# Detecting water using UAV-based GNSS-Reflectometry data and Artificial Intelligence

# Detecting water using UAV-based GNSS-Reflectometry data and Artificial Intelligence

Alfredo Favenza
*Space and Navigation Technologies*
*LINKS Foundation*
Turin, Italy
alfredo.favenza@linksfoundation.com

Rayan Imam
*Dept. of Electronics and Telecommunciations*
*Politecnico di Torino*
Turin, Italy
rayan.imam@polito.it

Fabio Dovis
*Dept. of Electronics and Telecommunciations*
*Politecnico di Torino*
Turin, Italy
fabio.dovis@polito.it

Marco Pini
*Space and Navigation Technologies*
*LINKS Foundation*
Turin, Italy
marco.pini@linksfoundation.com

***Abstract*— GNSS Reflectometry (GNSS-R) is a consolidated remote sensing technique that exploits the back-scattered GNSS signals to retrieve information about the Earth surface. By the use of GNSS-R sensors onboard UAVs, this tool can bring significant advantages in agriculture, especially for the detection of the presence of water/moisure in specific rural areas. However, to perform this detection, GNSS-R needs a priori calibration of a threshold on the received reflected power in order to distinguish the presence/no prence of water. This is a limitation for the adoption of this approach at large scale. The work presented in this paper aims at overcoming such a limitation, proposing a novel approach based on artificial intelligence for the automatic water detection on the Earth surface, avoiding a priori, empirical, thresholding.**

***Keywords—GNSS-R, ICT, agriculture, k-means.***

## I. INTRODUCTION

Active and passive microwave remote-sensing techniques are important sources of data for agriculture [1][2], complementing the optical remote sensing techniques which are however strongly affected by weather conditions and the diurnal cycles [3]. Among the passive microwave techniques lies the GNSS-Reflectometry (GNSS-R) which is a technique based on bistatic radar observations that makes opportunistic use of the signals broadcast by Global Navigation Satellite Systems (GNSS) [4].

GNSS-R was originally developed for oceanographic studies [5], but sooner proved to be a valuable tool for a variety of land surface applications [6][7] including agriculture, where GNSS-R has been used for vegetation detection [8], soil moisture sensing [9][10], vegetation characteristics monitoring [11], vegetation development conditions monitoring [9], and vegetation biomass sensing [10]. In these studies, both air and space borne GNSS-R sensors were utilized, providing moderate resolutions in the former and global coverage in the latter. Recently, the use of GNSS-R sensors on-board UAVs has been proved as a feasible [12] and valuable tool for many applications [13][14] including agriculture [15].

On the other hand, machine learning [16] is the systematic study of intelligent algorithms and systems that improve their knowledge or performance by experience. In its general concept, machine learning process refers to the ability of solving a task, processing right features describing the domain of interest, according to a model.

The use of machine learning is becoming popular for many GNSS applications, such as scintillation [17][18][19] and multipath detection [20] and it is also sustained by the increasing volume of available data collected at remote sites through low cost GNSS software receivers [21].

The main elements of machine learning are:

- domain, the problem to be solved;
- features, the description of the objects of the domain;
- task, the abstract representation of the problem which reflects in the mapping between the input and the output;
- model, the output of the machine learning when the training set is fed to the algorithms.

Machine learning offers a large number of algorithms, to build models from a given dataset of input observations, and to make predictions and decision expressed as output. These algorithms are mainly grouped under three big families:

- supervised learning as shown in Fig. 1, where input data (training set) has a known label or result;
- unsupervised learning as shown in Fig. 2, where input data is not labelled and does not have a known result;
- semi-supervised learning, where input data is a mixture of labelled and unlabelled examples.

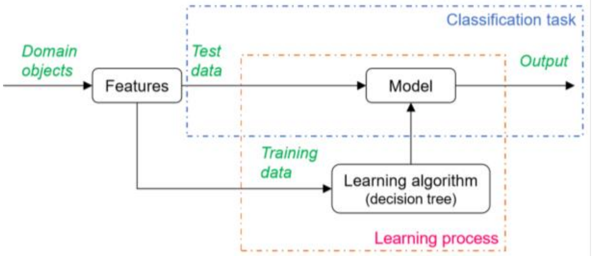


*Fig. 1 Flow diagram of the machine learning process (supervised learning)*

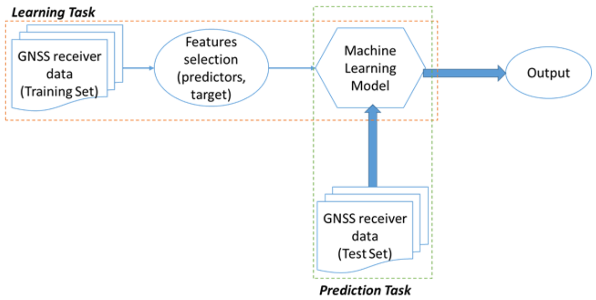


*Fig. 2 Flow diagram of the machine learning process (unsupervised learning)*

In this paper, we are presenting the use of GNSS-R sensors on-board UAVs as a tool for detecting water in agriculture fields. The objective is to investigate the possibility of providing reliable detection of water presence, derived from a low-cost GNSS-R sensor on-board small UAVs. The water detection is based on an unsupervised machine learning technique that is capable of detecting water without any prior knowledge of the sensor used. This eliminates the need for calibrating the measurements and thus setting detection thresholds. The unlabelled nature of the involved datasets, leads us to consider only techniques based on an unsupervised learning process, since they do not need to have a ground truth to perform the classification task.

In particular, we focused on a specific type of clustering algorithm, the K-means, which will be introduced in the section III.

## II. Data Collection and Processing

The data utilized in this paper were collected by a custom made low cost GNSS-R sensor. The sensor was mounted onboard a small UAV and collected data related to water content in fields. The collected data was processed using GNSS signals processing techniques and a metric for the reflected power was extracted. This metric was fed into an unsupervised machine-learning algorithm, which automatically detected the presence of water (or absence). In this section, the experimental setup for data collection is detailed. Then the post-processing procedure is described.

### A. *Experimental system setup*

The custom GNSS-R sensor described in [22] was originally designed to be mounted onboard a piloted aircraft. It was modified as described in [14] to be mounted onboard a small UAV. The modified sensor complies with the low power consumption and low weight constraints of a small UAV.

The original sensor receives GNSS signals using three antennas. The first antenna receives the direct left hand circular polarized (RHCP) GNSS signals. This signal is important because it is processed to determine the position of the UAV. The other two antennas capture the back-scattered LHCP and right hand circular polarized (RHCP) GNSS signals. These reflected signals allow monitoring the parameters of earth surface. Fig. 3 shows the basic architecture of the GNSS-R sensor.

In the modified sensor flown onboard the UAV, only the direct RHCP and the reflected LHCP signals were recorded. Fig. 4 shows the UAV with the sensor mounted onboard.

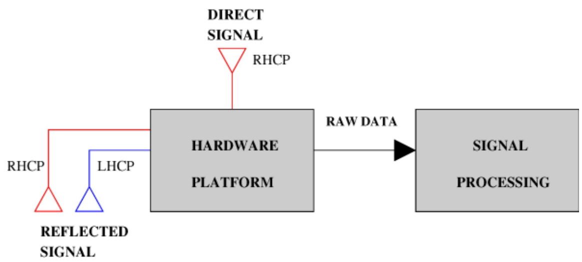


*Fig. 3 Basic architecture of a GNSS-reflectometer*

*Fig. 4 UAV equipped with on-board GNSS-R sensor*

### B. *Post-processing methodology*

The full details of the post processing methodology can be found in [14]. Samples of the LHCP reflected signal were processed by evaluating the cross ambiguity function CAF over a reduced search space. The reflected power metric chosen is the peak-to-noise floor seperation $\alpha_{mean}$ defined as:

$$\alpha_{mean} = \frac{R_p}{M_c} \;\; [dB]$$

Where $\mathrm{R_p}$ is the correlation peak and $\mathrm{M_c}$ is the mean value of the correlation noise. When $\alpha_{\mathrm{mean}}$ is close to zero, it means that the surface did not back-scatter LHCP power. The higher $\alpha_{\mathrm{mean}}$, the higher the reflected power.

### C. *The dataset*

The data used for this experiment represent 30 minutes of reflected GNSS signals from 13 satellites measured with a 20Hz rate. Together, they form approximately half a million observations. The data were collected during a flight over the Avigliana's lake area (North-west of Italy). Fig. 5 reports the reflected signals of one of these satellites. The *x*-axis is the time of data collection from the beginning of the flight. The *y*-axis is the peak-to-noise floor separation in dBs. It can be noticed that, for parts of the flight, the reflected signal power was around zero dB (800-1100 [s]), which indicate that no significant power was back-scattered by the ground surface. For other parts of the flight, the reflected signal was very strong (170-220 [s]) which are typical reflections from water. While on other parts, the reflected power was in between (500-570 [s]). The objective of using the machine-learning algorithm is to find the threshold for water detection automatically without prior knowledge of the sensor used.

## III. K-MEANS UNSUPERVISED LEARNING

The K-means algorithm [16] automatically identifies and forms clusters of similar data samples, iterating between partitioning the data using the nearest-centroid decision rule, and recalculating the centroids from a partition.

It is called K-means because it finds K unique clusters, and the centre of each cluster is the mean of the values in that cluster. In K-means, the K centroids are randomly assigned to a point. Next, each point in the dataset is assigned to a cluster. This assignment is done by finding the closest centroid and assigning the point to a cluster. After this step, the centroids are all updated by taking the mean value of all the samples in that cluster.

More formally, the algorithm consists in the iteration of these two steps:

1. each data is assigned to its nearer centroid, based on the squared Eulcidean distance; if $c_i$ is the collection of centroids in the set $C_i$ then each data point *x* is assigned to a cluster based on

   $$\arg\min_{c_i \in C} dist(c_i, x)^2$$

   where dist(•) is the standard Euclidean distance.

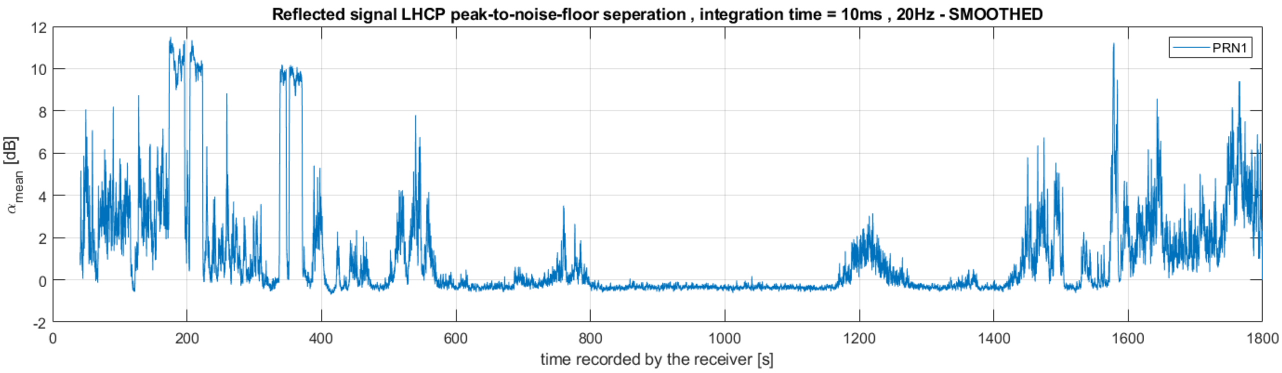


*Fig. 5 peak-to-noise floor seperation for one of the satellites in the data set*

2. Let the set of data point assignment for each $i^{th}$ cluster centroid be $S_i$. Now the centroids are recomputed taking the mean of all data points assigned to the centroid's cluster.

$$c_i = \frac{1}{|S_i|}\sum_{x_i \in S_i} x_i$$

The algorithm iterates between the two steps until no more data points change clusters, the sum of the distances is minimized, or the maximum number of iterations is reached. The algorithm is guaranteed to converge to a local optimum.

### A. *Model Evaluation Metrics*

In the cluster-predict methodology, we can evaluate how well the model performs based on how the different K clusters are selected and assigned to our data. To this aim, we use metrics that are consolidated in the literature [16], as described in the following.

#### 1) *Elbow method and Sum of Squared Error*

One well-known limitation of K-Means is that K the number of clusters has to be a priori declared. The value chosen for K can significantly affect the performance of the model. The goal is to define K such that the total intra-cluster variation or total within-cluster sum of square errors is minimized. One method to select the optimal K is the Elbow method: it is based on the computation of the Sum of Squared Errors (SSE) between data samples and their assigned clusters' centroids and gives an indication of what is the right number of cluster to reach good performances in our specific classification problem.

#### 2) *Silhouette Analysis*

Another metric that can provide clear indication about the quality of the clustering is the Silhouette Analysis. This tool can be used to determine the degree of separation between samples in the clusters by determining a coefficient that can take values in the interval [-1, 1]:

- If the value it is equal to 0, the sample is very close to the neighbouring clusters;
- if it is 1, the sample is far away from the neighbouring clusters;
- if it is -1, the sample is assigned to the wrong clusters.

In order to have a good clustering, this coefficients should be as larger as possible and close to 1 for all the samples belonging to a cluster.

## IV. VI. RESULTS

In this section, the water detection experiment based on GNSS-R and unsupervised machine learning technique is described. A quantitative and qualitative analysis of the results is provided, comparing the performances of the proposed approach with the traditional one based on the manual threshold setting.

### A. *Experimental Results*

In order to assess the K-means performance different types of analysis have been performed.

In this paper, in order to demonstrate the validity of the proposed method only the results for the analysis of a single-satellite (PRN1), are presented and assessed..

#### 1) *Clustering Performance Evaluation*

For GPS PRN1, Fig. 6 shows the results of the Elbow computation. It can be noticed that the SSE value does not decrease significantly when the number of cluster is above 3. This means that K=3 is the optimal number of clusters to classify data in water, not-water and weak-reflection clusters. Increasing the value to K=4 or K=5 will not significantly improve the performance but it would help to identify different degrees of humidity in the soil, which is outside the scope of this paper. Fig. 7 shows the result of the silhouette analysis for PRN1. It can be seen that the data samples $\alpha_{mean}$ have been split over K=3 identified clusters and the silhouette coefficient is, for the majority of the samples, higher than 0.8, that is an evidence of good clustering. In fact, it means that the samples are much closer, on average, to the other members of their cluster than to the members of neighbouring clusters.

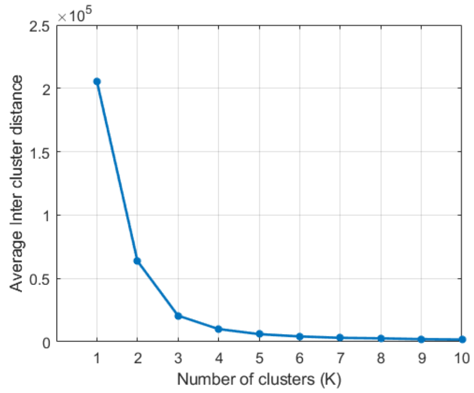


*Fig. 6 Elbow method to determine the optimal K of K-means*

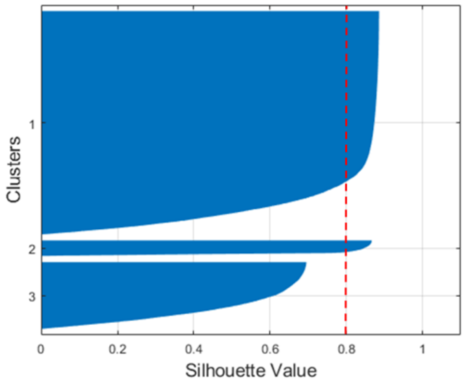


*Fig. 7 Silhouette values for clustering of Reflective Powers for PRN1*

### 2) *Water Detection Analysis*

Fig. 8 reports the time series of $\alpha_{mean}$ (in blue) for the satellite with PRN1 and the class assigned by the K-means algorithm to each $\alpha_{mean}$ value (in red). On the right axis, the cluster assigned for each of these points is shown, depicting (green lines) the threshold values separating the three clusters at 1.533 and 6.27 dB as decided by the machine-learning algorithm. In particular, the latter value corresponds to the value for detecting water presence, which is close to the 6dB value used by the author in [14], decided by the machine learning algorithm without any calibration or prior knowledge about the sensor. The same result can be visually appreciated in and Fig. 10 where the empirical classification and the ML classifications are respectively shown. Different colours are associated to different classes of reflected power (and thus different surfaces types). As it can be noticed in Fig. 9 and Fig. 10, the yellow segments in the middle of the figure identify the presence of two big surfaces of water which are accurately detected by the ML automatic detection. It can be noticed the capability of the proposed approach to identify the presence of the two lakes together with the small portion of land separating them.

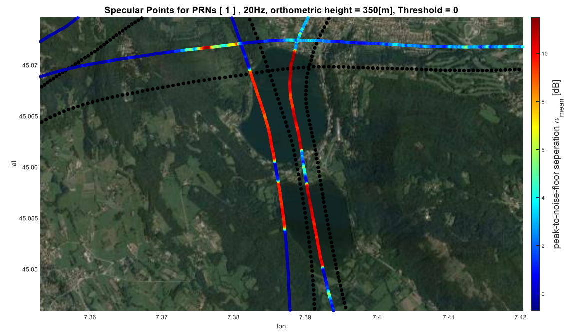


*Fig. 9 Earth surface water detection using the empirical threshold*

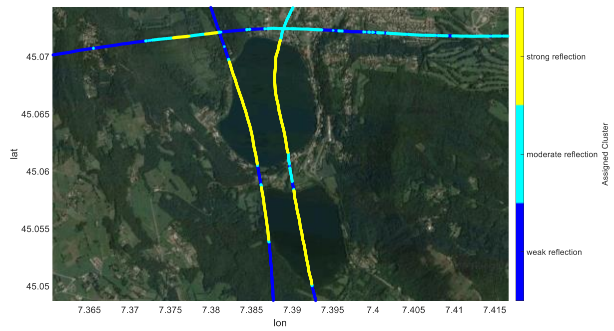


*Fig. 10 Earth surface water detection by the ML approach*

## V. Conclusions

In this paper we presented a method based on a machine learning clustering algorithm for the automatic detection of water presence in GNSS-R data. We showed how the unsupervised algorithm is able to clusterize the data of the surface reflected power, without the a priori knowledge of the threshold.

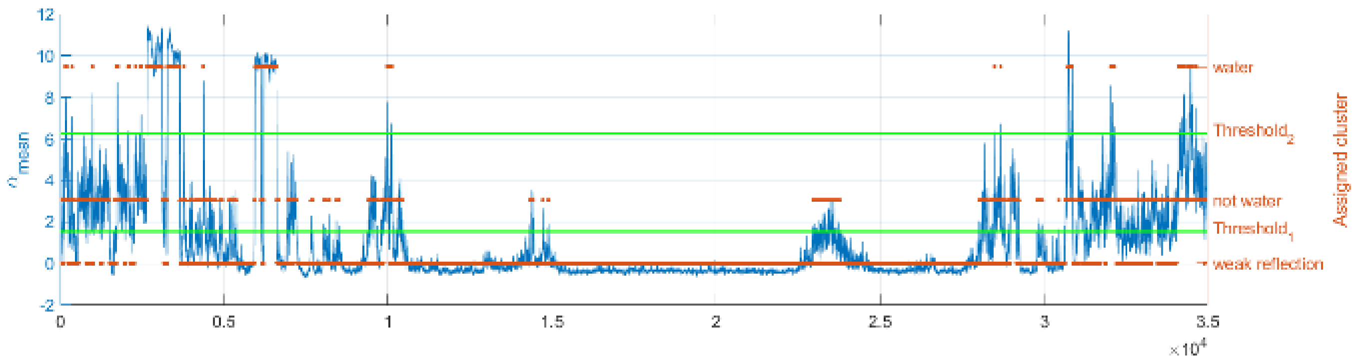


*Fig. 8 Quantitative comparison between traditional thresholding and machine learning classification*

# REFERENCES


[1] W. Wagner, G. Lemoine, and H. Rott, “A Method for Estimating Soil Moisture from ERS Scatterometer and Soil Data,” Remote Sens. Environ., vol. 70, no. 2, pp. 191–207, Nov. 1999.

[2] G. Fontanelli, S. Paloscia, M. Zribi, and A. Chahbi, “Sensitivity analysis of X-band SAR to wheat and barley leaf area index in the Merguellil Basin,” Remote Sens. Lett., vol. 4, no. 11, pp. 1107–1116, Nov. 2013.

[3] R. Fieuzal et al., “Combined use of optical and radar satellite data for the monitoring of irrigation and soil moisture of wheat crops,” Hydrol. Earth Syst. Sci., vol. 15, no. 4, pp. 1117–1129, Apr. 2011.

[4] V. U. Zavorotny, S. Gleason, E. Cardellach, and A. Camps, “Tutorial on Remote Sensing Using GNSS Bistatic Radar of Opportunity,” IEEE Geosci. Remote Sens. Mag., vol. 2, no. 4, pp. 8–45, Dec. 2014.

[5] M. Martin-Neira, “A passive reflectometry and interferometry system (PARIS): Application to ocean altimetry,” ESA J., vol. 17, no. 4, pp. 331–355, 1993.

[6] D. Masters, P. Axelrad, and S. Katzberg, “Initial results of land-reflected GPS bistatic radar measurements in SMEX02,” Remote Sens. Environ., vol. 92, no. 4, pp. 507–520, Sep. 2004.

[7] Y. Jia, P. Savi, D. Canone, and R. Notarpietro, “Estimation of Surface Characteristics Using GNSS LH-Reflected Signals: Land Versus Water,” IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 9, no. 10, pp. 4752–4758, Oct. 2016.

[8] A. Camps et al., “Sensitivity of GNSS-R Spaceborne Observations to Soil Moisture and Vegetation,” IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 9, no. 10, pp. 4730–4742, Oct. 2016.

[9] A. Egido et al., “Global Navigation Satellite Systems Reflectometry as a Remote Sensing Tool for Agriculture,” Remote Sens., vol. 4, no. 8, pp. 2356–2372, Aug. 2012.

[10] S. Paloscia et al., “Grass: An experiment on the capability of airborne GNSS-R sensors in sensing soil moisture and vegetation biomass,” in 2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS, 2013, pp. 2110–2113.

[11] M. Zribi et al., “Potential Applications of GNSS-R Observations over Agricultural Areas: Results from the GLORI Airborne Campaign,” Remote Sens., vol. 10, no. 8, p. 1245, Aug. 2018.

[12] T. Hobiger, J. Strandberg, and R. Gähwiler, “Versatile and low-cost GNSS-R receivers by means of software defined radio,” Am. Geophys. Union, Fall Meet. 2018, Abstr. #G51D-0511, 2018.

[13] R. Notarpietro, S. De Mattia, M. Campanella, Y. Pei, and P. Savi, “Detection of buried objects using reflected GNSS signals,” EURASIP J. Adv. Signal Process., vol. 2014, no. 1, p. 132, Dec. 2014.

[14] R. Imam, M. Pini, G. Marucco, F. Dominici, and F. Dovis, “Data from GNSS-based passive radar to support flood monitoring operations,” in Proc.of 9th International Conference on Localization and GNSS (ICL-GNSS), Nuremberg, 2019.

[15] P. F. Silva et al., “Biomass Applications with Galileo Signals,” 6th Int. Colloq. Sci. Fundam. Asp. GNSS/Galileo, vol. 9.

[16] P. Flach, *Machine learning: the art and science of algorithms that make sense of data*. Cambridge University Press, 2012.

[17] N. Linty, A. Farasin, A. Favenza and F. Dovis, "Detection of GNSS Ionospheric Scintillations Based on Machine Learning Decision Tree," in *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 1, pp. 303-317, Feb. 2019.

[18] Y. Jiao, J. J. Hall, and Y. T. Morton, “Automatic equatorial GPS amplitude scintillation detection using a machine learning algorithm,” IEEE Transactions on Aerospace and Electronic Systems, vol. PP, no. 99, pp. 1–1, 2017.

[19] Caner Savas and Fabio Dovis, Comparative Performance Study of Linear and Gaussian Kernel SVM Implementations for Phase Scintillation Detection in 9th International Conference on Localization and GNSS (ICL-GNSS), 2019. in Proc.of 9th International Conference on Localization and GNSS (ICL-GNSS), Nuremberg, 2019.

[20] Caner Savas and Fabio Dovis, “Multipath Detection based on K-means Clustering” to appear in Proceedings of International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2019), Sept 2019.

[21] Favenza, A., Linty, N., Dovis, F., "Exploiting Standardized Metadata for GNSS SDR Remote Processing: A Case Study," Proceedings of the 29th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2016), Portland, Oregon, September 2016, pp. 77-85.

[22] M. T. Gamba, G. Marucco, M. Pini, S. Ugazio, E. Falletti, and L. Lo Presti, “Prototyping a GNSS-based passive radar for UAVs: An instrument to classify the water content feature of lands,” Sensors (Switzerland), vol. 15, no. 11, pp. 28287–28313, 2015.