Plant Archives Vol. 25, No. 2, 2025 pp. 263-268 e-ISSN:2581-6063 (online), ISSN:0972-5210

Plant Archives

Journal homepage: http://www.plantarchives.org
DOI Url : https://doi.org/10.51470/PLANTARCHIVES.2025.v25.no.2.038

SATELLITE SOIL MOISTURE ACCURACY: ASSESSMENT AND ENHANCEMENT OVER PUNE, INDIA

Harshita Rani Ahirwar[1]*, Anupam Kumar Nema[1] and Manish Kumar Nema[2]

[1]Department of Farm Engineering, Institute of Agricultural Sciences, Banaras Hindu University, Varanasi, India.
[2]Water Resources Systems Division, National Institute of Hydrology, Roorkee, Haridwar, India.
*Corresponding author E-mail : writetoharshita.bhu@gmail.com
(Date of Receiving-21-05-2025; Date of Acceptance-28-07-2025)

ABSTRACT

Soil moisture plays an essential role in the energy and mass balance between the land and atmosphere by influencing agriculture, hydrology, and climate. The present study evaluates the performance of four soil moisture products-AMSR-2, MERRA-2, SMAP, and SMOS—against in-situ soil moisture data provided by the Soil Moisture Indian Network at the Pune station from 2016 to 2020. The comparison statistics indicate that AMSR-2 has the least correlation, whereas MERRA-2 exhibits the highest correlation (R = 0.91), followed closely by SMAP (R = 0.89) and SMOS (R = 0.87). SMOS and SMAP achieved the lowest RMSE (13.6 mm^3/mm^3 and 14.0 mm^3/mm^3, respectively), while AMSR-2 had the most significant errors. Bias adjustment methods were applied to each satellite product to determine the best-fitting technique. It was observed that different satellites require different approaches to suit them well. The study supports the use of soil moisture estimation products from MERRA-2, SMAP and SMOS for regional applications.

Key words : AMSR-2, MERRA-2, SMAP, SMIN COSMOS, SMOS.

Introduction

Soil moisture is one of the critical variables in the hydrological cycle, influencing land–atmosphere interactions, agricultural productivity, and weather and climate systems. It governs the sectionalization of rainfall into infiltration and runoff, affects evapotranspiration rates, and plays a key role in drought and flood forecasting (Entekhabi *et al.*, 2010 and Kerr *et al.*, 2012). Accurate soil moisture information is essential for improving agricultural water management, climate modelling, and early warning systems for hydrological hazards (Brocca *et al.*, 2010).

In-situ soil moisture monitoring delivers precise and continuous ground-based observations. However, measuring soil moisture on the ground is often a costly process and is spatially limited. On the contrary, the remote sensing method of dataset assimilation can overcome these limitations by consistently offering soil moisture for a large area at regular intervals. Among the prominent satellite and reanalysis sources are the Advanced Microwave Scanning Radiometer 2 (AMSR-2), the Soil Moisture Active Passive (SMAP) mission, the Soil Moisture and Ocean Salinity (SMOS) satellite, and the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA-2) dataset (Reichle *et al.*, 2004; Njoku *et al.*, 2005; Kerr *et al.*, 2010).

Site-specific validation of these products becomes crucial due to the dependency of soil moisture on local conditions such as climatic regime, vegetation cover, and soil type, which leads to variability in accuracy. For a data-scarce region like India, it becomes increasingly crucial to validate these global products for planning at the regional level (Dorigo *et al.*, 2015; Srivastava, Petropoulos and Kerr, 2016).

This study aims to evaluate the consistency and accuracy of AMSR-2, SMAP, SMOS, and MERRA-2 soil moisture products against COSMOS observations at the Pune site over 5 years (2016–2020). In addition to standard statistical measures such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), bias, and Pearson correlation, we apply six regression models—linear, exponential, logarithmic, polynomial, power and quantile regression—to determine the best-fitting relationships. The study not only contributes to the validation of satellite products against in situ data but also provides insights into the correction of data to best fit ground data for greater accuracy and reliability.

Materials and Methods

Study area

Pune station is geographically situated at 18.32° N latitude and 73.48° E longitude at approximately 560m of elevation above mean sea level. It is located in a semi-arid agroclimatic zone in India. Pune experiences a tropical wet and dry climate (Köppen: Aw), with distinct wet (monsoon) and dry seasons. The average annual rainfall ranges from 700 to 800 mm. The summer season (March to May) sees a gradual increase in temperature, whereas the winter season (November to February) is typically dry and cool. The study area at the station features a soil type of sandy clay loam, characterized by a moderate water-holding capacity and variable infiltration rates.

Data used

The present study utilizes in-situ soil moisture and remotely sensed soil moisture by satellite/reanalysed data products collected from 2016 to 2020. The combination of both types of datasets enables a comprehensive assessment of the behaviour of moisture products at the Pune Station.

In-situ data – COSMOS Probe (SMIN)

Soil Moisture Indian Network (SMIN)—an initiative by Indian scientific agencies to establish long-term monitoring of soil moisture using advanced technologies. Among the eight stations, Pune station plays a vital role in this network, representing the semi-arid climatic zone and contributing high-quality data for research and validation studies.

The *in-situ* data used in this study were obtained from the COSMOS (Cosmic Ray Soil Moisture Observing System) probe installed at Pune as part of the Soil Moisture Indian Network (SMIN). This sensor measures neutron flux in the lower atmosphere, which is inversely related to soil moisture content within a radius of approximately 150–200 meters, covering an area of around 20 hectares.

The COSMOS probe provides daily average volumetric soil moisture values (in mm^3/mm^3), with a representative sensing depth of ~0–20 cm, depending on soil type and moisture conditions. The data span from January 2016 to December 2020. These measurements are considered ground truth due to their established reliability, spatial representativeness, and calibration to site-specific soil conditions.

Satellite and Reanalysis Soil Moisture products

To study the performance of different remotely sensed soil moisture products available, we have employed the following widely used satellite and reanalysed datasets. These datasets differ in spatial resolution, temporal frequency and sensing technique. Table 1 contains information regarding the soil moisture products used in the present study.

Methodology

This section outlines the approach used to evaluate the performance of satellite and reanalysis soil moisture products relative to *in-situ* COSMOS observations at

Table 1 : Details of Remote Sensing Soil Moisture data.

Parameter	AMSR-2	SMAP	SMOS	MERRA-2
Sensor Type	Passive microwave	L-band radiometer (active component disabled post-2015)	L-band radiometer	Reanalysis product
Resolution	~25 km	~9 km	~40 km	0.5° × 0.625°
Temporal Frequency	3-hourly (aggregated to daily)	2–3 days revisit	2–3 days revisit	Daily
Source	JAXA (Japan Aerospace Exploration Agency)	NASA	ESA (European Space Agency)	NASA's Global Modeling and Assimilation Office (GMAO)
Notes	Limited accuracy in vegetated and semi-arid regions due to coarse resolution and surface emissivity effects	Designed for topsoil moisture estimation; reliable in low-vegetation areas	Uses synthetic aperture; suitable for large-scale moisture variability	Integrates satellite data into land-surface models; spatially continuous and consistent over time

Fig. 1 : Number of paired observations.

Pune station. Satellite data were reprojected and extracted at the COSMOS station's geographic coordinates. For meaningful comparison, only days with soil moisture values available from both the COSMOS station and each satellite product were considered. This step ensured temporal consistency and allowed for unbiased statistical comparisons. The number of paired observations (n) for each dataset is shown in Fig. 1. The methodology includes data pairing, statistical performance assessment, regression modelling, and visualization techniques. A flowchart explains the work process (Fig. 2).

Statistical Performance Metrics

To assess how closely satellite soil moisture data match the *in-situ* COSMOS observations, the following performance metrics were computed (Table 2). These metrics collectively provide insight into the precision, accuracy, and consistency of each dataset.

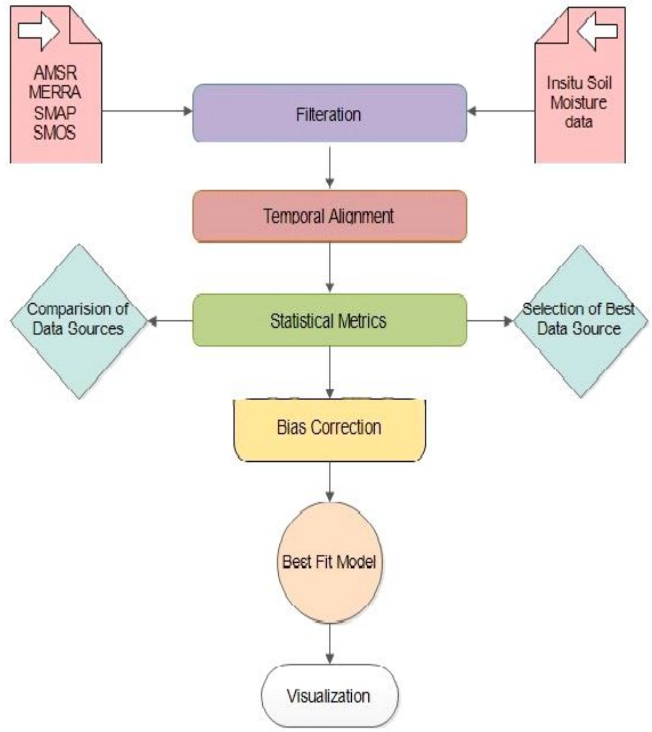

Fig. 2 : Flow chart of methodology adopted for the study.

Regression Model Fitting

To better understand the relationship between satellite-derived soil moisture estimates and in-situ observations from the COSMOS probe, six distinct regression models were employed: Linear, Exponential, Logarithmic, Polynomial, Power and Quantile regression. Each model has a different ability to read various aspects, including linearity, curvature, multiplicative effects, and distributional asymmetries. The motive of applying multiple regression types is to know the best mathematical formulation that defines the relationship between the two datasets and helps in removing the bias value from the

Table 2 : Statistical parameters used for comparison.

Metrics	Formula	Description
Mean Absolute Error (MAE)	$\frac{1}{n}\sum_{i=1}^{n}\lvert SM_{SAT} - SM_{SMIN}\rvert$	Measures the average magnitude of error without considering direction.
Root Mean Square Error (RMSE)	$\sqrt{\frac{1}{n}\sum_{i=1}^{n}(SM_{SAT} - SM_{SMIN})^2}$	Emphasizes larger errors more than MAE due to the squaring operation.
Bias	$\frac{SM_{SAT} - SM_{SMIN}}{n}$	Indicates systematic overestimation (positive) or underestimation (negative).
Pearson Correlation Coefficient (R)	$\frac{\sum_{i=1}^{n}\left(SM_{SAT} - \overline{SM}_{SAT}\right)\left(SM_{SMIN} - \overline{SM}_{SMIN}\right)}{\sqrt{\sum_{i=1}^{n}\left(SM_{SAT} - \overline{SM}_{SAT}\right)^2 \sum_{i=1}^{n}\left(SM_{SMIN} - \overline{SM}_{SMIN}\right)^2}}$	Measures linear relationship between satellite and *in-situ* observations.

read dataset.

Linear Regression : Linear regression assumes a direct relation between insitu (observed) and satellite (estimated) soil moisture. It is one of the most widely used models due to its interpretability and simplicity.

Exponential Regression : Exponential models capture multiplicative growth, where small changes in input (*in-situ* data) can lead to significant changes in output. The model may be suitable for areas where the dynamics of soil moisture are governed by rapid influx, such as drying or wetting events, where data displays exponential decay or rise. The method is advantageous in areas with well-drained soil with non-linear responses.

Logarithmic Regression : Logarithmic models are suitable when the relationship increases rapidly at low values and levels off at higher values. The model implies diminishing returns with increasing soil moisture, as seen in saturated soils where supplementary moisture does not significantly change radiometric signals. This method is helpful for lands with vegetative cover and clay-rich soil.

Polynomial Regression (2nd Degree) : This model extends linear regression by introducing curvature, capturing relationships that are parabolic due to factors such as heterogeneous soil profiles, sensor sensitivity, or vegetation interference. It allows for different rates of change at low vs. high moisture values and is more flexible than simple linear forms

Power Regression : This model assumes a multiplicative, scale-invariant relationship. It is commonly used in geophysics and environmental sciences where variables are related through power laws, particularly when both low and high ranges are important.

Quantile Regression : Unlike ordinary least squares, which estimates the mean response, quantile regression estimates conditional quantiles (e.g., median, 25th or 75th percentile). It is robust to non-normality, skewness and outliers, which are frequent in environmental datasets. Especially useful for understanding performance at extremes (like droughts or floods). For example, quantile regression can show whether a satellite underestimates soil moisture during very dry or wet periods, supporting bias correction efforts.

Results and Discussion

Time Series analysis

Fig. 3 illustrates the daily time series comparison between in-situ soil moisture (dotted line) and the four satellite products (continuous line). A seasonal trend is evident in all datasets, with increased moisture during the monsoon periods (JuneSeptember) and lower values during the pre-monsoon season (MarchMay). MERRA-2 shows strong temporal coherence with the COSMOS data, closely matching seasonal amplitude and peak timing. SMAP and SMOS also follow the seasonal dynamics, though their magnitude is slightly lower during peak monsoon, consistent with the observed negative bias. AMSR-2 diverges the most, with both a noticeable lag and underestimation during key seasonal peaks. The time series plot supports the hypothesis that satellite and reanalysis products, especially SMAP, SMOS and MERRA-2, are capable of capturing the temporal evolution of soil moisture but require localized calibration for amplitude correction.

Overview of Satellite Product performance

The accuracy and reliability of satellite- and reanalysis-based soil moisture products were assessed by comparing them with *in-situ* measurements from the COSMOS station located in Pune. Four datasets were evaluated: AMSR-2, MERRA-2, SMAP, and SMOS. The evaluation period spanned from 2016 to 2020, during which statistical performance metrics, including MAE, RMSE, bias and Pearson correlation coefficient (R), were computed and evaluated (Fig. 4).

MERRA 2 showed a strong linear relationship with in-situ COSMOS data, with the highest correlation coefficient (R = 0.91) among the four satellite datasets. With R values of 0.89 and 0.87, respectively, SMAP and SMOS closely trailed behind, while AMSR-2 showed the lowest correlation (0.65). In terms of RMSE, SMOS exhibited the lowest error at 13.55 mm^3/mm^3, followed by SMAP (13.98 mm^3/mm^3) and MERRA-2 (15.19 mm^3/mm^3). AMSR-2 again performed poorly, with the highest RMSE of 22.49 mm^3/mm^3, indicating a substantial deviation from the observed soil moisture values. The MAE followed a similar trend, confirming these insights. Bias analysis revealed that AMSR-2 and MERRA-2 tended to overestimate soil moisture, whereas SMAP and SMOS exhibited a slight underestimation of soil moisture. This has practical implications: while the magnitude of errors was lower for SMAP and SMOS, the sign of the bias needs to be considered for operational purposes.

The Taylor diagram (Fig. 5) provides a comprehensive statistical comparison of COSMOS in-situ observations and soil moisture products derived from satellites. The centred root mean square difference, standard deviation, and correlation coefficient are all shown at the same time. SMAP, SMOS, and MERRA-2 exhibit comparable variability and high correlation with the reference, indicating good agreement among the evaluated products.

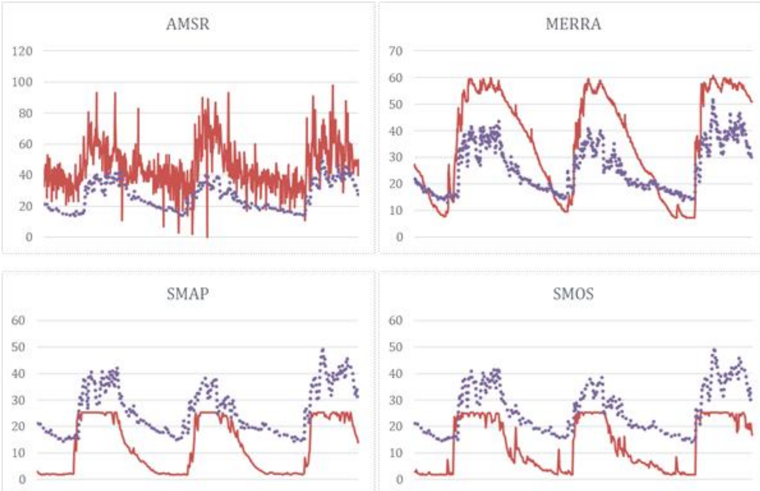

Fig. 3 : Time series graph at Pune station.

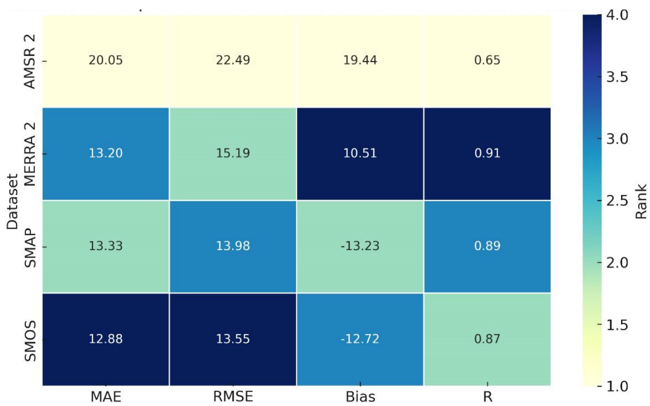

Fig. 4 : Performance Metrics for Pune Station.

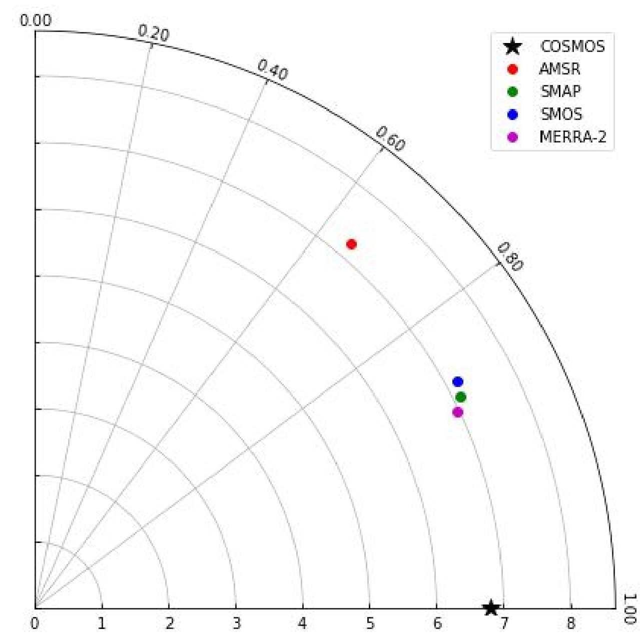

Fig. 5 : Taylor Diagram-based evaluation of Satellite Soil Moisture Accuracy.

AMSR, on the other hand, underestimates variability and shows less correlation. The relative performance of each satellite dataset is clearly illustrated in the diagram (Fig. 5), where the radial distance represents the standard deviation and the angle from the horizontal axis represents the correlation coefficient. The star-marked value is the centred root mean square difference of the in situ dataset.

Regression Model performance

To better understand the functional relationship between in-situ and satellite-based soil moisture, six regression models were fitted for each dataset: Linear, Exponential, Logarithmic, Polynomial, Power and Quantile regression. The correlation coefficients (R) obtained from these models are summarized in Fig. 6.

Polynomial and quantile regression models consistently produced the highest correlation values across most datasets. For SMAP and SMOS, polynomial regression (second degree) achieved R-squared values of 0.894 and 0.879, respectively, highlighting their non-linear relationships with the COSMOS data. This suggests that a quadratic correction function may be helpful for local calibration of satellite products.

Firm performance was found for AMSR 2 using quantile regression (R = 0.951), suggesting that this approach is more sensitive to outliers and better able to manage the extremes of soil moisture variability. Despite its poor performance under conventional models, AMSR 2's high quantile regression fit raises the possibility that it could be beneficial in some hydrologically extreme situations, such as post-monsoon saturation or drought. On the other hand, logarithmic regression performed the worst for all products, suggesting that log-linear models do not adequately capture soil moisture dynamics at this location, perhaps due to their asymptotic behaviour.

Strengths of SMAP, SMOS and MERRA-2

The overall performance of SMAP and SMOS reflects their advanced sensor technology (L-band radiometers), which are less affected by vegetation and surface roughness. Their relatively higher spatial

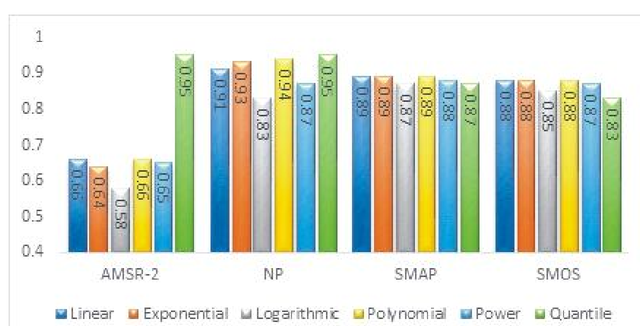

Fig. 6 : Correlation Coefficients (R) for Regression Models.

resolution (940 km) also contributes to improved accuracy at the Pune site. The integration of various observation types and model physics is advantageous for MERRA 2, as it is a product of reanalysis. Particularly in situations where *in-situ* data are limited, MERRA-2's strong correlation and reliable performance imply that it can be used as a reference dataset for regional hydrological studies.

Conclusion

This study conducted a comprehensive comparison between in-situ soil moisture observations from the COSMOS probe at Pune station and four satellite- or reanalysis-based soil moisture datasets: AMSR-2, SMAP, SMOS and MERRA-2, over the 20162020 period. The aim was to evaluate their accuracy, understand functional relationships and identify suitable models for bias correction or data integration in hydrological studies.

Statistical analyses showed that MERRA-2 emerged as the most consistent and accurate product, exhibiting the highest Pearson correlation coefficient (R = 0.91), moderate RMSE (15.19 mm³/mm³), and a manageable positive bias (+10.51 mm³/mm³). SMAP and SMOS followed closely, with strong correlations (0.89 and 0.87, respectively) and the lowest RMSE values (~13.5 mm³/mm³). However, both showed a slight underestimation of soil moisture compared to COSMOS data. In contrast, AMSR-2 underperformed, with a lower correlation (R = 0.65), the highest RMSE (22.49 mm³/mm³) and a significant positive bias, which limits its reliability at this site.

In addition to standard statistical metrics, the study applied six different regression modelsLinear, Exponential, Logarithmic, Polynomial, Power, and Quantileto capture the underlying relationships between in-situ and satellite observations. Polynomial and Quantile regressions consistently provided better fits, especially for MERRA-2 and SMAP data. Quantile regression was particularly valuable in capturing tail behaviour, such as extreme wet or dry periods. It significantly improved the correlation for AMSR-2 (R = 0.951), suggesting it may still offer valuable information under specific conditions.

Despite being site-specific, this study demonstrates the critical value of ground validation, particularly through advanced ground instruments like COSMOS, in assessing the accuracy and applicability of global soil moisture products.

Future work should include spatial expansion to multiple stations across different climatic and land-cover zones in India, as well as seasonal performance analysis and the incorporation of advanced modelling techniques, such as machine learning or hybrid regression, to further improve downscaling and correction of satellite-derived soil moisture.

In summary, this work not only identifies the best-performing satellite products for Pune but also advances the methodological approach to soil moisture validation through the use of regression modelling. This study adds depth to the global effort of enhancing satellite remote sensing accuracy for local-scale hydrological applications.

References

Brocca, L. *et al*. (2010) Improving runoff prediction through the assimilation of the ASCAT soil moisture product. *Hydrol. Earth Syst. Sci.*, **14(10)**, 18811893.

Dorigo, W.A. *et al*. (2015) Evaluation of the ESA CCI soil moisture product using ground-based observations. *Remote Sensing of Environment*, **162**, 380395.

Entekhabi, D. *et al*. (2010) The soil moisture active passive (SMAP) mission. *Proceedings of the IEEE*, **98(5)**, 704716.

Kerr, Y.H. *et al*. (2010) The SMOS mission: New tool for monitoring key elements ofthe global water cycle. *Proceedings of the IEEE*, **98(5)**, 666687.

Kerr, Y.H. *et al*. (2012) The SMOS soil moisture retrieval algorithm. *IEEE Trans. Geosci. Remote Sensing*, **50(5)**, pp. 13841403.

Njoku, E.G. *et al*. (2005) Global survey and statistics of radio-frequency interference in AMSR-E land observations. *IEEE Trans. Geosci. Remote Sensing*, **43(5)**, 938947.

Reichle, R.H. *et al*. (2004). Global soil moisture from satellite observations, land surface models and ground data: Implications for data assimilation. *J. Hydrometeorol.*, **5(3)**, 430442.

Srivastava, P.K., Petropoulos G.P. and Kerr Y.H. (2016). Satellite soil moisture retrieval: techniques and applications. *Elsevier*.