AIT Series

Trends in earth observation

Volume 3

Earth Observation: current challenges and opportunities for environmental monitoring

Edited by

Associazione Italiana di Telerilevamento (AIT)

INVESTIGATING THE DEPENDENCY BETWEEN SENTINEL-2 MULTISPECTRAL IMAGES AND GROUND-BASED FIELD MEASUREMENTS OF SOIL MOISTURE IN MENDATICA, LIGURIA, ITALY

A. Iacopino [1], S. Gachpaz [1,2], G. Boni [1], R. Bovolenta [1], G. Moser [3], B.Federici [1] *

[1] Department of Civil, Chemical and Environmental Engineering (DICCA), University of Genoa, Italy
(alessandro.iacopino, rossella.bovolenta, giorgio.boni, bianca.federici)@unige.it, saba.gachpaz@edu.unige.it

[2] Dipartimento di Ingegneria Civile, Edile E Ambientale (DICEA), Sapienza University, Italy – saba.gachpaz@uniroma1.it

[3] Department of Electrical, Electronics and Telecommunication Engineering and Naval Architecture (DITEN), University of Genoa, Italy
gabriele.moser@unige.it

KEY WORDS: Dependency analysis, Machine learning, Multispectral images, Sentinel-2, Surface Soil Moisture

ABSTRACT:
In the present study, the potential of Sentinel-2 (S-2) multispectral satellite images for Surface Soil Moisture (SSM) estimate is investigated. For this purpose, dependency is looked for between S-2 images and an 18-months (from 1st of January 2020 to 30th of June 2021) dataset of hourly SSM measurements, acquired at four different depths (-10cm, -35cm, -55cm, -85cm) from each of the nodes of a monitoring network in Mendatica (Liguria, Italy). Data acquired by the sensors were previously calibrated, considering the soil-specific characteristics of the areas, and the reliability of the dataset was verified. After performing the required preprocessing on satellite images, the performance of three nonlinear regression methods, when applied to four different types of inputs (12 spectral channels, NDVI, NDWI and NDMI), was quantitatively assessed.

1. INTRODUCTION

Surface Soil Moisture (SSM) is an Essential Climate Variable (ECV) that crucially influences rainfall-triggered landslides, where slope stability can be markedly affected by the propagation of the saturation front within the unsaturated zone (Viaggio et al., 2022). SSM can be monitored using traditional methods such as ground-based measurements through contact sensors; they provide accurate but single-point measurements, require manual placement and intensive maintenance and are therefore particularly onerous over wide areas (Nguyen et al., 2022). Since SSM is a heterogeneous variable in terms of space and time, data acquisition with traditional single-point measurement methods is limited to the local scale. Remote Sensing (RS) offers the possibility to continuously observe the land surface and characterize the spatio-temporal variation of the SSM (Adab et al., 2020), because it is one of the influential factors that control the radiation emitted from the Earth's surface (Gao et al., 2013). All parts of the electromagnetic (EM) spectrum normally used for Earth Observation (EO) can be analyzed for quantitative SSM estimate. RS-based methods for SSM retrieval can be classified into three categories, as a function of the type of input data: thermal, microwave, and optical. Most globally available SSM products are derived from microwave RS, due to the ability of microwave radiation to penetrate cloud cover, and their potential to provide all-weather all-time sensing (Yuana et al., 2020). However, soil moisture products obtained from input space-borne passive microwave data are sensitive to surface roughness and have coarse spatial resolution (in the range of km), making them inefficient for studies over small areas (Fang et al., 2019). Active microwave data can be used for retrieving soil moisture at higher spatial resolution (e.g., 20m), although their potential for this retrieval task often decreases in the case of vegetation covered areas (Cui et al, 2023, Graldi and Vitti, 2022). Thermal RS usually exploits the differences between Land Surface Temperature (LST) and air temperature to estimate evaporative fraction as a proxy of the SM (Sini et al, 2008). Separating LST from canopy temperature is anyway a difficult task (Gao et al., 2013). Optical RS, in the visible, near-infrared (NIR), and shortwave infrared (SWIR) ranges, measures the reflected radiation from the earth surface, which can be correlated with soil moisture to provide very high spatial resolution data (Adab et al., 2020). Numerous researches focused on multi-sensor data fusion, satellite-derived vegetation, water and soil indices, and different Machine Learning (ML) techniques (Nguyen et al., 2022). Multispectral indices, like Normalized Difference Vegetation Index (NDVI) and Normalized Difference Water Index (NDWI), have demonstrated strong associations with SSM (Ramat et al, 2022, Hachani et al., 2023, Serrano et al., 2019). Several ML techniques, such as Gradient Boosting Regression (GBR), Support Vector Regression (SVR), Elastic Net Regression (ENR), and Random Forest (RF), showed great potential (Adab et al., 2020). In the present study, the potential of multispectral satellite images acquired by Sentinel-2 (S-2) for SSM extraction at the spatial resolution of 10 meters is investigated. For this purpose, an 18-months dataset of hourly SSM measurements, acquired by sensors placed at four different depths in four nodes of a monitoring network in Mendatica (Liguria, Italy), from 1st of January 2020 to 30th of June 2021, is used to look for dependency with S-2 images. For this purpose, the ML algorithms RF, SVR, and GBR, have been trained for each measurement node and at the various depth, and the resulting regression performance have been evaluated and compared.

2. SOIL MOISTURE MONITORING NETWORK: CHARACTERISTICS, CALIBRATION AND RELIABILITY ANALYSIS

Capacitive sensors are the most versatile and economically sustainable soil moisture sensors. They are relatively easy to install and replace and can be installed in the soil at different depths and locations in the study area, thus creating a monitoring network (Bovolenta et al., 2020). Such monitoring network was installed in Mendatica (Liguria, Italy), using the WaterScout SM100 (Spectrum Tec.). The network consists of five measurement nodes (M1, M2, M3, M4, M5) and a retriever node, that collects and sends data. Each node is connected with four sensors placed at different depths (-10cm, -35cm, -55cm, -85cm), providing information on soil water content along a vertical measuring line (Viaggio et al., 2022). However, the M2 node was not considered in the further analysis, because unrepresentative of the study area. The adopted soil moisture sensors need to be site-specifically calibrated considering the characteristics of the soil samples taken from the study area at the measuring points. The results of the soil-specific calibration are shown in Table 1, in which θ defines the volumetric water content of the soil while the ratio between output and input voltage (V_{out}/V_{in}) represents the raw data from the soil moisture sensors.

* Corresponding author

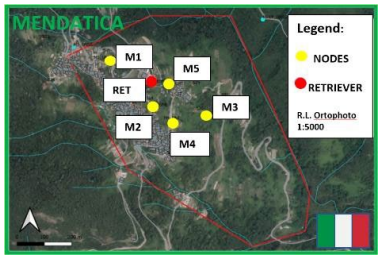

Figure 1. Mendatica soil moisture monitoring network (Regione Liguria orthophoto on the background)

Equation	Number of data samples	a	b	R^2	RMSE [%]
$\theta = \alpha \frac{V_{OUT}}{V_{IN}} + b$	64	286.8	-89.3	0.95	3.1

Table 1. Calibration function for SM100 in Mendatica. R^2: coefficient of determination; RMSE: root mean square error.

The reliability verification of the acquisitions obtained through the SM monitoring network was conducted through the analysis of the correlation between rainfall and volumetric water content θ variations recorded by the four sensors along the vertical of each measurement node. As expected, a higher correlation between rainfall and SM is evident in the shallower sensors, with maximum values positioned at short lags. A decrease in correlation is evident for the deepest sensors, with a progressive delay of the peaks (Figure 2). In Figure 3 an example of SM acquisition performed by M1 node is presented.

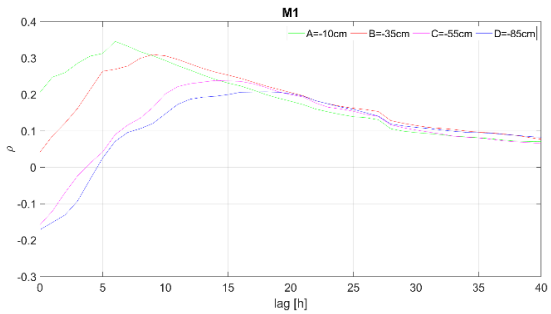

Figure 2. Rainfall-θ cross-covariance referred to May 2021.

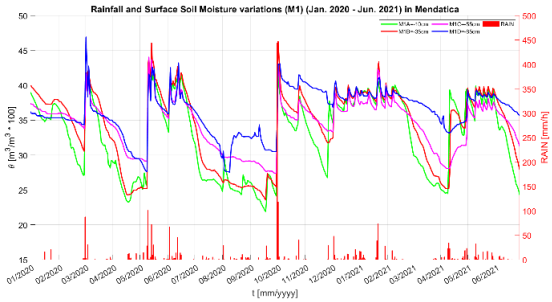

Figure 3. SM data measured by M1 measurement node at different depth, and rainfall data during the study period.

3. SATELLITE REMOTE SENSING DATA

S-2 provides systematic global acquisition of high-resolution multispectral images (MSI) with 12 spectral bands at spatial resolutions of 10m, 20m and 60m. The availability of cloud-based platforms for acquiring and processing satellite images has significantly simplified the implementation of large-scale RS applications like time series analysis (Duan et al., 2020). In this study, we used the Harmonized S-2 MSI Level 2A ("COPERNICUS/S2-SR-HARMONIZED") image collection in Google Earth Engine over the considered study period. This image collection contains Surface Reflectance (SR) values of 12 spectral bands. SR values for all bands of S-2 were used in this study to train regression models. SR was exported for all bands from the image collection with 10 meter spatial resolution on each measurement node in the monitoring network through the study period. Accordingly, a dataset containing values of SR of each band based on the specific geometry for the whole study period was generated.

3.1. Machine Learning for Supervised Regression

In order to train the model, all bands of S-2 were set as input observations and the field-measured SSM at each node (M1, M3, M4 and M5) was considered as the target variable. Separate regression models were trained with measurement data at distinct depths. In the following, working SM sensors (11 in total) connected to measurement nodes are identified with a notation that defines their measurement depth XX in cm (SSMXX). Optical reflectance can serve as an indirect measure of root-zone SM. Most vegetation indices, related to biomass and leaf area index like Normalized Different Vegetation Index (NDVI), or canopy water-based indices like Normalized Different Water Index (NDWI) and Normalized Difference Moisture Index (NDMI), are associated with root-zone SM (Liu et al, 2018). Subsequently, each regression method was trained using three indices in input: the NDVI, the NDWI, and the NDMI. Each site was considered independent because of the different land-use types at the various locations. Hence, soil moisture content has been estimated at four sites and the available depths, by developing eleven individual models for each regression algorithm used, i.e. RF, SVR, and GBR. The following model selection settings were chosen: for RF, the maximum number of features to be used in each split was set using the well-known rule of thumb of the square root, while the number of trees in the ensemble was fixed to 10.000. For SVR, four different kernels, i.e., linear, sigmoidal, polynomial, and Gaussian radial basis function (RBF) were used. The regularization parameters C and ε were set to 1 and 0.1, respectively. For GBR, the number of estimators, the learning rate (shrinkage), the minimum number of observations at each node and the maximum depth were 100, 0.1, 10 and 3, respectively. In this study, the dataset was split as 80% for training the models and 20% for testing their performances. The total number of acquisitions for sites M1 (land use: agriculture), M3 (land use: woods and brambles), M4 (land use: agriculture close to houses) and M5 (land use: agriculture close to woods) were 106, 113, 108, and 108, respectively. Discrepancies in acquisition numbers arose due to the cloud cover or shadow obscuring the desired pixels. External tests help to ensure the robustness of the model. Accuracy assessment was done using the root mean squared error (RMSE) to evaluate the difference between the observed values of SSM and the retrieved values computed by the different regression algorithms on the test samples.

4. RESULTS AND CONCLUSIONS

The main goal of this study was to explore the relationship between S-2 MSI data and ground-based SSM measurements. Initially, 12 different spectral channels (B1 to B12) were

considered as potential variables for model training. Subsequently, common vegetation indices were examinated, due to the significant impact of vegetation cover on the received surface reflectance from earth surface. Table 2 shows that the accuracy of the results is considerably dependent on the training dataset, hence the land-use types at each location and the analysed depth. The highest dependency between the measured soil moisture and the inputs of the regression models was observed at the deepest measurement depth across all monitoring nodes (M1, M3, M4, and M5). Except for the M4 sensor, which operates in areas characterized by high heterogeneity in land use, both NDVI and NDMI exhibited a more pronounced dependency on soil moisture for the remaining measurement nodes. This highlights the substantial impact of vegetation dynamics on soil moisture assessment, particularly as measurement depth increases. Previous studies (Liu et al, 2018) have emphasized the effectiveness of vegetation indices such as NDVI in elucidating the relationship between SSM, root-zone SM, and crop water content. Root-zone SM significantly influences vegetation cover and alters the surface energy balance.

			M1 Node				M3 Node		M4 Node			M5 Node	
	Method	kernel	SSM10	SSM30	SSM50	SSM80	SSM10	SSM85	SSM10	SSM35	SSM55	SSM10	SSM50
12 Channels	RF		4.21	5.12	3.42	**2.62**	3.64	2.13	**6.43**	6.43	2.55	**2.89**	1.87
	SVR	Linear	**3.39**	**3.67**	3.11	3.26	**2.69**	**1.76**	6.71	5.31	5.80	4.90	2.62
		Sigmoid	4.18	4.46	**3.08**	3.11	2.99	2.43	8.40	4.82	4.71	4.97	2.84
		RBF	4.34	4.72	3.16	**2.62**	3.33	2.27	8.54	4.51	5.00	4.77	1.78
		Polynomial	4.58	5.05	3.46	3.36	2.87	2.05	13.82	21.99	16.1	6.00	2.89
	GBR		5.47	5.35	4.04	3.29	3.35	2.39	6.74	**3.02**	**2.25**	3.16	**1.68**
NDVI	RF		**2.93**	**3.64**	**1.81**	**1.76**	**2.15**	**1.38**	7.63	6.12	6.59	**2.15**	**1.42**
	SVR	Linear	5.71	6.76	3.64	3.43	3.94	2.79	7.61	7.61	7.66	4.60	3.15
		Sigmoid	5.71	6.85	5.47	5.69	5.47	4.72	7.86	8.06	8.20	6.12	5.82
		RBF	5.66	6.68	3.54	3.37	3.85	2.75	**7.32**	7.56	7.60	4.52	2.91
		Polynomial	5.72	6.78	3.61	3.46	3.95	2.78	8.03	8.12	8.05	4.68	3.20
	GBR		4.63	6.03	3.96	5.09	3.27	2.27	8.65	**5.02**	**5.22**	3.70	2.41
NDWI	RF		**2.50**	**3.09**	**1.82**	**1.93**	**2.15**	**1.29**	**4.52**	**3.05**	**3.12**	**2.17**	**1.45**
	SVR	Linear	4.91	6.09	3.21	3.49	3.99	2.85	8.12	5.54	5.82	4.58	2.99
		Sigmoid	4.93	6.14	3.23	3.51	5.92	5.86	8.58	7.19	7.40	4.68	4.65
		RBF	4.97	6.15	3.21	3.46	3.85	2.66	8.17	5.33	5.63	4.48	2.87
		Polynomial	4.89	6.07	3.26	3.49	3.96	2.79	8.26	5.77	6.22	4.61	3.09
	GBR		3.69	4.64	2.47	2.73	3.28	2.15	6.51	4.34	4.38	3.64	2.29
NDMI	RF		**2.78**	**3.29**	**1.74**	**1.86**	**1.88**	**1.27**	**4.52**	**3.05**	**3.12**	**2.31**	**1.61**
	SVR	Linear	5.59	6.53	3.51	3.51	3.81	2.82	8.79	6.19	6.62	4.60	3.24
		Sigmoid	5.62	6.49	3.74	5.51	5.18	4.00	8.85	6.15	7.18	5.00	3.41
		RBF	5.51	6.33	3.35	3.26	3.68	2.71	8.39	5.92	6.02	4.46	2.96
		Polynomial	5.64	6.77	3.65	3.47	3.88	2.89	8.92	6.12	6.59	4.53	3.35
	GBR		4.46	5.13	2.73	2.77	3.03	2.07	7.14	4.77	4.75	3.57	4.72

Table 2. RMSE of Volumetric Water Content data (expressed as percentage) using different ML algorithms. The lowest values for each sensor and each training are highlighted in bold.

The findings of this study indicate that the greenness-based vegetation index (NDVI) is more reliant on variations in soil moisture in the root-zone area in agricultural land use cases (M1 and M5), compared to canopy water content indices like NDWI and NDMI, particularly evident in M3 with denser vegetation cover. Among the three algorithms used to estimate soil moisture, the RF method demonstrated the highest proficiency in soil moisture retrieval, particularly when combined with NDVI, consistent with findings from other studies (Wang and Fu, 2023). For future works, it is suggested to consider the homogeneity of land use within each pixel and verify that the selected pixel accurately represents the area surrounding the sensor.

ACKNOWLEDGEMENTS

The authors acknowledge the support given by AD-VITAM Partners. A particular acknowledgement goes to the Unione dei Comuni della Valle Argentina e Armea and to the lab technicians of the Department of Civil, Chemical and Environmental Engineering of the University of Genoa.

REFERENCES

Adab, H., Morbidelli, R., Saltalippi, C., Moradian, M., and Fallah Ghalhari, G.A., 2020. “Machine Learning to Estimate Surface Soil Moisture from Remote Sensing Data”. Water, 12(11). https://doi.org/10.3390/w12113223.

Bovolenta, R.; Iacopino, A.; Passalacqua, R.; Federici, B., 2020. “Field Measurements of Soil Water Content at Shallow Depths for Landslide Monitoring”. Geosciences 2020, 10, 409. https://doi.org/10.3390/geosciences10100409.

Cui, H., Jiang, L., Paloscia, S., Santi, E., Pettinato, S., Wang, J., Fang, X., Liao, W., 2021. “The potential of ALOS-2 and Sentinel-1 radar data for soil moisture retrieval with high spatial resolution over agroforestry areas, China”. IEEE transactions on geoscience and remote sensing.

Duan, Y., Li, X., Zhang, L., Chen, D., Liu, S., and Ji, H., 2020. “Mapping national-scale aquaculture ponds based on the Google Earth Engine in the Chinese coastal zone”. Aquaculture, 520. https://doi.org/10.1016/j.aquaculture.2019.734666

Fang, B., Lakshmi, V., Jackson, T, J., Bindlish, R., Coliander, A., 2019. “Passive/ active microwave soil moisture change disaggregation using SMAPVEX12 data”. Journal of Hydrology, 574, 1085-1098.

Gao, Z., Xu, X., Wang, J., Yang, H., Huang, W., Feng, H., 2013. “A method of estimating soil moisture based on the linear decomposition of mixture pixels”. Mathematical and computer modelling, 58(3-4), 606-613. https://doi.org/10.1016/j.mcm.2011.10.054.

Graldi, G., Vitti, V., 2022. “Identifying time patterns at the field scale for retrieving superficial soil moisture on an agricultural area with a change detection method: a preliminary analysis”. The international archives of photogrammetry, remote sensing and spatial information science, volum XLIII-B3-2022. XXIV ISPRS congress, 6-11 June 2022, Nice, France. https://doi.org/10.5194/isprs-archives-XLIII-B3-2022-879-2022.

Hachani, A., Ramat, G., Paloscia, S., Santi, E., Baroni, F., Fontanelli, G., Lapini, A., Pettinato, A., Pilia, S., Santurri, L., 2023. “Remote sensing techniques for soil humidity monitoring in drought areas: case study of the Wadi Hallouf/Oum Zessar watershed (Tunesia)”. International workshop on meteorology for agriculture and forestry.

Liu, L., Yang, X., Liu, S., Zhou, L., Li, X., Yang, J., Wu, J., 2018. “Relationship of root zone soil moisture with solar-induced chlorophyll fluorescence and vegetation indices in winter wheat: a comparative study based on continuous ground-measurements”. Ecological indicators, 90, 9-17. https://doi.org/10.1016/j.ecolind.2018.02.048

Nguyen, T., Ngo, H, H., Guo, W., Chang, S, W., Nguyen, D, D., Nguyen, C, T., Zhang, J., Liang, S., Bui, X, T., Hoang, N, B., 2022. “A low-cost approach for soil moisture prediction using multi-sensor data and machine learning algorithm”. Science of The Total Environment, 833.

Ramat, G., Santi, E., Paloscia, S., Fontanelli, G., Pettinato, S., Santurri, L., Souissi, N., Da Ponte, E., Abdel Wahab, M., Khalil, A., H.Essa, Y., Ouessar, M., Dhaou, H., Sghaier, A., Hachani, A., Kassouk, K., Chabaane, Z., 2022. “Remote sensing techniques for water management and climate change monitoring in drought areas: case studies in Egypt and Tunisia”. European Journal of Remote Sensing, 56(1). https://doi.org/10.1080/22797254.2022.2157335

Serrano, J., Shahidian, S., Da Silva, J., 2019. “Evaluation of normalized difference water index as a tool for monitoring pasture seasonal and inter annual variability in a mediterranean agro-silvo-pastoral system”. Water, 11(1). https://doi.org/10.3390/w11010062

Sini, F., Boni, G., Entekhabi, D. , 2009. “Measurements of Hydrological Variables from Satellite: Application to Mediterranean Regions”. In: Sorooshian, S., Hsu, KL., Coppola, E., Tomassetti, B., Verdecchia, M., Visconti, G. (eds) Hydrological Modelling and the Water Cycle. Water Science and Technology Library, vol 63. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-77843-1_5

Viaggio, S.; Iacopino, A.; Bovolenta, R.; Federici, B., 2022. “Landslide Susceptibility Assessment: Soil Moisture Monitoring Data Processed by an Automatic Procedure in GIS for 3D Description of the Soil Shear Strength”. The international archives of photogrammetry, remote sensing and spatial information sciences. XLVIII-4/W1-2022: 517–523. https://doi.org/10.5194/isprs-archives-XLVIII-4-W1-2022-517-2022.

Wang, S., and Fu, G., 2023. “Modelling soil moisture using climate data and normalized difference vegetation index based on nine algorithms in alpine grasslands”. Front.Environ.Sci. 11:1130448.doi: 10.3389/fenvs.2023.1130448

Yuana, Q., Xua, B., H., Lic, T., Shenb, C., D., H., and Zhang, L., 2020. “Estimating surface soil moisture from satellite observations using a generalized regression neural network trained on sparse ground-based measurements in the continental U.S”. Hydrology, 580