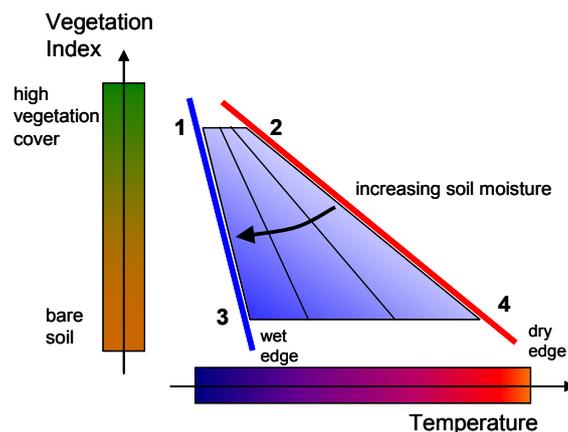


Figure 14: Principle of soil moisture mapping from a Temperature-Vegetation Index (T-VI) plot (Moran 1994 [6], Sandholt 2002 [8], Krpez, 2005, 2009, 2012, [1] [4] Chatelard, 2018) [5].

Practically, the temperature and VI data that are combined in such scatterplot should be obtained from one or more flight lines satisfying the following conditions:

- The environmental energy-forcing conditions should be the same for all analyzed areas for a past period of time in the range of say, one hour (in particular, the same solar radiation flux and the same wind speed history).
- The data set should contain a high diversity of situations in terms of soil moisture and vegetation cover. Notice that the latter condition is not compulsory. Indeed, an exception is when the region is uniform in terms of vegetation cover and vegetation fraction. In that case the trapezoid scatterplot is restricted to a thin horizontal strip and the identification is almost exclusively based on the temperature data.

The T-VI method, its multiple variants and their justifications have been reviewed in D2.1 and D3.1. Next we will merely describe the essential features for an efficient implementation of the empirical approach.

#### 4.2.1 2D scatterplot construction

In principle, to construct the 2D T-VI scatterplot, we must plot the temperature versus the vegetation index for all points that has been viewed during the remote sensing experiment, namely for all points of the mosaicked and registered temperature and VI images. As pointed out in § 3, because of the *relative* character of the method, the absolute temperature can be replaced by any variable that is in linear relation with it. It is then tempting to replace the temperature by the raw TIR signal: at any VI level, as a first approximation, the pixels are similarly distributed between the extreme values  $T_{low}$  and  $T_{high}$  on one side and between  $S_{low}$  and  $S_{high}$  on the other side. Nevertheless, it is not excluded that the trapezoid scatterplot is slightly distorted (horizontal shear) when changing from temperature to raw TIR signal. An emissivity dependence on VI is a possible source of such distortion. This point will be discussed later. Anyway, for the while, we will assume that the T-VI scatterplot is built directly from the raw TIR signal.

The mosaic of TIR images that has been considered for the illustrative case is reported in Figure 15.

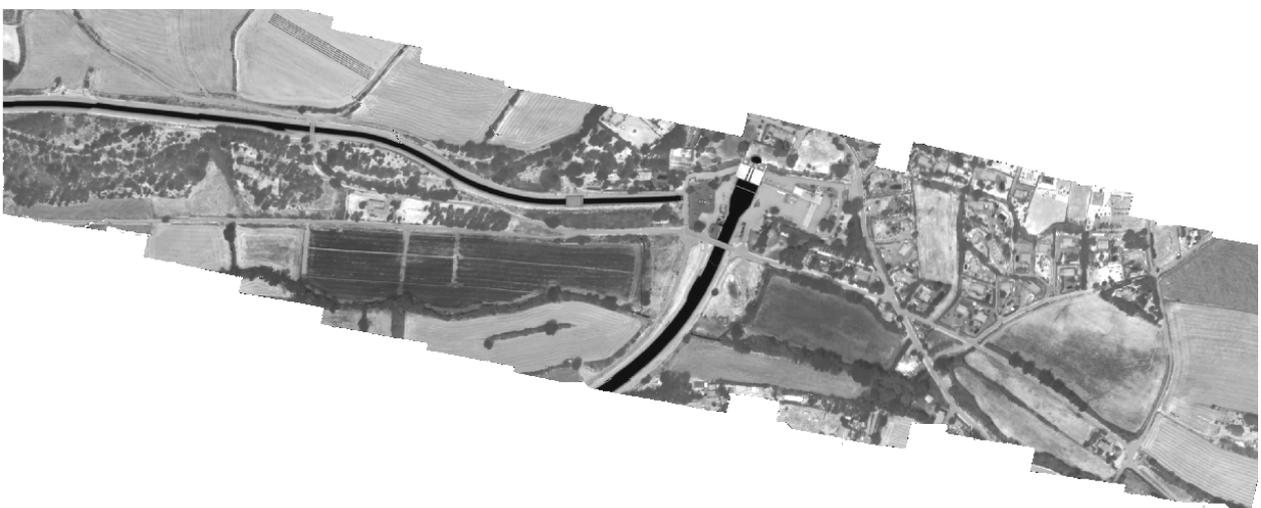


Figure 15: Mosaic of TIR images. Raw signal expressed in grey tones. The darkest areas correspond to the coldest surfaces (i.e. water surfaces like canals).

When reporting the normalized temperature (actually the normalized TIR signal) against the vegetation index  $NDVI_{\rho}$  we obtain the 2D scatterplot in Figure 16. We can observe a rather compact cloud of data in the top right part of the plot and a very thin vertically elongated spindle in the bottom left part (as delineated with a blue ellipse). This group of data of negative NDVI values and very low temperature actually corresponds to wide water surfaces (canals) (see the corresponding blue pixels in Figure 17). A small ghost cloud (see the red ellipse) appears to be related to the edges of these water surfaces (see the corresponding red pixels in Figure 17). Most likely it is an artefact due to non-perfect matching of the TIR and VNIR images during the registration process. In that instance an inappropriate association is made between water pixels in the TIR image and bank pixels in the VI image and vice-versa. This makes that in this location, water temperature values are (wrongly) associated with soil VI values and, symmetrically, soil temperature values are (wrongly) associated with water VI values.

In the same way, the top right region of the 2D scatter plot is suspected to contain “ghost” points originating from another registration mismatch as well. All points above the green slanted line in Figure 16 have been marked as green pixels in the temperature map in Figure 17. As a matter of fact, most of them are related to field edges, individual tree edges, grove edges and paths/trails edges. Nevertheless, there are other groups of green pixels that are not related to edge effects and that, on the contrary, cover large portions of fields; for this reason, they should not be rejected.

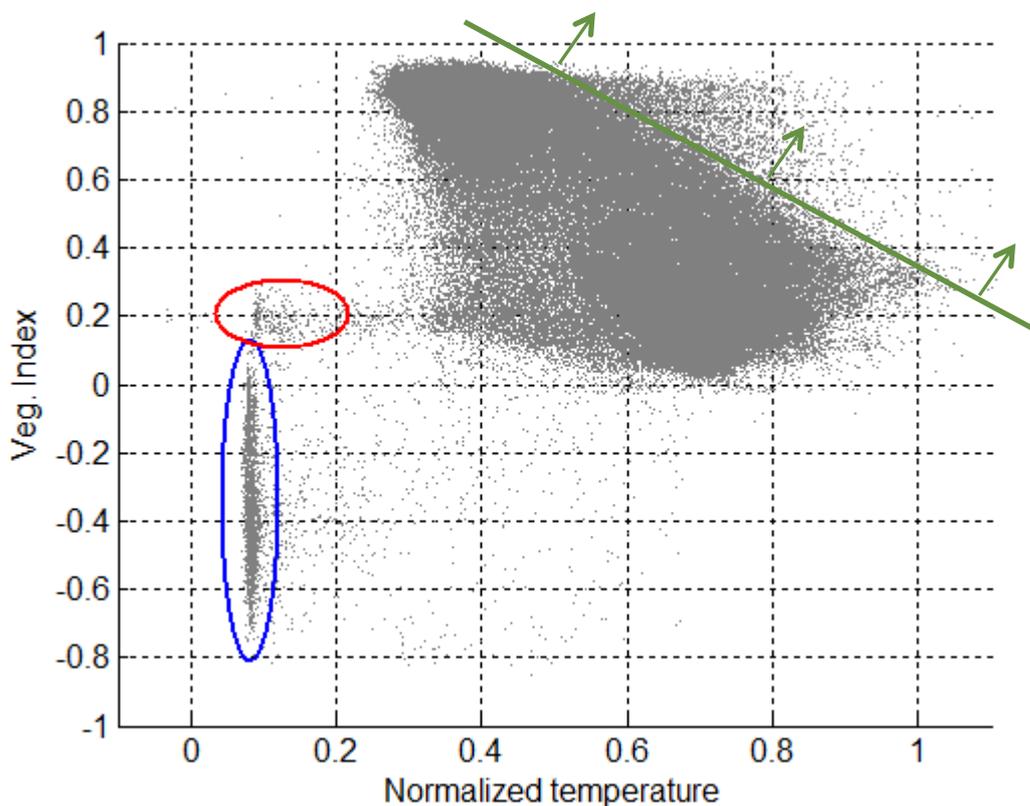


Figure 16: 2D scatterplot when considering 1 in 10 pixels in the T and VI mosaics. The points marked with a blue ellipse correspond to wide water surfaces (canals). The points marked with a red ellipse are likely artifacts due to non-perfect matching between T and VI images at the boundaries of these water surfaces (see Figure 17). The region above and right of the green line is also suspected to contain artifacts.

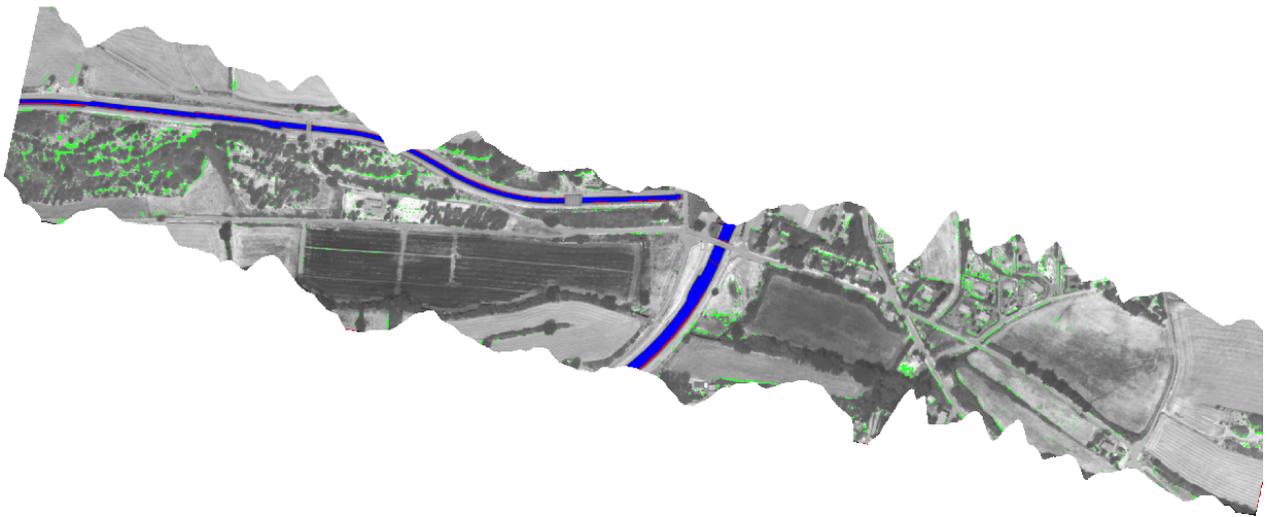


Figure 17: Temperature image, actually the “apparent” temperature over the common areas that have been sensed by the VNIR camera and the TIR camera) where have been marked in blue and red the points delineated with blue and red ellipses in Figure 16. The green pixels correspond to the top right region in Figure 16 (above the green line). Most of them (but not all) correspond to edge effects.

Depending on the size of the mosaics, the 2D scatterplot may be bulky. The density of points can be so high that it becomes difficult to estimate the right position of the boundaries. As a matter of fact, there are always outliers originating from areas of no interest (water surfaces, roads, paths, shadows, buildings, ...). When their representation is too dense they may be confused with areas of interest and thus integrated within the trapezoid boundaries. An easy solution consists in reducing the number of points in the 2D scatterplot. Incidentally, it provides files easier to handle because of smaller size. To fix the ideas, one in every 50 pixels up to one in every 100 pixels has been considered when processing the images of the 2017 campaign. This parameter should be adjusted depending on the size of the probed region, the number of pixels of each camera, and finally, the number of pixels in the resulting mosaics (see Figure 18 to Figure 20 where we considered 1/10, 1/50 and 1/100 pixels).

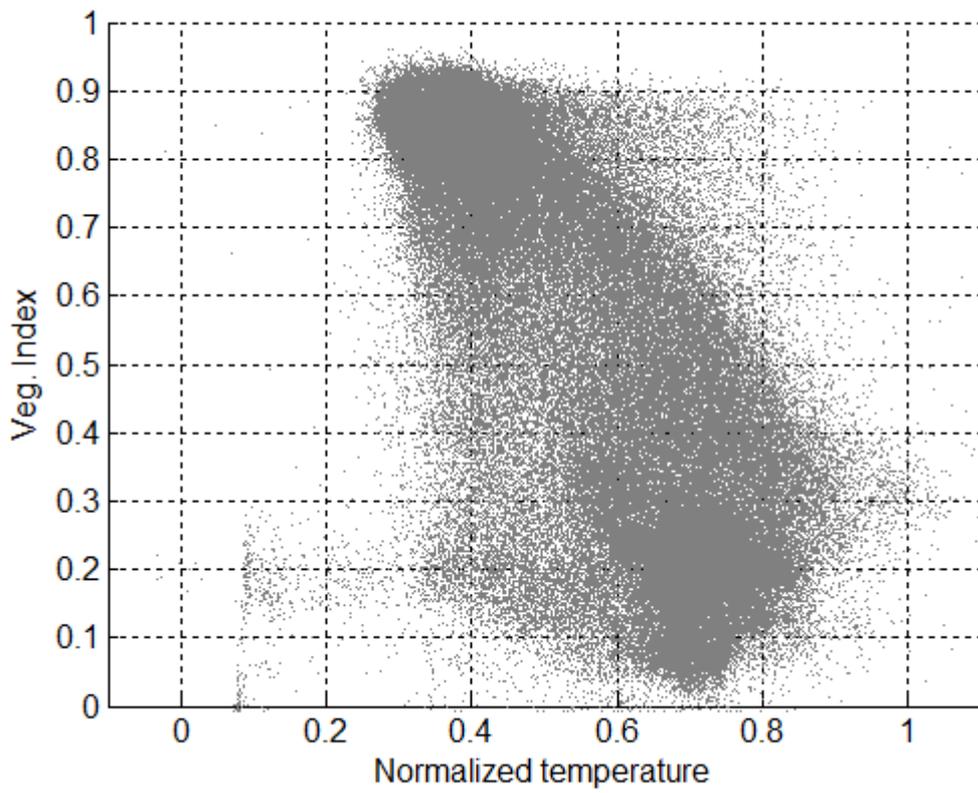


Figure 18: 2D scatterplot when considering 1 in every 10 pixels in the T and VI mosaics.

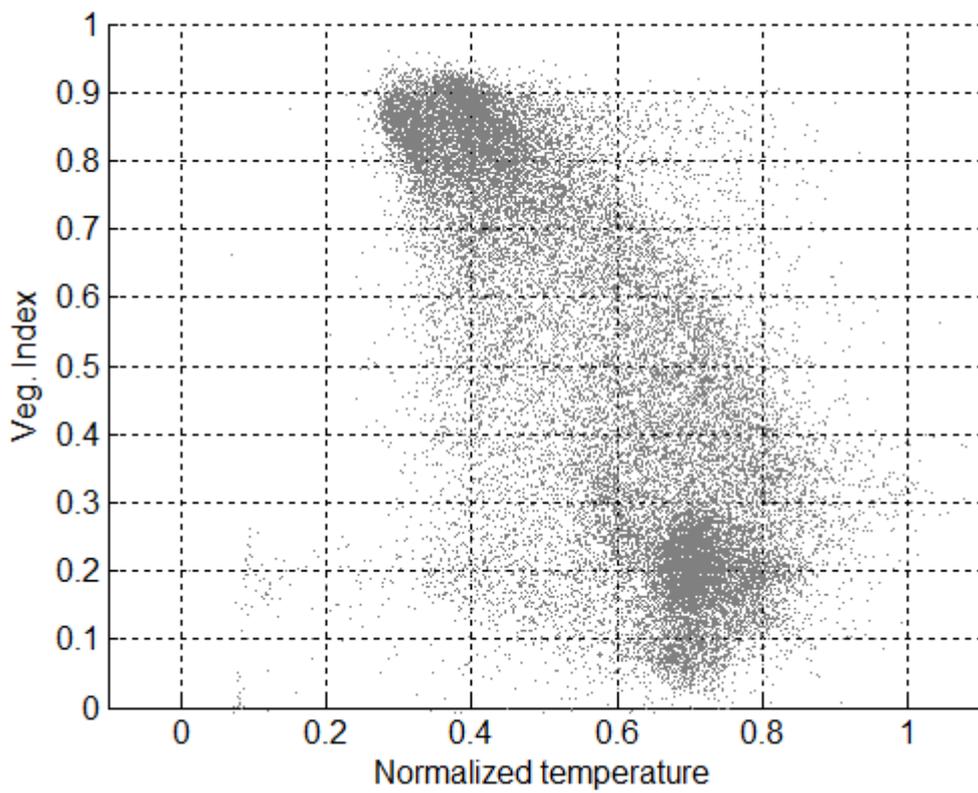


Figure 19: 2D scatterplot when considering 1 in every 50 pixels in the T and VI mosaics.

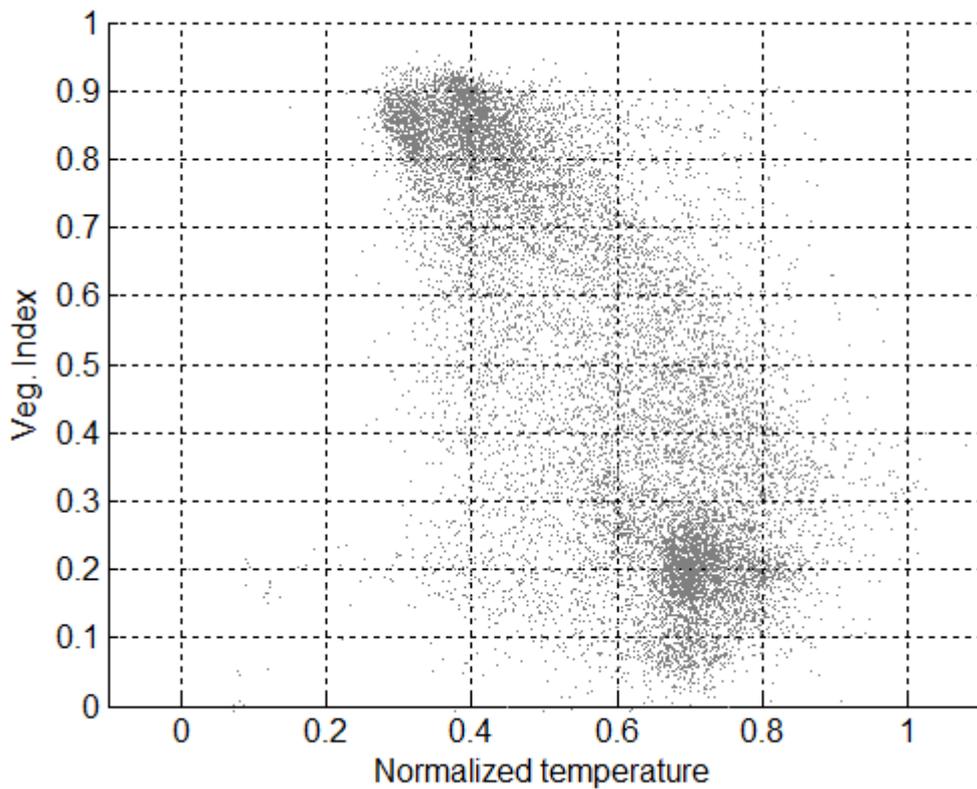


Figure 20: 2D scatterplot when considering 1 in every 100 pixels in the T and VI mosaics.

#### 4.2.2 Trapezoid boundary determination and Water Index calculation

There are several possibilities to set the cold and warm edges of the 2D scatterplot. We can first try an automatic method and if the results are not satisfactory, we can fall back on a manual method.

Figure 21 is the result of the application of an automatic method. Notice that we need first to define the sub-domain that will be taken into account for the fittings. For this purpose, it may be necessary to discard the areas that have no interest as for example those marked with blue and red ellipses in Figure 16. The principle is first described for the cold edge (dashed blue line in Figure 21):

- Select an amplification factor  $K$ . This factor will be applied to the temperature difference of the points that are left from the blue curve (no amplification factor will be applied to the points that are right from the blue curve). A typical value for  $K$  is 50.
- Solve the non-linear optimization problem of finding the equation of the straight line  $T(VI) = \alpha VI + \beta$  with unknown parameters  $\alpha$  and  $\beta$  such that the following cost function is minimized:  $J = \sum_{d_{i,j} > 0} d_{i,j} - K \sum_{d_{i,j} < 0} d_{i,j}$  where  $d_{i,j}$  is the algebraic distance to the cold edge defined by:  $d_{i,j} = T_{i,j} - T(VI_{i,j})$

To split the sum into two parts and to multiply the sum of negative algebraic distances by  $-K$  where  $K \gg 1$ , allows the cold edge to stick to the 2D scatterplot on the left side without much penetration. Some points (outliers) may stay to the left of the cold edge. By playing with the  $K$  factor we may tolerate more or less points on the left side of the cold edge. In a way, this method amounts to finding the centroid line of the whole 2D-scatterplot knowing that the points at left of the (blue) line are artificially stunned by a factor  $K$  (note that a pseudo-torque is applied since instead of the true distance to the line we use the temperature difference).

A symmetrical analysis has been performed for the warm edge (dashed red line in Figure 21). This time, the points at right of the (red) line are artificially stunned by a factor  $K$ .

The choice for the factor  $K$  is much empirical. Anyway, there is no single solution since the problem boils down to finding a boundary to a diffuse cloud. There is a lot of subjectivity in determining the appropriate limit for a scatterplot presenting a gradual density on both sides.

However, the reassuring point is that the points that come to be outside of the trapezoid are not discarded from the Water Index analysis. They will simply appear as out of scale. As such, there is no danger of missing a small leak by mishandling the cold edge and interpreting a pixel of very low temperature as an outlier (i.e. left from the cold edge).

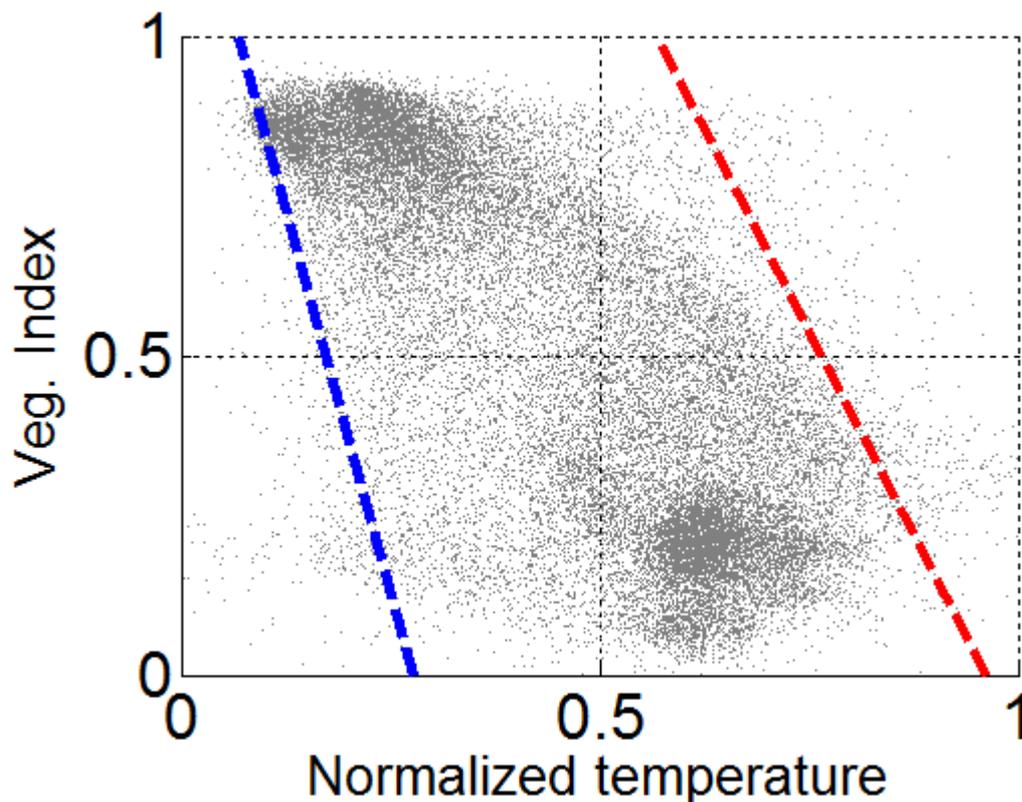


Figure 21: 2D scatterplot when considering 1 in every 50 pixels in the T and VI mosaics.

Once the cold edge and the warm edge have been set, the water index (WI) is defined for each data point as the relative distance to the dry edge (more exactly as the relative abscissa along the temperature coordinate):

$$WI_{i,j} = \frac{T_{i,j} - T_{dry}(VI_{i,j})}{T_{wet}(VI_{i,j}) - T_{dry}(VI_{i,j})}$$

By construction, the water index is essentially between 0 and 1. Only for the “outliers” left from the cold edge and right from the warm edge can it take values that are slightly higher than 1 or slightly lower than 0.

The Water Index map as inferred from Figure 21 has been reported in Figure 22. A close-up near the artificial leak location is in Figure 23. The Water Index profile along a South/North line passing through the leak is in Figure 24.

We can easily distinguish the moisture increase due to the leak. The wet area extends over 15m. The soil moisture has been measured with a TDR probe. Moisture was 38%, 13% and 4%wt at respectively 5, 10 and 20m from the leak source. The close presence of a tree south from the leak area is also seen on the WI map. The water index is even slightly higher over the tree/shadow area than over the leak-affected area.

Different modifications to the data processing have been considered:

- The warm edge of the T- $NDVI_{\rho}$  has been set manually instead of automatically (see the continuous red line in Figure 25 and the corresponding WI maps in Figure 26 and Figure 27). The WI profile is in continuous black line in Figure 24. The warm edge was manually set so as to exclude a larger part of the data on right of the green line in Figure 16. Practically, in the operational code, the top and bottom nodes of the cold/warm edges could be set through the GUI by specifying their temperature coordinate, by clicking on the right place in the plot or with the help of sliders.
- The index  $OSAVI_{\rho}$  has been used instead of  $NDVI_{\rho}$ . The 2D scatterplot is somewhat crushed (see Figure 28), however, when applying the automatic edge determination, the resulting WI is not much affected (the WI maps are in Figure 29 and Figure 30). The WI profile along the leak direction is in green dashed line in Figure 24.
- The index  $OSAVI_{\rho}$  has been used instead of  $NDVI_{\rho}$  and the warm edge has been manually modified. The WI profile along the leak direction is in green continuous line in Figure 24.
- For the T- $NDVI_{\rho}$  scatter, broken lines have been used for the cold/warm edges (see the scatterplot in Figure 31 and the WI maps in Figure 32, Figure 33). The WI profile along the leak direction is in black dotted line in Figure 24.

The latter method has been proposed in Krapez, 2011 [2]. It consists in the following procedure:

- Subdivide the VI space in  $N$  intervals (say 15). The central value of the VI in the  $I$ -th interval is  $VI_I$ ,  $I = 1, N$

- In each interval, calculate the x and 100%-x percentile in (relative) temperature; typically  $x=1\%$ ; the two temperature values are  $T_i^x$  and  $T_i^{100-x}$
- The cold edge is obtained by linear interpolation between two successive nodes ( $VI_i, T_i^x$ )
- The warm edge is obtained by linear interpolation between two successive nodes ( $VI_i, T_i^{100-x}$ )
- Remark: linear interpolation is the simplest option; alternatively, spline interpolation can be used to obtain smooth boundaries.

The result is illustrated in Figure 31. Of course, the new cold/warm boundaries better fit to the scatterplot than when straight lines are used. The choice between the two options regarding the cold/warm edges (straight lines or broken/curved lines) is not obvious. At first sight one would be tempted to choose the broken/curved boundary option. As a matter of fact, since the WI scale is defined in such a way that in a given VI range, WI evolves linearly from the highest to the lowest temperature observed in this VI range, the best is to select these high and low temperature level locally, i.e. with respect to a narrow VI interval. However, a problem may arise in the case where the 2D scatterplot is not fully representative of the temperature/moisture/vegetation diversity in the region under consideration. There may be holes in the scatter graph. The situation is actually critical when a hole is at one boundary (left or right). Obviously the data fitting will try to follow the missing area and the resulting edge will present a cavity. As a consequence, the WI calculated for the corresponding VI values will be biased. In such a situation, the use of straight lines for the edge fitting makes up for the lack of data.

As we can see by comparing Figure 22, Figure 23 on one side and Figure 32, Figure 33 on the other side, the WI maps are not affected very much by the type of lines used for the cold/warm edges. There is a global shift, but no appearance or disappearance of any detailed contrast area. In addition, when comparing the WI profiles along the leak line in Figure 24, we can again state that, in the present case, the choice of the type of edges in the triangle/trapezoid representation is not critical.

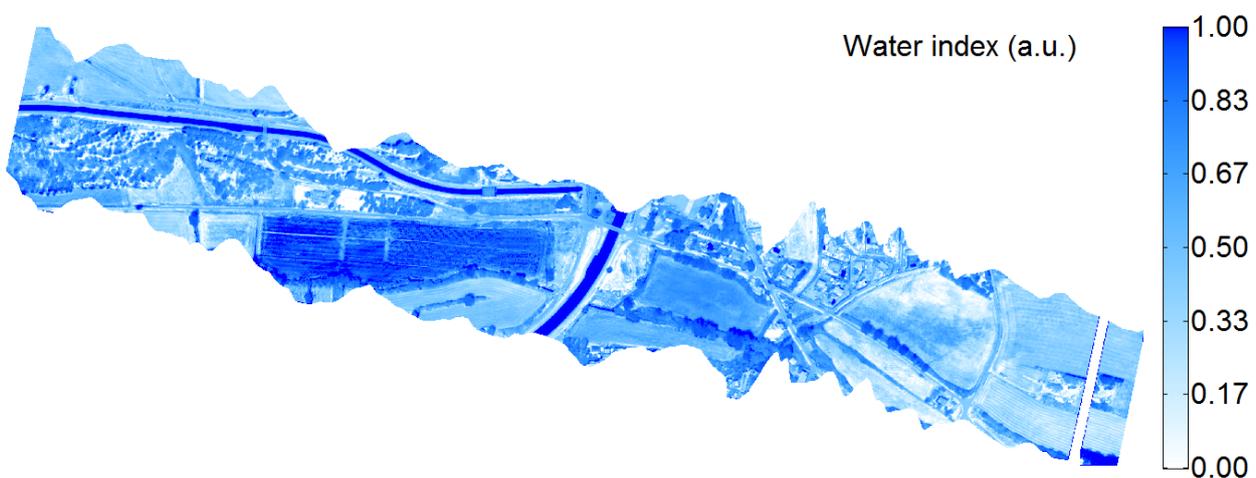


Figure 22: Water index map inferred from Figure 21 ( $NDVI_\rho$ ; automatic cold/warm edges).

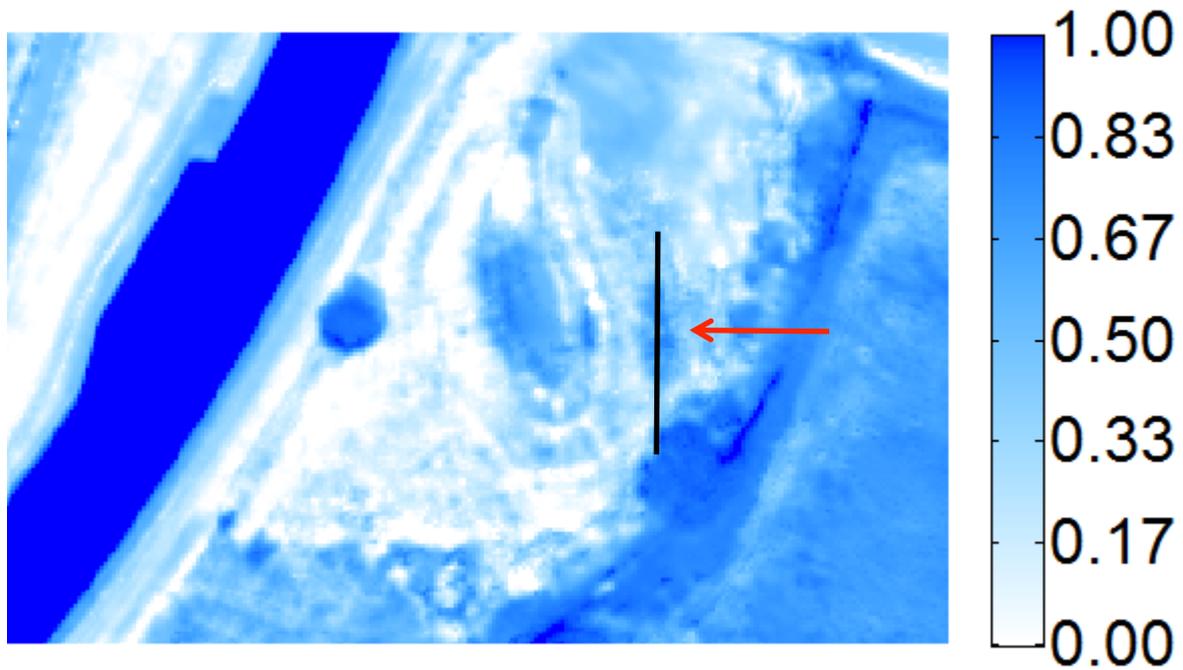


Figure 23: Close-up of the Water Index map in Figure 22 centered in the area containing the artificial leak. The leak is indicated by a red arrow (dark vertically elongated area). A vertical black line passing through the leak area has been added. The water index profile along this line is indicated in Figure 24.

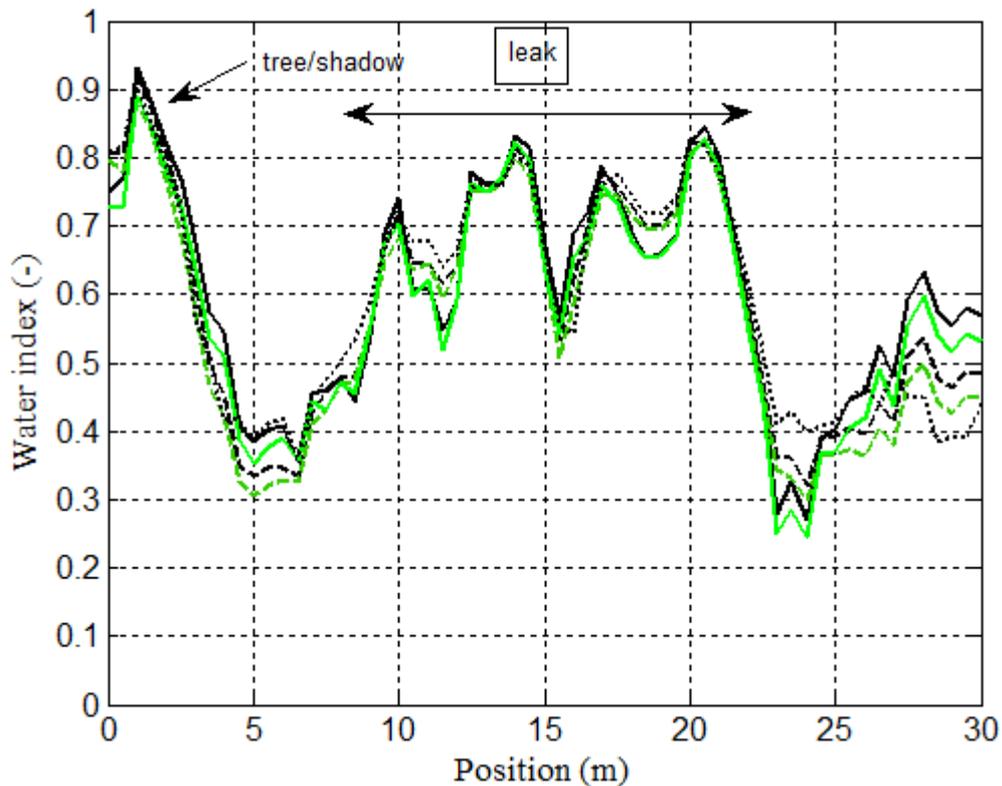


Figure 24: Water index profile along the black line in Figure 23. The consequence of the leak is a significant increase of the soil moisture along a distance of about 15 m. The soil moisture measurements at ground gave 38%, 13% and 4%wt at respectively 5, 10 and 20m from the leak source. The high WI

value at the left of the plot (resp. at the bottom end of the black line in Figure 23) is actually the signature of a tree (vegetation canopy+shadow).

**In black:** The considered VI is  $NDVI_{\rho}$ .

**In green:** The considered VI is  $OSAVI_{\rho}$ .

**Dashed line:** Automatic determination of the cold/warm edges (straight lines).

**Continuous line:** Manual determination of the warm edge (straight line).

**Dotted line:** Cold/warm edges with broken/curved lines.

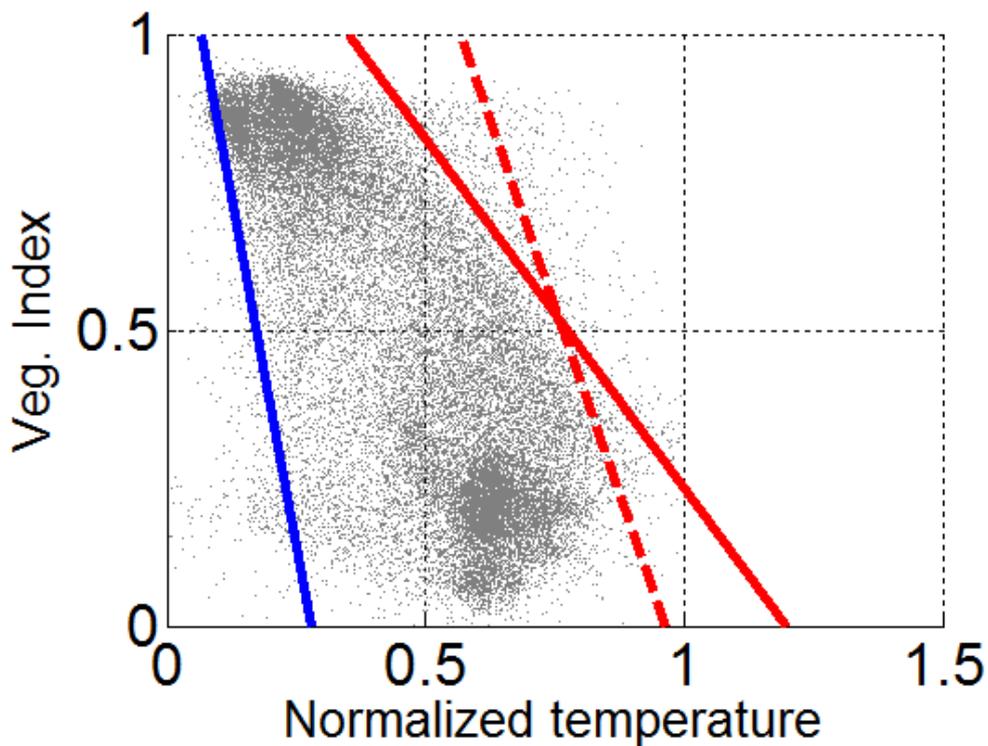


Figure 25: 2D scatterplot when considering 1 in every 50 pixels in the T and VI mosaics. Dashed lines: edges determined automatically. Validated edges and possibly manually transformed edges are in continuous lines.

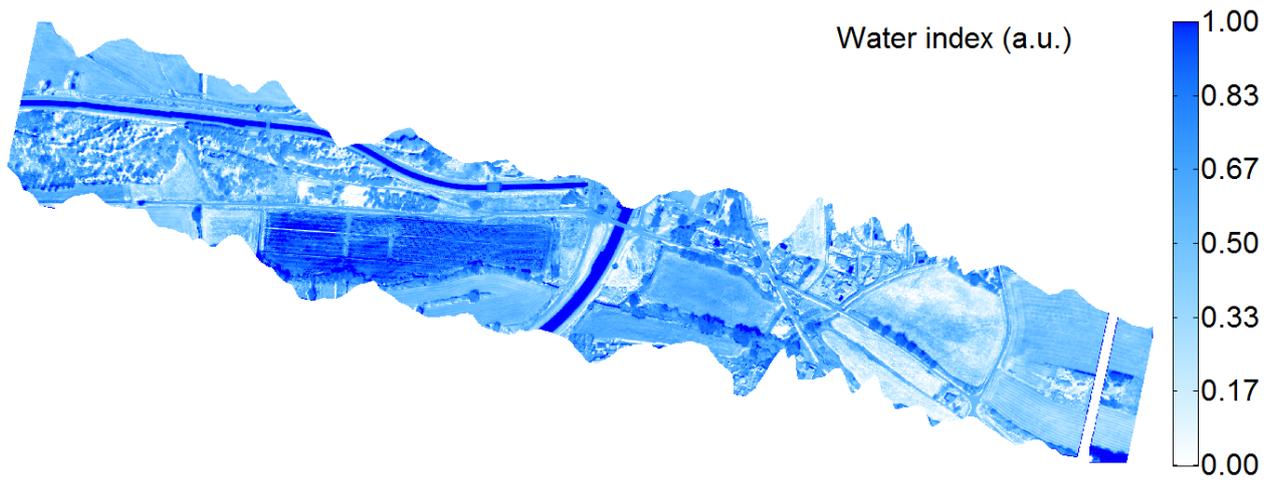


Figure 26: Water index map inferred from Figure 25 ( $NDVI_{\rho}$ ; manually selected warm edge- straight lines).

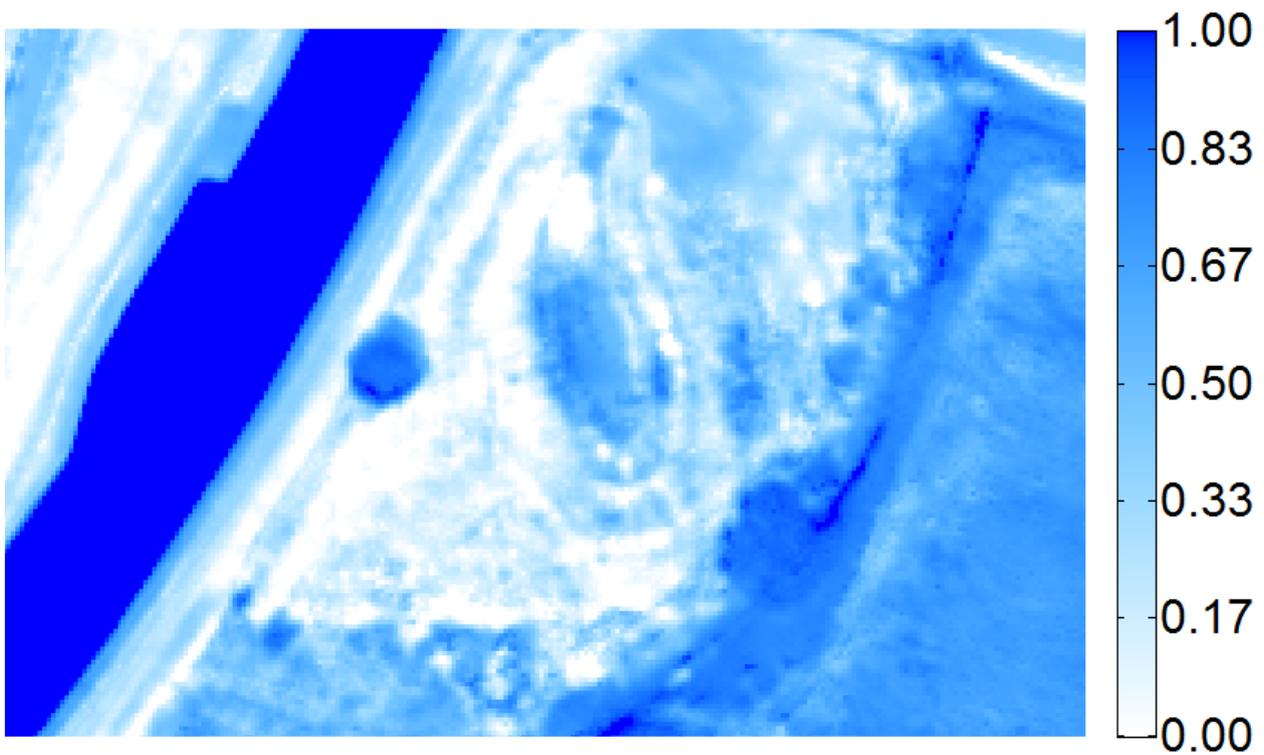


Figure 27: Close-up of the Water Index map in Figure 26 ( $NDVI_{\rho}$ ; manually selected warm edge - straight lines).

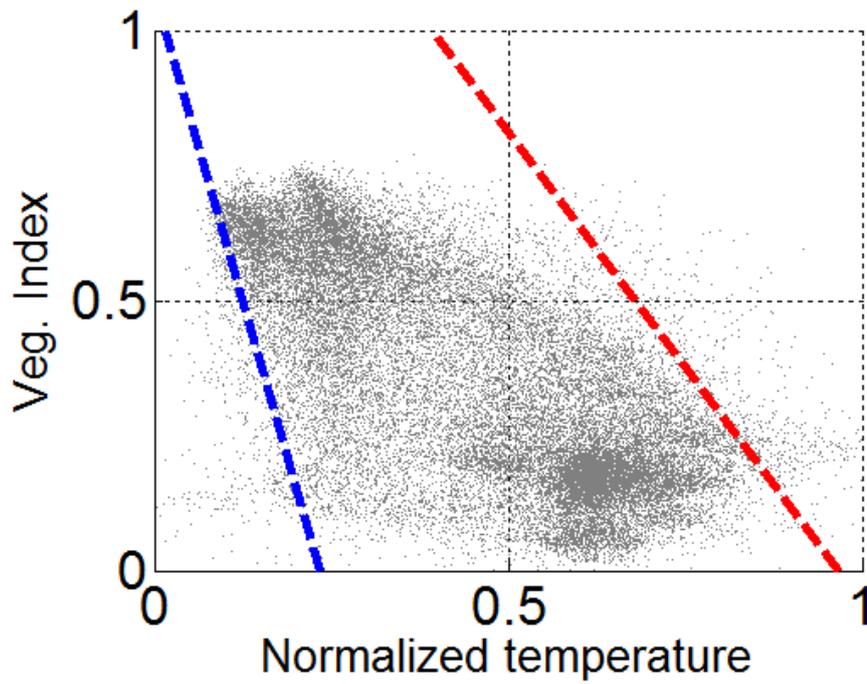


Figure 28: 2D scatterplot when considering the index  $OSAVI_{\rho}$ . The dashed edges have been determined automatically.

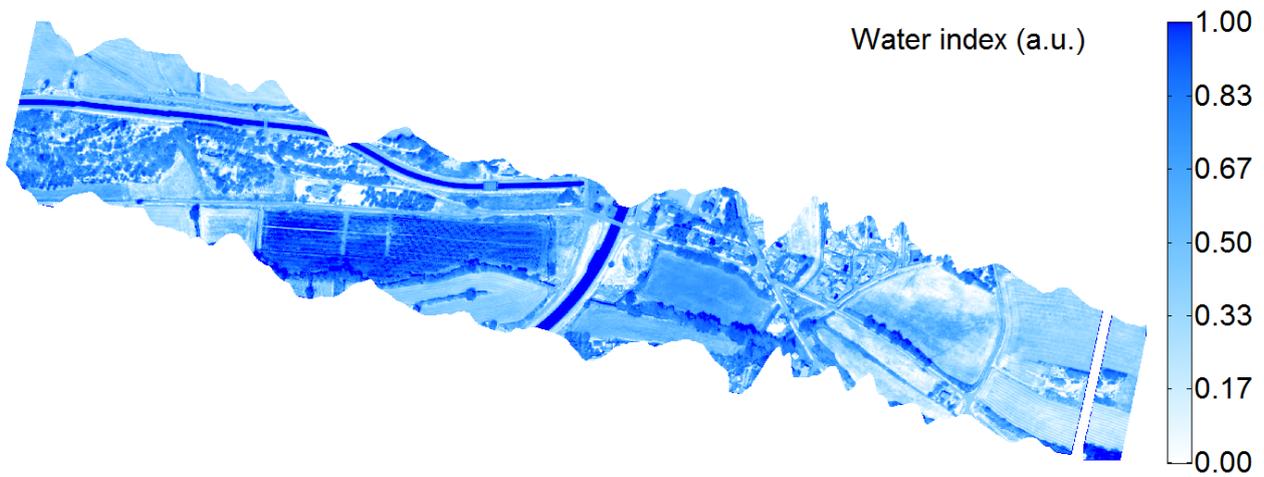


Figure 29: Water index map inferred from Figure 28 ( $OSAVI_{\rho}$ ; automatic cold/warm edges - straight lines).

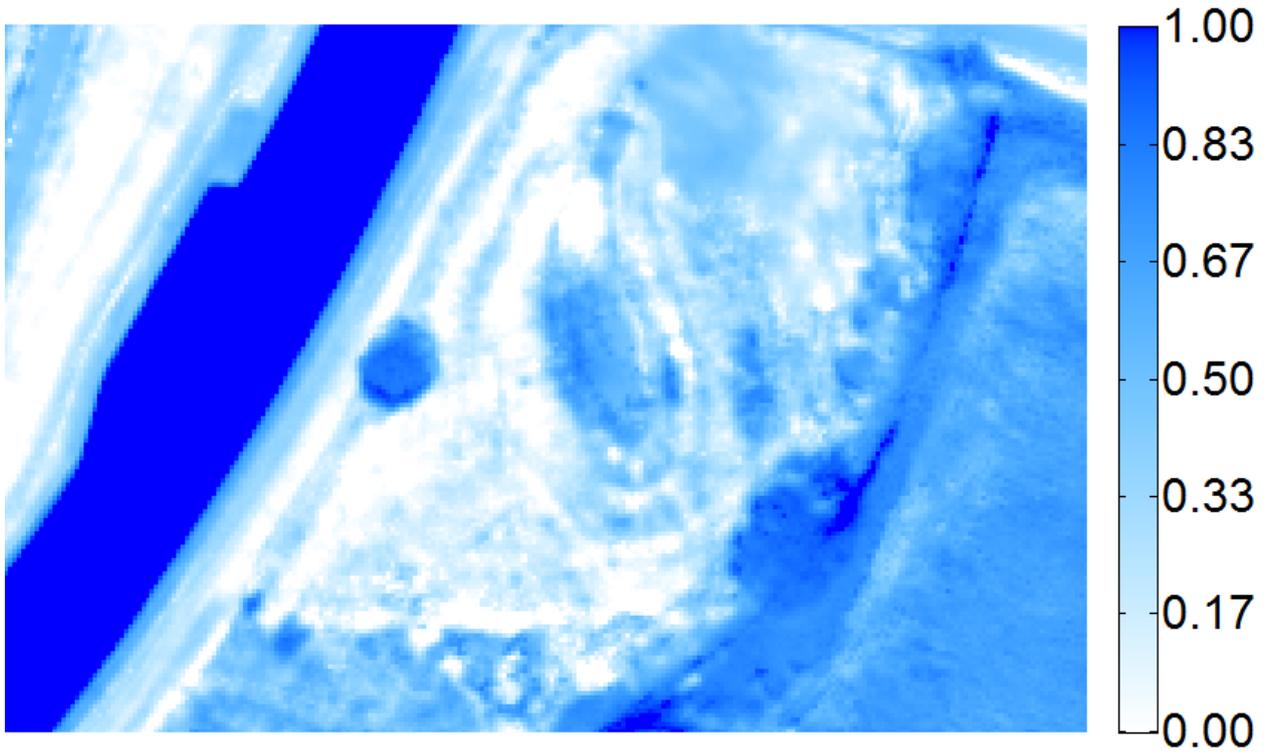


Figure 30: Close-up of the Water Index map in Figure 29 ( $OSAVI_p$ ; automatic cold/warm edges – straight lines).

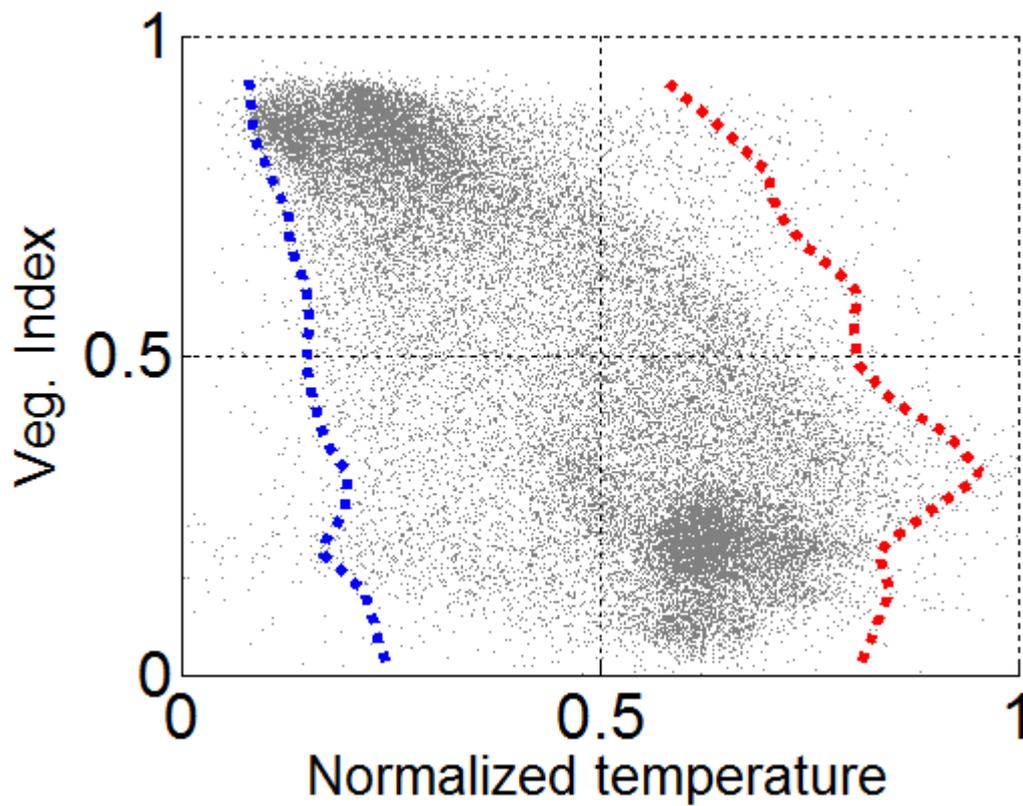


Figure 31: 2D scatterplot when considering the index  $NDVI_p$ ; cold/warm edges with broken/curved lines.

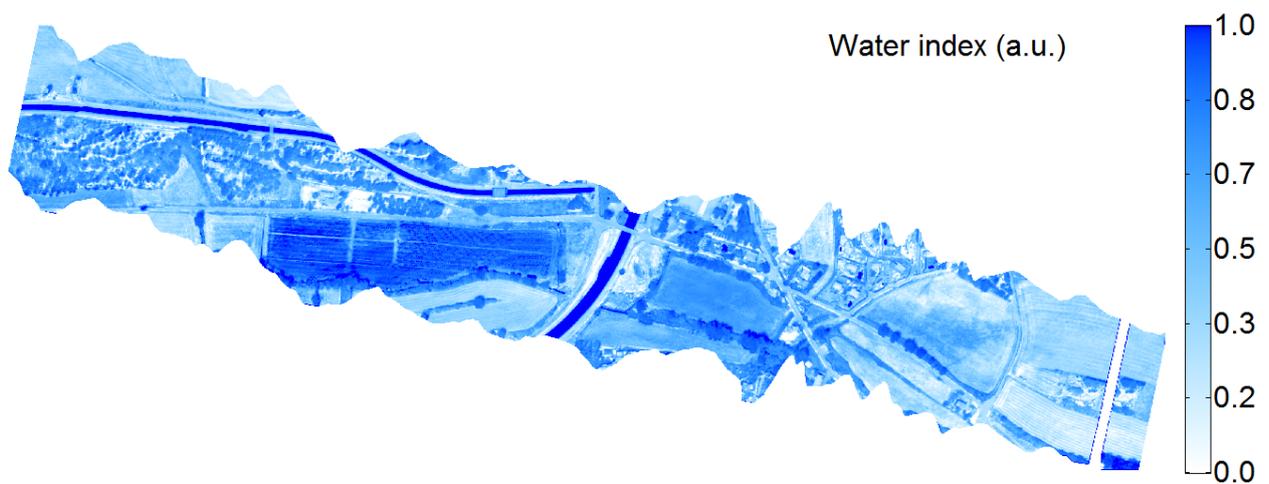


Figure 32: Water index map inferred from Figure 31 ( $NDVI_{\rho}$ ; cold/warm edges with broken/curved lines).

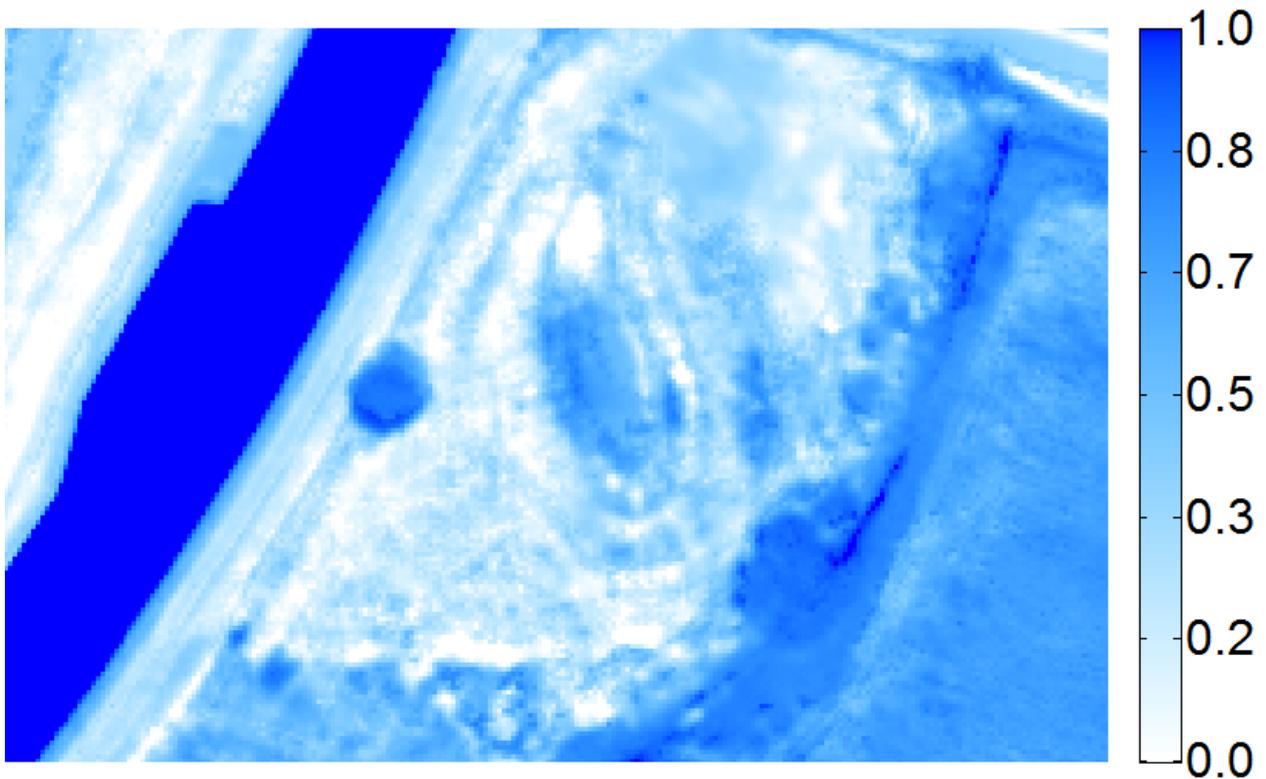


Figure 33: Close-up of the Water Index map in Figure 32 ( $NDVI_{\rho}$ ; cold/warm edges with broken/curved lines).

In Figure 34 the pixels in the vicinity of the leak area showing a significant contrast with the surroundings have been delineated with a blue line. Those pixels are highlighted in the 2D scatterplot in Figure 35 with blue large dots. Two important things can be noticed:

- The blue dots are quite scattered, both in temperature and in VI.

- The cloud of blue dots representative of the leak area is not very close to the cold/wet edge. On the average, the Water Index is about 0.65 and at most it reaches about 0.85.

This last observation makes that the leak would not be easy to detect if we didn't know where it is. A significant portion of the 2D scatterplot is closer to the cold/wet edge than the set of blue dots. To illustrate this problem, we reported in red in the WI map in Figure 36 those pixels showing simultaneously  $WI > 0.85$  and  $VI < 0.7$ . By this way we highlight all areas that present a stronger signature than do the artificial leak. In the same time, we excluded too high VI values since highly vegetated areas present an even more difficult problem. The canals are obviously revealed as well as a large irrigated field in the central part of the image. Some artefacts are revealed in the peri-urban zone (roofs, metallic sheds, swimming pools and a strong and wide signature from a seemingly bare soil and very cold area). The other false alarms come essentially from the edges of groves and trees and from trails and roads. In the former case, the high WI value can originate from the shadows (shadows are unavoidably colder areas which are here interpreted as areas with high evaporation/transpiration, hence surfaces presenting a high moisture, which is not necessarily the case). In all cases however, the spurious signal can come from a residual error of the image registration process.

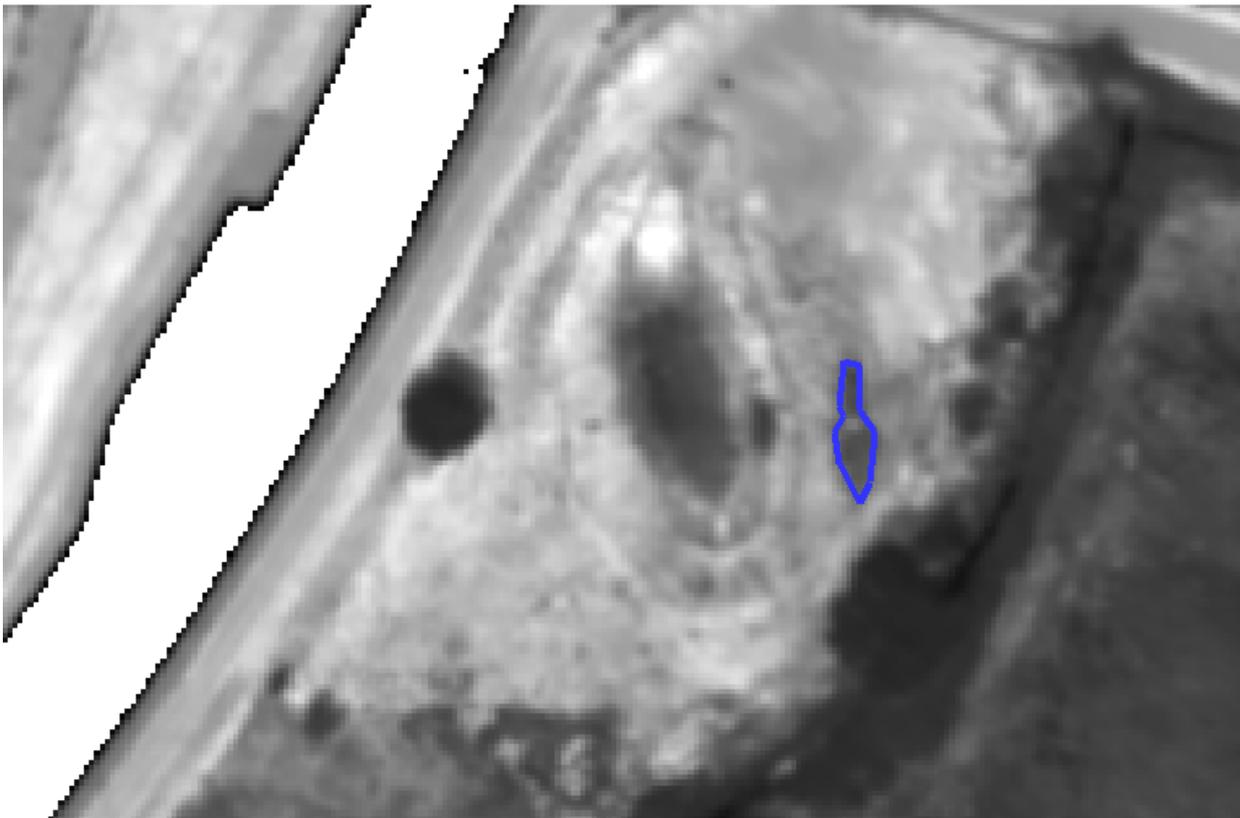


Figure 34: Close-up of the apparent temperature map. The "leak" area has been delineated with a blue line.

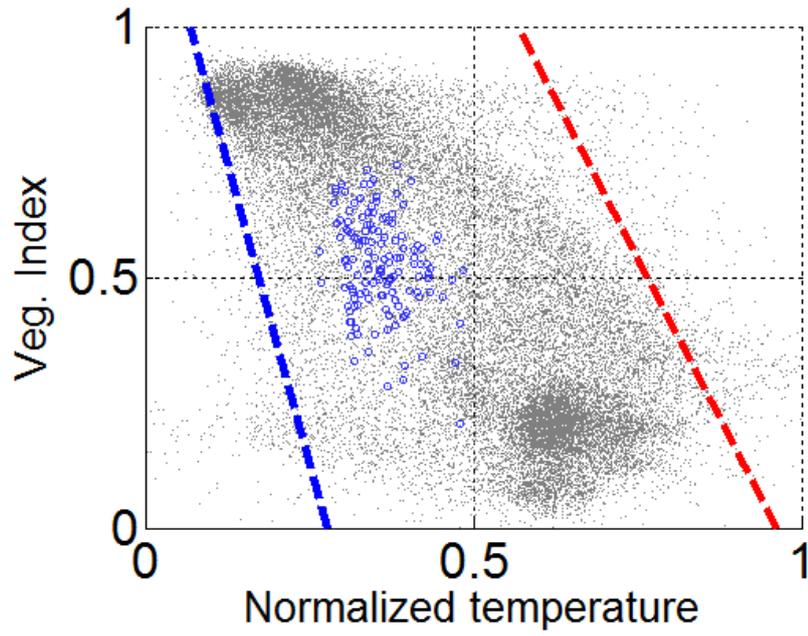


Figure 35: 2D scatterplot as in Figure 21 where the pixels from the leak area have been highlighted with blue dots.

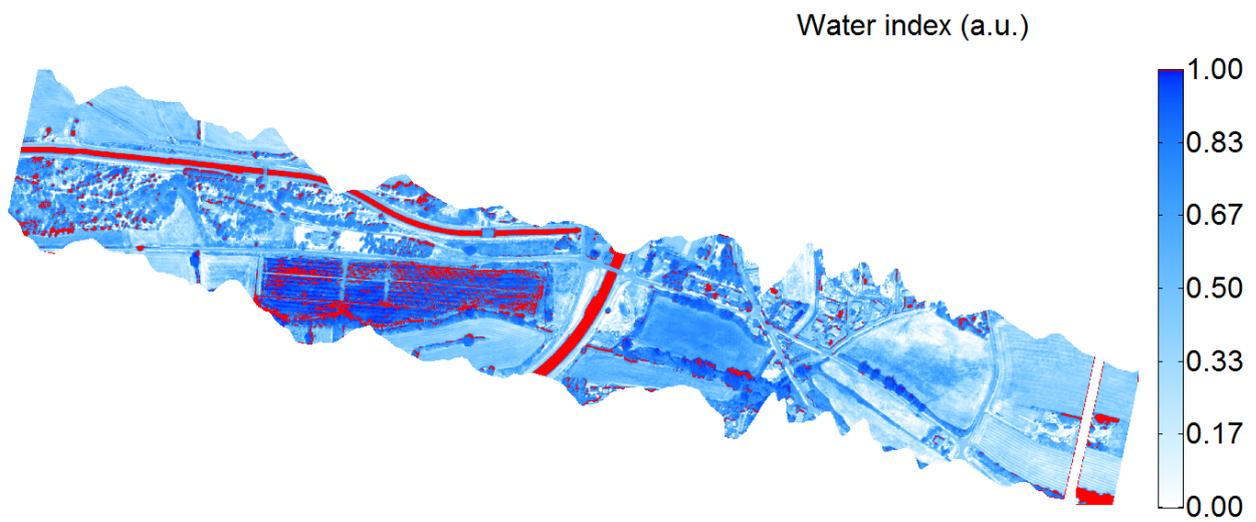


Figure 36: Water index map inferred from Figure 21 ( $NDVI_{\rho}$ ; automatic cold/warm edges). Areas with  $WI > 0.85$  and  $VI < 0.7$  have been highlighted in red.

## 5 Data processing validation

The illustrations given before are from one among the six areas that have been considered during the flight campaigns in 2017. However, moisture was not always measured at ground. If we retain only those flights during which TDR probes have been used to measure soil moisture in the vicinity of the artificial leaks, there are three sites left: Rians (as shown before), le Tholonet and Vauvenargues.

For all of them, ONERA can provide files containing:

- Pushbroom-reconstituted images of the reflectance in several bands: 0.501, 0.648, 0.662, 0.676, 0.752, 0.818, 0.858, 0.900 and 0.970  $\mu\text{m}$ .
- Mosaic of images built from the raw TIR signal as obtained with a microbolometer camera.

These data have been resampled to provide synthetic images of the same size and same resolution (0.5m). In the end, two files in BSQ format (one for the multispectral VNIR images and one for the TIR mosaic) are available for each site.

It will be possible to use the three pairs of files to validate the new code (which will have to be adapted to accept the former files since they don't correspond in size and format to those provided by the new cameras). The results will be compared against any of the results presented earlier:

- Vegetation index map.
- 2D scatterplot of (apparent) temperature and VI.
- Cold and warm edge construction.
- Water index calculation.

The validation will be done by comparing the values retrieved at a series of pixels (some of them being in the leakage area and in its vicinity, others being in various interesting locations, i.e. canals, buildings, field boundaries, etc...).

In particular, different options regarding the calculation of the scatterplot boundaries will be considered (manual and automatic settings).

## 6 Methodology for data reliability

While the performance evaluation of the WADI service is 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 will be built by analyzing indexes based on Sentinel images and further complemented with a terrain-following water pathways model. This methodology for data reliability in WADI intends to promote confidence in WADI service and is applicable after the post-processing step.

The Sentinel data 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 5 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 can be 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.

In the first step of the methodology, Sentinel 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. This algorithm can then 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 can be 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 the implementation of this methodology, which is illustrated in the following figures, 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. An initial implementation of this methodology to the EDIA site is illustrated below and can be found in [9].

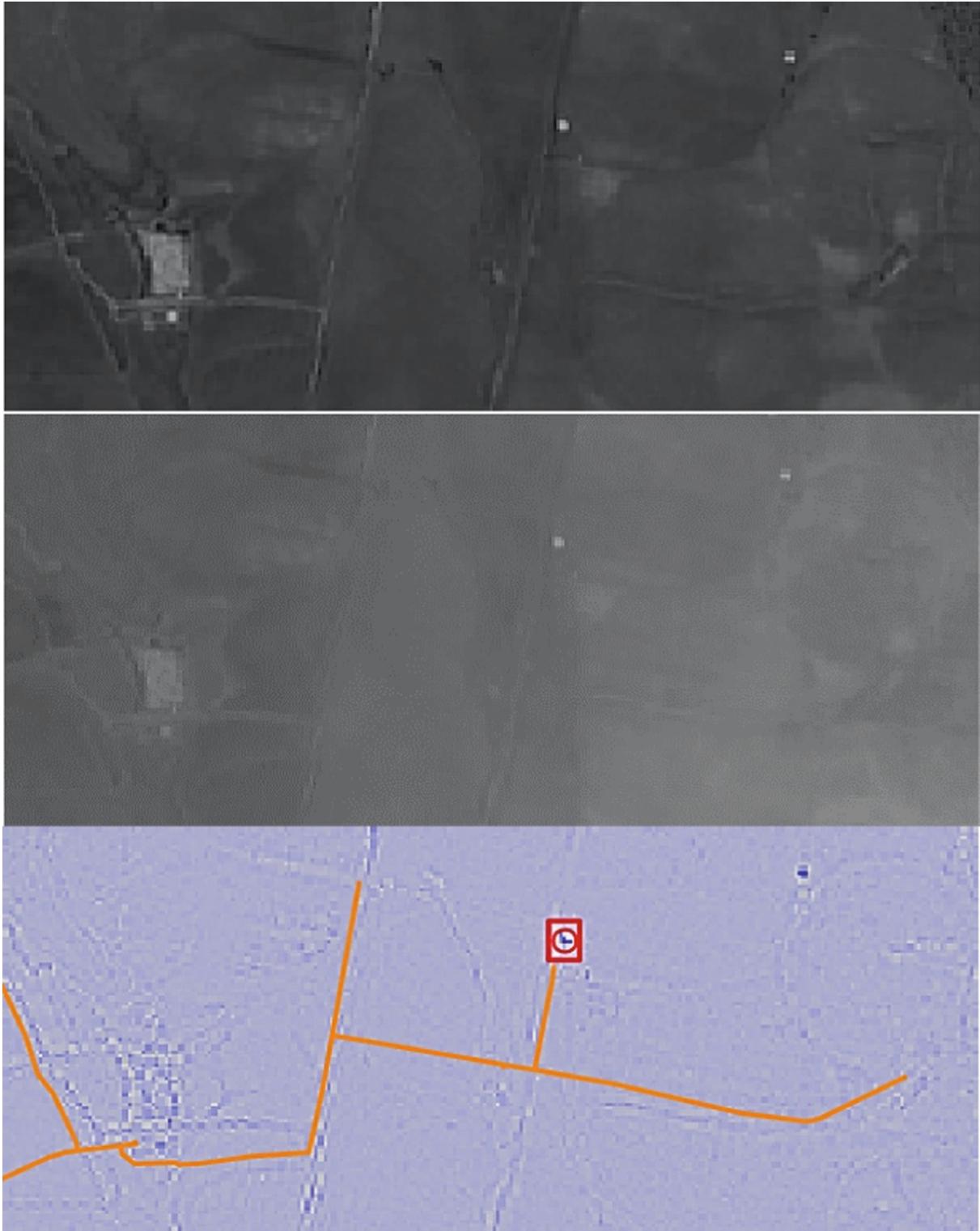


Figure 37: Example of leak detection: MNDWI (top), Anomaly (middle), Laplacian and possible leaks (bottom) of a small area of the AoS. The network location is shown in orange (extracted from [9]).

## 7 Data File Formats

### 7.1 Best practices for file formats

The format of the files in which data is stored plays a very important role because it determines how we can access them in the future and how we can share them with others. It is important to take into account the following considerations [10]:

1. Is the file format part of a standard?
2. Is the format open (open source) or closed (proprietary)?
3. Is there any special software required to open the file?
4. Does the format imply multiple files? (for example, shapefile).

Data producers should choose a format that can be read in the future and that is independent of changes in the applications. Whenever possible, a non-proprietary format should be chosen. When the conversion to an open format results in the loss of some information the data should be saved in both formats.

If it is necessary to save the data in a proprietary format, it is advisable to create a `readme.txt` that will be used to store the name and version of the software that generated the data, as well as the name of the company and authors that developed the software. Finally, it is advisable to keep data files updated with new versions of the software, and migrate the data to its latest versions if possible, to promote its longer use in the future.

### 7.2 Guidelines for choosing formats

When selecting file formats for archiving, the formats should ideally be [11]:

- Non-proprietary.
- Unencrypted.
- Uncompressed.
- Commonly used by the research community.
- Adherent to an open, documented standard.
- Interoperable among diverse platforms and applications.
- Fully published and available royalty-free.
- Fully and independently implementable by multiple software providers on multiple platforms without any intellectual property restrictions for necessary technology.

In the following section a review of standard file formats is presented, aiming at establishing the desired formats for all types of data produced in WADI that comply with the INSPIRE directive and other requirements.

Based on this review, WADI partners can select the formats that better fit their field of expertise while promoting data reuse and interoperability.

### 7.3 File format standards

The data format is directly related to the process of information generation and use through a particular type of software. The selected format and software depend on the work area of each organization or the scientific area of the project where the data is being produced.

Usually data formats are chosen because they are the most commonly used in the subject field. However, this is not always the most recommended choice. This format may be difficult to convert to other formats or to preserve the data in the future. Indeed, it may be impossible in the future to find a software that can open this data.

A good example of this problem is spreadsheet Lotus 1-2-3 which was in common use some years ago. However, the last stable release is from 16 years ago and its official sale ended in 2013. Data in this type of formats may be rendered useless soon. Therefore, it is recommended for data producers in WADI to use standard formats to save data to promote data share and data usability in the future.

In the table below, the standard formats are presented, defined for the different data types. To facilitate data management, the most commonly used alternative formats are also listed. If a non-standard format is chosen because it is considered easier to handle by the users, it is recommended to also record data in a standard format for future use [12].

	Standard	Common use
<b>Containers</b>	ZIP file (.zip) ISO/IEC 21320-1:2015 Tape Archive (.tar) ISO/IEC/IEEE 9945:2009	GNU zip (.gzip)
<b>Tabular Data</b>	Comma separated values file (.csv) ISO/IEC 8859-1:1998 OpenDocument Spreadsheet (.ods) ISO/IEC 26300-1:2015 eXtensible Markup Language (.xml) extended from Standard Generalized Markup Language ISO 8879:1986	Microsoft Office Excel (.xlsx)
<b>Textual Data</b>	Plain text format (.txt) ISO/IEC 8859-1:1998 eXtensible Markup Language (.xml) extended from Standard Generalized Markup Language ISO 8879:1986 OpenDocument Text (.odt) ISO/IEC 26300-1:2015 PDF/A ISO 19005-1:2005	PDF Microsoft Office Word (.docx)
<b>Geospatial Data</b>	Geographic Markup Language (.gml) ISO 19136:2007 Network Common Data Form (.nc) OGC standard Keyhole Markup Language (.kml) OGC standard	ESRI Shapefile (.shp, .shx, .dbf) Geo referenced TIFF (.tif)
<b>Images</b>	TIFF (.tif) ISO 12639:2004 JPEG (.jpg) ISO/IEC 10918-1 PNG (.png) ISO/IEC 15948:2004 JPEG 2000 (.jp2) ISO/IEC 15444-1:2016	GIF (.gif)
<b>Video</b>	MPEG4 (.mp4) ISO/IEC 14496-14:2003 Motion JPEG 2000 (.mjp2) ISO/IEC 15444-3:2007	Audio Video Interleave (.avi) Windows Media Video (.wmv) Apple Quick Time Movie (.mov)
<b>Audio</b>	MPEG audio layer 3 (.mp3) ISO/IEC 11172-3:1993 and ISO/IEC 13818-3:1998	Audio Interchange Format (.aif) Waveform Audio File Format (.wav)
<b>Computer Aided Design (CAD)</b>	Extensible 3D (.x3D, .x3dv) ISO/IEC 19775-1:2013 Standard for the Exchange of Product model data (.step) ISO 10303-21:2016	Drawing Exchange Format (.dxf)
<b>Web archive</b>	WARC ISO 28500:2017	

Table 1: Standard and common used file formats [13][14].

## 8 Metadata

### 8.1 Introduction

The shortest definition of metadata is “data that describes other data”. It consists of a set of information that describes and gives information about other data, providing the necessary information for this data to be understandable by humans and computers. It is typically encoded using XML documents. The concept of metadata is not new: it's a term that emerged in the late 1960's. It has been used in libraries for classifying and categorizing materials based on title, author and subject so that they can more easily be located on shelves from among many thousands of others [15].

Nowadays, the amount of information that is produced and stored is huge. It is thus necessary to look for a way to describe this information in order to be easier to find it both by who produces and by all those with whom we want to share our information. One of the most important uses for metadata is to locate a resource [16].

Nowadays all the information has some kind of associated metadata even if we do not realize it. For example, when we take a photo, the size and shape of the image, the number of colors and the date are automatic metadata that allows us to better describe this photo and, in the future, to find and access it more easily.

The next section describes the current metadata standards and the suggestions for metadata editing in the WADI project, given the type and nature of the information being produced in this project.

### 8.2 Metadata Standards

Metadata is the documentation that accompanies the research data, making it discoverable and usable over time. Metadata standards exist to provide standardized descriptions of data that can easily be searched for and used for other purposes than the original one. Through the use of metadata standards, computer software is able to find and combine data from several sources, based on their metadata [17].

Metadata elements are typically grouped into sets designed for a specific purpose, e.g., for a specific domain or a particular type of information resource. They are called metadata schemes. For every element, the name and the semantics (the meaning of the element) are specified. Content rules (how content must be formulated), representation rules (e.g., capitalization rules), and permitted element values (e.g., from a controlled vocabulary) can be specified optionally. Some schemes also specify in which syntax the elements must be encoded, in contrast to syntax independent schemes. Many current schemes use Standard Generalized Markup Language (SGML) or XML to specify their syntax.

Metadata schemes that are developed and maintained by standard organizations (such as ISO) or organizations that have taken on such responsibility (such as the Dublin Core Metadata Initiative) are called metadata standards.

Many different metadata schemes are being developed as standards across disciplines, such as library science, education, archiving, e-commerce, and arts [18].

### 8.3 WADI metadata standards

WADI will produce and reuse a variety of data types such as text documents, worksheets, databases, aerial photographs or maps (see the WADI DMP report, Oliveira et al., 2017). Metadata will be produced for all data using standards. Herein, three options for metadata generation are proposed, based on different standards and tools.

#### 8.3.1 ZENODO - Research data repository

According to the Data Management Plan (DMP, Oliveira et al., 2017), ZENODO will be used as the project's open data repository. In ZENODO, metadata use a formal, accessible, shared and broadly applicable language for knowledge representation. It uses JSON Schema as internal representation of metadata. It also offers the capacity to export these metadata to other popular formats such as Dublin Core or MARC XML [19]. The core of the MARC XML framework is a simple XML schema which contains MARC data which is a structured record used to catalog bibliographic information [20]. This base schema output can be used where full MARC records are needed or act as a "bus" to enable MARC data records to go through further transformations such as to Dublin Core and/or processes such as validation. MARC standards are a set of fields for the description of items catalogued by libraries, such as books [21].

This type of standard is good for describing publications, presentations, images or videos but it has many limitations when we want to describe more complex data such as geographic data. For this data, in addition to the indication of title, author, subject or date, the metadata should indicate the geographic location, the coordinate reference system and the geographical transformation used to represent spatial information. The figure below illustrates the information fields provided by ZENODO to document the data.

The image shows a screenshot of the ZENODO upload interface. It is divided into two main sections: 'Communities' and 'Upload type'. The 'Communities' section has a search bar with the placeholder text 'Start typing a community name...' and a dropdown menu currently set to 'recommended'. The 'Upload type' section is labeled 'required' and contains a grid of radio button options for selecting the data type. The options are: Publication, Poster, Presentation, Dataset, Image, Video/Audio, Software, Lesson, and Other. The 'Dataset' option is selected, indicated by a filled radio button.

Figure 38: ZENODO upload data type.

The majority of the information generated in WADI will be datasets (from remote sensing, local data or performance matrix indicators, making the dataset type the most important. Although there is a Dataset data type, the fields to be filled in the ZENODO upload form do not differ much from those used for the other data types, providing only poor, free-text fields. Figure 39 illustrates the available documentation fields for a Dataset.

The screenshot shows the 'Basic information' section of the ZENODO upload form. It includes the following fields and options:

- Digital Object Identifier:** A text input field with the placeholder 'e.g. 10.1234/foo.bar'. Below it is a 'Reserve DOI' button and explanatory text: 'Optional. Did your publisher already assign a DOI to your upload? If not, leave the field empty and we will register a new DOI for you. A DOI allows others to easily and unambiguously cite your upload. Please note that it is NOT possible to edit a Zenodo DOI once it has been registered by us, while it is always possible to edit a custom DOI.'
- Publication date:** A date input field containing '2018-03-19'. Below it is explanatory text: 'Required. Format: YYYY-MM-DD. In case your upload was already published elsewhere, please use the date of first publication.'
- Title:** A text input field containing 'Hourly forecast of the sea surface height for the Tagus estuary'. Below it is the label 'Required'.
- Authors:** A section with three input fields: 'LNEC', 'Affiliation', and 'ORCID (e.g.: 0000-0001-8285-2489)'. Below the ORCID field is the label 'Optional'. There is a '+ Add another author' link below the fields.
- Description:** A rich text editor with a toolbar containing icons for bold, italic, strikethrough, subscript, superscript, link, unlink, list, ordered list, indent, outdent, quote, and image. Below the toolbar is the text: 'Hourly forecast maps of the sea surface height in meters for the Tagus estuary region. The maps are generated once a day with forecasts of the following 48 hours (2 days).'

Figure 39: ZENODO Basic information example for Dataset data type.

If we want to describe any geographical information in ZENODO, the description element, which is a free text field, is the only option. Thus it will be very difficult to extract the relevant information in an automatic way so it can be used in a geographic portal or application software. It will be also very difficult to exchange or overlay information with other geographic data producers.

### 8.3.2 Dublin Core Metadata Initiative

The Dublin Core Schema is a small set of vocabulary terms that can be used to describe digital resources such as video, images, web pages as well as physical resources such as books or CDs (<http://www.dublincore.org/documents/dcmi-type-vocabulary/>). The original Dublin Core Schema consists of 15 metadata elements. This original set has been progressively been enlarged to extend or refine its scope. This new set of elements (Figure 40) is called Qualified Dublin Core. Metadata Implementing Rules and is based on ISO 15836 [22].

- abstract
- accessRights
- accrualMethod
- accrualPeriodicity
- accrualPolicy
- alternative
- audience
- available
- bibliographicCitation
- conformsTo
- contributor
- coverage
- created
- creator
- date
- dateAccepted
- dateCopyrighted
- dateSubmitted
- description
- educationLevel
- extent
- format
- hasFormat
- hasPart
- hasVersion
- identifier
- instructionalMethod
- isFormatOf
- isPartOf
- isReferencedBy
- isReplacedBy
- isRequiredBy
- issued
- isVersionOf
- language
- license
- mediator
- medium
- modified
- provenance
- publisher
- references
- relation
- replaces
- requires
- rights
- rightsHolder
- source
- spatial
- subject
- tableOfContents
- temporal
- title
- type
- valid

Figure 40: Qualified Dublin Core Metadata Element Set.

Similarly to ZENODO, Dublin Core Metadata has very generic elements that allow the user to document different data types. However, it is limited at the level of geographic information. One of the possible elements that rises from the original set is data coverage. It represents the spatial or temporal information of the resource, its spatial applicability, or the jurisdiction under which the resource is relevant. Spatial information may be a place's name or a location specified by its geographic coordinates. Temporal information may be a period's name, date, or date range. A jurisdiction may be a designated administrative entity or a geographic place to which the resource applies [23].

Examples of use of the coverage element are:

- `<meta name = "DC.Coverage" content = "Portuguese discoveries era; 1415-1543">`
- `<meta name = "DC.Coverage" content = "Lisboa, Portugal; Lat: 38 43 N Long: 9 7 W">`

As seen in the examples, the coverage element is a free text field with different meanings that goes from spatial to temporal coverages. This is enough if we just want to catalog data. However, if spatial queries are needed, it will be very hard to analyze a text field and determine a spatial coverage.

### 8.3.3 INSPIRE – Infrastructure for spatial information in Europe

Geographical data, its standardization and its sharing across member states has been an EC concern. To address this concern and promote a spatial data infrastructure for Europe, the INSPIRE Directive was created. According to [24] the INSPIRE Directive aims to create a

European Union spatial data infrastructure for the purposes of EU environmental policies and policies or activities which may have an impact on the environment. This European spatial data infrastructure will enable the sharing of environmental spatial information among public sector organizations, facilitate public access to spatial information across Europe and assist in policy-making across boundaries.

The INSPIRE Directive is a legal act (Directive 2007/2/EC) of the Council of the European Union that all state members must implement (Figure 41). The INSPIRE Directive sets the objectives to be achieved and asks the Member States to transpose it to their own national legislation establishing their spatial data infrastructure. This mechanism of European plus national legislation allows each country to define its own way to achieve the objectives agreed upon, taking into account its own institutional characteristics and history of data infrastructure development [25].

### INSPIRE in your Country

In this section you can find information about INSPIRE in the countries implementing INSPIRE.

Click here to [view the Map Mode](#)



#### Europe



[Summary Report on Status of Implementation of the INSPIRE Directive in EU](#)

#### Member States

- |                                |                            |                                |
|--------------------------------|----------------------------|--------------------------------|
| <a href="#">Austria</a>        | <a href="#">Germany</a>    | <a href="#">Netherlands</a>    |
| <a href="#">Belgium</a>        | <a href="#">Greece</a>     | <a href="#">Poland</a>         |
| <a href="#">Bulgaria</a>       | <a href="#">Hungary</a>    | <a href="#">Portugal</a>       |
| <a href="#">Croatia</a>        | <a href="#">Ireland</a>    | <a href="#">Romania</a>        |
| <a href="#">Cyprus</a>         | <a href="#">Italy</a>      | <a href="#">Slovakia</a>       |
| <a href="#">Czech Republic</a> | <a href="#">Latvia</a>     | <a href="#">Slovenia</a>       |
| <a href="#">Denmark</a>        | <a href="#">Lithuania</a>  | <a href="#">Spain</a>          |
| <a href="#">Estonia</a>        | <a href="#">Luxembourg</a> | <a href="#">Sweden</a>         |
| <a href="#">Finland</a>        | <a href="#">Malta</a>      | <a href="#">United Kingdom</a> |
| <a href="#">France</a>         |                            |                                |

Figure 41: INSPIRE in your country.

INSPIRE is based on the infrastructures for spatial information established and operated by the Member States of the European Union. The Directive addresses 34 spatial data themes for use in environmental applications. The Directive started its application on the 15<sup>th</sup> of May 2007 and is being implemented in various stages, with full implementation required by 2021 [24].

Metadata Implementing Rules for INSPIRE are based on ISO 19115, which defines the schema required for describing geographic information and services by means of

metadata [26], and ISO 19119, which defines requirements for how platform services shall be created. These rules aim at allowing for one service to be specified independently of the distributed computing platforms [27].

The INSPIRE Implementing Rules on interoperability of spatial data sets and services and Data Specifications specify common data models, code lists, map layers and additional metadata on the interoperability to be used when exchanging spatial datasets. Inspire-compliant datasets are those which fit under one or more of the 34 spatial data themes set out in this Directive. Interoperability in INSPIRE means the possibility to combine spatial data and services from different sources across the European Community in a consistent way without involving large efforts of humans or machines [28].

INSPIRE metadata elements provides information about identification, extent, quality, spatial and temporal aspects, content, spatial reference, portrayal, distribution and other properties of digital geographic data and services [26]. It has some core elements like the title, author, abstract or date that are similar to the ones used with ZENODO or Dublin Core metadata elements but it also has other elements that are very important to describe the geographic data in a very precise way.

Although ISO 19115 is applicable to digital data and services, its principles can be extended to many other types of resources such as maps, charts, and text documents as well as non-geographic data [26].

## 8.4 Metadata Editors

Metadata comes in many different varieties, each with their own structures and protocols [29]. As noted above, it is typically encoded using XML documents. Although we can work on metadata with any text editor, it is a complicated task to write an XML file using a simple text editor.

Therefore, metadata files are typically generate using metadata editors. A metadata editor is a tool that facilitates the documentation of resources, focusing on the description of information resources and producing automatically an xml file that complies with standards [30].

As there are many metadata structures, it would be almost impossible to create a single editor that would satisfy all the requirements of all metadata schemas. A number of proprietary and freeware metadata editing programs have thus emerged. There are editors for Windows, Linux or Mac systems. These editors can be desktop versions or online applications accessible using a web browser.

Geographic information systems like QGIS or ArcGIS have their own metadata editor compliant with several metadata standards.

Herein, we focus on 4 editors, selected by the type of metadata files produced (taking into account the type of data in WADI) and for being freely available.

### 8.4.1 ZENODO metadata editor

When uploading a dataset to ZENODO's repositories, the user needs to fill in some information to document this dataset. This task leads to the creation of metadata file that is stored together with the data.

When uploading a new resource, we have to fill an online form with several text fields describing the resource we are about to publish. After filling all required fields and publishing a resource we can then export its metadata to one of the existing formats such as BibTeX, DataCite or Dublin Core. This file is typically created in Dublin Core format for interoperability reasons.

The screenshot shows the ZENODO metadata editor interface. It consists of several sections:

- Version:** A text input field. Below it, a note states: "Optional. Mostly relevant for software and dataset uploads. Any string will be accepted, but semantically-versioned tag is recommended. See [semver.org](https://semver.org) for more information on semantic versioning."
- Language:** A dropdown menu with "Portuguese" selected and a checkmark icon. Below it, a note states: "Optional. Primary language of the record. Start by typing the language's common name in English, or its ISO 639 code (two or three-letter code). See [ISO 639 language codes list](#) for more information."
- Keywords:** A text input field containing "Oceanographic geographical features" and a close button (X). Below it, a link says "+ Add another keyword".
- Additional notes:** A large text area. Below it, a note states: "Optional."
- License:** A section with a "License" label and a "required" dropdown menu.
- Access right \*:** A section with four radio button options:
  - Open Access
  - Embargoed Access
  - Restricted Access
  - Closed AccessBelow these options, a note states: "Required. Open access uploads have considerably higher visibility on Zenodo."

Figure 42: ZENODO metadata form example.

After filling in the example of Figure 42, the corresponding metadata xml file is produced but is quite limited to describe georeferenced data as shown in the following figure:

# Hourly forecast of the sea surface height for the Tagus estuary

LNEC

## Dublin Core Export

```
<?xml version='1.0' encoding='utf-8'?>
<oai_dc:dc xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:oai_dc="http://www.openarchives.org/OAI/2.0/oai_dc/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.openarchives.org/OAI/2.0/oai_dc/ http://www.openarchives.org/OAI/2.0/oai_dc.xsd">
  <dc:creator>LNEC</dc:creator>
  <dc:date>2018-03-19</dc:date>
  <dc:description>Hourly forecast maps of the sea surface height in meters for the Tagus estuary region. The maps are generated once a day with forecasts of the following 48 hours (2 days).</dc:description>
  <dc:identifier>https://zenodo.org/record/1203234</dc:identifier>
  <dc:identifier>10.5281/zenodo.1203234</dc:identifier>
  <dc:identifier>oai:zenodo.org:1203234</dc:identifier>
  <dc:language>por</dc:language>
  <dc:relation>doi:10.5281/zenodo.1203233</dc:relation>
  <dc:rights>info:eu-repo/semantics/restrictedAccess</dc:rights>
  <dc:subject>Oceanographic geographical features</dc:subject>
  <dc:title>Hourly forecast of the sea surface height for the Tagus estuary</dc:title>
  <dc:type>info:eu-repo/semantics/other</dc:type>
  <dc:type>dataset</dc:type>
</oai_dc:dc>
```

Figure 43: ZENODO – Dublin Core export example.

### 8.4.2 Dublin Core metadata editor

There are several Dublin Core metadata editors since it is a generic standard and it is used to describe many different resources. One of the most commonly used tools for generating Dublin Core standard metadata is available online at <http://dublincoregenerator.com>. It has a Simple Generator and only allows the user to generate Dublin Core metadata using the 15 main elements.

**Input**

Title? [ + ] [ - ]  
 Teste

Creator? [ + ] [ - ]  
 LNEC

Subject? [ + ] [ - ]  
 Agitação marítima

Description? [ + ] [ - ]  
 Dados sobre agitação marítima

Publisher? [ + ] [ - ]  
 LNEC

Contributor? [ + ] [ - ]  
 IPMA

Date? [ + ] [ - ]  
 24-01-2018

Type? [ + ] [ - ]  
 Simulação

Format? [ + ] [ - ]  
 txt/csv

Identifier? [ + ] [ - ]  
 URI ou DOI ou ISBN

Source? [ + ] [ - ]  
 IH - Bóias Ondógrafo

Language? [ + ] [ - ]

Figure 44: Dublin Core Simple Generator.

This tool also has an Advanced Generator option when a more advanced Dublin Core metadata code is necessary. This advanced option used the more detailed qualified elements and encoding schemes, as illustrated below.

Publisher? [ + ] [ - ]

Contributor? [ + ] [ - ]

Date? [ + ] [ - ]  
 Date [ v ] [ v ]

Type? [ + ] [ - ]

Format? [ + ] [ - ]  
 Format [ v ] [ v ]

Identifier? [ + ] [ - ]  
 Identifier [ v ] [ v ]

Source? [ + ] [ - ]

Language? [ + ] [ - ]

Relation? [ + ] [ - ]  
 Relation [ v ] [ v ]

Coverage? [ + ] [ - ]  
 Coverage [ v ] [ v ]

Rights? [ + ] [ - ]  
 Rights [ v ] [ v ]

Audience? [ + ] [ - ]  
 Audience [ v ] [ v ]

Provenance? [ + ] [ - ]

Rights Holder? [ + ] [ - ]

Instructional Method? [ + ] [ - ]

Figure 45: Dublin Core Advanced Generator.

As its name implies this tool only generates metadata not being a real editor because it does not allow opening previously created files. Since the structure of the both types of created metadata files are not very complex, it is possible to later edit the file in any XML editor such as Notepad++.

Like ZENODO's editor, the editors for Dublin core cannot address all the complexity of information from the WADI georeferenced data.



```
<?xml version="1.0" encoding="UTF-8"?>
<metadata
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:dc="http://purl.org/dc/elements/1.1/">
<dc:title>Teste</dc:title>
<dc:creator>LNEC</dc:creator>
<dc:subject>Agitação marítima</dc:subject>
<dc:description>Dados sobre agitação marítima</dc:description>
<dc:publisher>LNEC</dc:publisher>
<dc:contributor>IPMA</dc:contributor>
<dc:date>24-01-2018</dc:date>
<dc:type>Simulação</dc:type>
<dc:format>txt/csv</dc:format>
<dc:identifi er>URL ou DOI ou ISBN</dc:identifi er>
<dc:source>IH - Bóias Ondógrafo</dc:source>
</metadata>
```

Save Generated Metadata to File

Figure 46: Dublin Core output.

### 8.4.3 INSPIRE metadata editor

The INSPIRE Directive requires the Commission to establish a community geo-portal and the Member States shall provide access to their infrastructures through the geo-portal as well as through any access points they themselves decide to operate [31].

The INSPIRE geo-portal has several tools, one of them being a metadata editor specially developed for editing INSPIRE metadata. This editor was planned to be fully complainant with the INSPIRE standard. However, after the European editor initial creation, each country began to develop its own metadata editor that complies with the INSPIRE Directive and this initial editor stopped being updated. Therefore, it cannot be an adequate choice for use in WADI.

The European Open Source Metadata Editor (EUOSME) is a web application to create INSPIRE-compliant metadata in any of 22 European languages, being an adequate substitute for the INSPIRE geoportal editor. It is being developed and maintained by the Joint Research Centre as part of the EuroGEOSS project ([www.eurogeoss.eu](http://www.eurogeoss.eu)) [32].

This online editor is available at <http://eenvplus.sinergis.it/euosmegwt/>. Its interface is very similar to that of the original geoportal INSPIRE editor (see next Figure 47).

**eENUplus**

New Open Validate Save Save in GN Save as template Help About INSPIRE Spatial data series - en

Metadata Identification Classification Keyword Geographic Temporal Quality&Validity Conformity Constraints Interoperability Responsible party

**Metadata on metadata**

Metadata point of contact (\*)

Point of contact 1

Organisation name (\*)

LNEC

An example: <Protection and Conservation of European Seas (PROCAS)>-<Water Resources Unit (H01)>-<Institute for Environment and Sustainability (IES)>-<Joint Research Centre (JRC)-European Commission

E-mail (\*)

aoliveira@lneec.pt

aoliveira@lneec.pt

Metadata date

2018-03-15

Metadata language (\*)

english

\* This field is mandatory

Figure 47: European Open Source Metadata Editor (EUOSME).

EUOSME is a very complete editor as it allows you to create a file and then open it later for new editions. It also has a “Validate” button that uses the INSPIRE Geoportal Metadata Validator to test the compliancy of created metadata with INSPIRE Metadata Regulation.

**eENUplus**

New Open Validate Save Save in GN Save as template Help About INSPIRE Spatial data series - en

Metadata Identification Classification Keyword Geographic Temporal Quality&Validity Conformity Constraints Interoperability Responsible party

**GEOGRAPHIC LOCATION**

Geographic bounding box (\*)

North Bound Longitude	East Bound Longitude	South Bound Latitude	West Bound Longitude
42.15	-6.19	30.14	-31.27

42.15 -6.19 30.14 -31.27

Countries

Portugal

4.1. Geographic bounding box

This is the extent of the resource in the geographic space, given as a bounding box.

The bounding box shall be expressed with westbound and eastbound longitudes, and southbound and northbound latitudes in decimal degrees, with a precision of at least two decimals.

[Commission Regulation \(EC\) No 1205/2008 of 3 December 2008](#)

500 mi

Figure 48: European Open Source Metadata Editor (EUOSME).

After comparing the requirements and advantages of each editor and promoting the opinion of the project partners, the European Open Source Metadata Editor was selected for general use in WADI.

## 9 Conclusions

A description has been provided of the different computation steps the future code will have to perform to transform the pre-processed images (visible, near-infrared and thermal infrared) into a Water Index map. Different possible options have been presented that lead to quite similar results. In the end the choice will be made depending on the type of inputs that will be available (raw signal or absolute reflectance on one side, raw signal or absolute temperature on the other side).

Different methods have also been presented to build the triangle-trapezoid boundaries of the 2D T-VI scatterplot. Based on the results of the Rians campaign, we have shown that the final product only slightly depends on the chosen method. Since they are not very difficult to develop, it would be advisory to implement them all in the software. The GUI will help selecting the parameters for an easy evaluation of the Water Index map

The results of three campaigns are available for the data processing validation, at least for the post-processing steps (a validation of the pre-processing steps is not feasible since the optical instruments and the image characteristics are not the same).

Regarding the standards compliance, it has been defined which data and metadata standards adequate for the information produced in the WADI project. To promote the generation of FAIR data, a list of metadata editors is also reviewed and compared for possible use in the project. Finally, the European Open Source Metadata Editor was selected as the prime choice for metadata generation in WADI.

## 10References

- [1] Krapez, J. C., Olioso, A., & Coudert, B. (2009). Comparison of three methods based on the Temperature-NDVI diagram for soil moisture characterization. In Proc. of SPIE Vol (Vol. 7472, pp. 74720Y-1).
- [2] Krapez, J. C., & Olioso, A. (2011). Improvement of the Triangle method for soil moisture evaluation by adding a third index: albedo or Cellulose Absorption Index. In Proc. of SPIE Vol (Vol. 8174, pp. 81740L-1).
- [3] Krapez, J. C., & Olioso, A. (2011). A combination of temperature, vegetation indexes and albedo, as obtained by airborne hyperspectral remote sensing, for the evaluation of soil moisture. Quantitative InfraRed Thermography Journal, 8(2), 187-200.
- [4] Krapez J.-C., Chatelard C, Nouvel J.-F., Deliot P., "Combined Airborne Thermography And Visible-To-Near Infrared Reflectance Measurement For Soil Moisture Mapping", 11th Quantitative Infrared Thermography Conf., Naples, June 11-14, 2012.
- [5] Chatelard C., Krapez J.-C., Barillot P., Deliot P., Frederic Y.-M., Pierro J., Nouvel J.-F., Helias F., Louvet Y., Legoff I., Serra G., " Multispectral approach assessment for detection of losses in water transmission systems by airborne remote sensing", 13th Int. Hydroinformatics Conf., Palermo, July 1-6, 2018.
- [6] Moran M.S., Clarke T.R., Inoue Y., Vidal A., "Estimating Crop water deficit using the relation between surface air temperature and spectral vegetation index", Remote Sensing of Environment 49, 246-263 (1994).
- [7] Rondeaux G., Steven M., Baret F., "Optimization of soil adjusted vegetation indices", Remote Sensing of Environment, vol. 55, 95-107, 1996.
- [8] Sandholt I., Rasmussen K., Andersen J., "A simple interpretation of the surface temperature-vegetation index space for assessment of surface moisture status", Remote Sensing of Environment 79, 213-224 (2002).
- [9] R. Martins, A. Azevedo, A.B. Fortunato, E. Alves, A. Oliveira, A. Carvalho. (2019). An innovative and reliable water leak detection service supported by data-intensive remote sensing processing. International Conference on Computational Science (ICCS), João M. F. Rodrigues, Pedro J. S. Cardoso, Jânio Monteiro, Roberto Lam, Valeria V. Krzhizhanovskaya, Michael H. Lees, Jack J. Dongarra, Peter M.A. Sloot (eds), LNCS, volume 11540, LNTCS, volume 1154, pp. 96–108, DOI: 10.1007/978-3-030-22747-0\_8.
- [10] <https://www.dataone.org/best-practices/document-and-store-data-using-stable-file-formats>.
- [11] <https://library.stanford.edu/research/data-management-services/data-best-practices/best-practices-file-formats>.
- [12] <https://www2.usgs.gov/datamanagement/plan/dataformats.php>.
- [13] <http://www.loc.gov/preservation/resources/rfs/TOC.html>
- [14] <https://guides.library.upenn.edu/datamgmt/fileformats>.

- [15] <https://en.wikipedia.org/wiki/Metadata>.
- [16] <http://www.language-archives.org/documents/gentle-intro.html>.
- [17] <https://guides.library.uwa.edu.au/c.php?g=325196&p=2178564>.
- [18] [https://en.wikipedia.org/wiki/Metadata\\_standard](https://en.wikipedia.org/wiki/Metadata_standard).
- [19] <http://about.zenodo.org/principles/>.
- [20] <https://www.loc.gov/standards/marcxml/marcxml-design.html>.
- [21] [https://en.wikipedia.org/wiki/MARC\\_standards](https://en.wikipedia.org/wiki/MARC_standards).
- [22] [https://en.wikipedia.org/wiki/Dublin\\_Core](https://en.wikipedia.org/wiki/Dublin_Core).
- [23] [https://omeka.org/classic/docs/Content/Working\\_with\\_Dublin\\_Core/](https://omeka.org/classic/docs/Content/Working_with_Dublin_Core/).
- [24] <https://inspire.ec.europa.eu/about-inspire/563>.
- [25] <http://www.esri.com/news/arcnews/spring10articles/building-inspire.html>.
- [26] <https://www.iso.org/standard/53798.html>.
- [27] <https://www.iso.org/standard/59221.html>.
- [28] <https://inspire.ec.europa.eu/data-specifications/2892>.
- [29] <https://www.wikihow.com/Edit-Metadata>.
- [30] [https://www.dataone.org/software-tools/tags/metadata\\_editor](https://www.dataone.org/software-tools/tags/metadata_editor).
- [31] <http://inspire-geoportal.ec.europa.eu/>.
- [32] <http://eenvplus.sinergis.it/euosmegwt/userguide/about.html>.