Sentinel-1 high-resolution soil moisture product and validation

F. Mattia (1), A. Balenzano (1), G. Satalino (1), F. Lovergine (1), J. Peng (2&3), U. Wegmuller (4), M. Santoro (4), O. Cartus (4), K. Dabrowska-Zielinska (5), J. Musial (5), and M.W. J. Davidson (6)

(1) Consiglio Nazionale delle Ricerche (CNR), Ist. per il Rilev. Elettrom. dell'Ambiente (IREA), Italy
(2) Ludwig-Maximilians Universität München (LMU) - Department of Geography, Germany
(3) Max Planck Institute for Meteorology, Hamburg, Germany
(4) Gamma Remote Sensing Research and Consulting AG (GAMMA), Gümligen, Switzerland
(5) Institute of Geodesy and Cartography (IGiK) - Remote Sensing Centre, Poland
(6) ESA, ESTEC, 2201, Noordwijk, The Netherlands

ACKNOWLEDGEMENTS. This research is supported by the Scientific Exploitation of Operational Missions (SEOM) program of the European Space Agency, through the project "Exploitation of S-1 for Surface Soil Moisture Retrieval at High Resolution (Exploit-S-1)" contract n. 4000118762/16-NB (https://exploit-s-1.ba.issia.cnr.it).

Authors thank the teams providing soil moisture data of various hydrologic networks and the International Soil Moisture Networks for providing the REMEDHUS, HOBE and Elm Creek data.

Validation strategy for S-1 SSM product at 1km resolution

- Committee on Earth Observation Satellites (CEOS) - Working Group on Calibration and Validation (WGCV) – Land Product Validation (LPV) subgroup has established standards for validation procedures that are currently being used for satellite SSM product validation, which include:
  - **direct comparison of retrieved SSM values with independent measurements from <u>in situ networks</u> or satellite products**.

- For the low resolution SSM product validation, various **<u>experimental core sites</u>** have been identified or purposely set up in order to have multiple and representative SSM measurements within the SSM product pixels to estimate the average SSM with a confidence level (CL), e.g. at least 3 stations in 3km core site at 70%CL for 0.05m^3/m^3 SM uncertainty (Colliander at al., 2017).

- ESA Sentinel-1 observations to support systematic SSM product generation at regional/continental scale has risen the needs to tailor the validation strategy to high spatial resolution SM product (e.g 1km):
  - identify/set up **<u>1km core sites</u>** with adequate number of stations
  - alternatively, take into account the **<u>measurement sampling error (MSE)</u>** in the validation metrics (MSE bias the estimates of the slope and intercept in a linear regression, the correlation coefficient (Kelly, 2007; York et al., 2004) and affects the rmse)

Objective

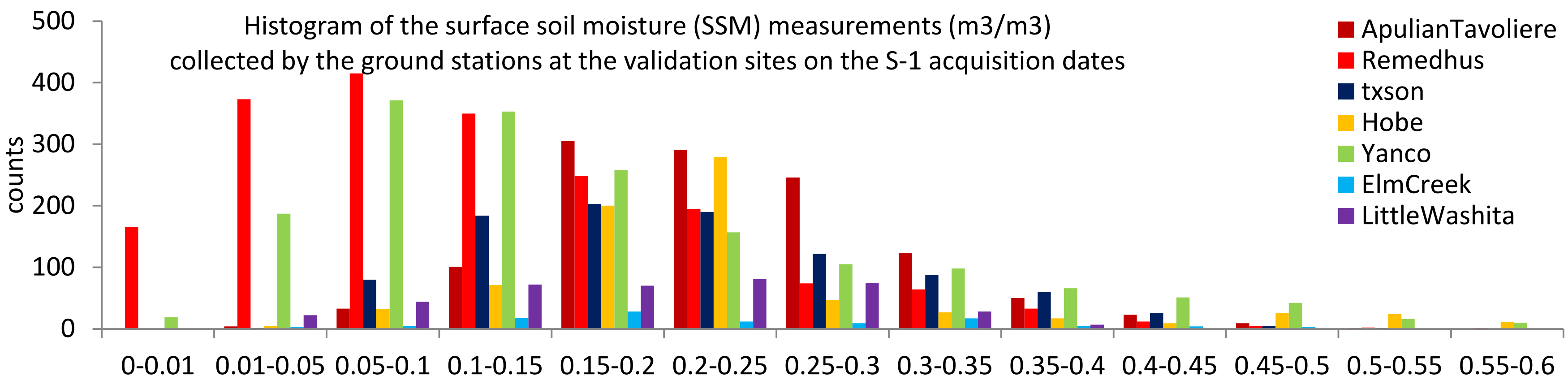
- Discuss a validation strategy for Sentinel-1 (S-1) Surface Soil Moisture (SSM) Product at 1km resolution

Outline:

- Experimental sites and ground&S-1 data
    - **"representativeness"** error (error made in assuming that the average of a limited number of point scale in situ measurements in the SM product pixel, corresponds to the true average SSM of the entire area) & its quantification over the validation sites
    - modelling the "representativeness" **error across scales**
- Retrieval Model Error Budget
- Validation results at 1 km resolution for a Short Term Change Detection Retrieval Algorithm
    - Example of the **impact of "representativeness" error at high resolution over a dense hydrological network in Southern Italy**
- Summary

Validation sites

Site	Lat., Lon (centre)	Climate zone (Köppen-Geiger climate classification)	Land cover (IGBP)	Hydrological network (extension and number of stations)	# S-1 IW images
Apulian Tavoliere (Italy)	41°22’13”N 15°29’23”E	Csa (warm temperature, summer dry, hot summer)	croplands	$2*2km^2$, 11 stations	119 (D124) Jan 2015- Jan2018 111 (A146) Jan2015 – Dec2017
Hobe (Denmark)	55°58'00"N 09°05'43"E	Dfb (Snow, fully humid, Warm Summer)	croplands forest	$30*30km^2$, 30 stations	58 (D139) Jan 2015- Feb 2017
Remedhus (Spain)	41°18'49"N 05°23'37"W	BSk (arid, steppe, cold arid)	croplands	$35*35km^2$, 20 stations	68 (D154) Jan 2015-Mar2017
Little Washita (USA, Oklahoma)	34°53’53”N 98°04’44”W	Cfa (warm temperature, fully humid, hot summer)	grasslands	$25*25km^2$, 20 stations	29 (A107) Apr 2016-Mar 2017
TxSON (USA, Texas)	30°18'26"N 98°46'13"W	Cfa (warm temperature, fully humid, hot summer)	grasslands	$36*36km^2$, 40 stations	21 (A107) Apr 2016- Feb 2017
Yanco (Australia)	34°50’39”S 146°07’13”E	BSk (arid, steppe, cold arid)	croplands grasslands	$60*60km^2$, 13 permanent (oznet) and 24 semi-permanet (SMAP network)	59 (D118) Jan2015 - Apr2017
Elm Creek (Canada)	49°40’48” N 98°0’23”W	Dfb (Snow, fully humid, warm summer)	croplands	$17*17 km^2$, 9 stations	15 (A136) Aug 2016 – Feb2017



Measurement sampling error (MSE) & its quantification

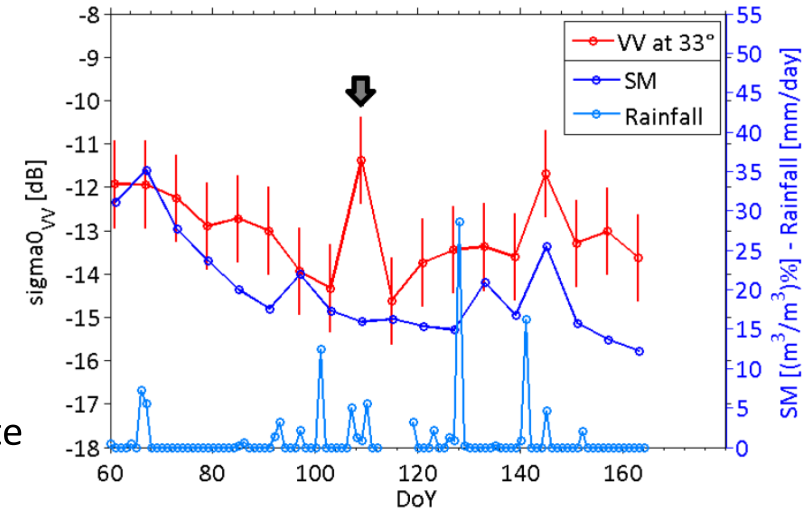
Modelling observations:

$$ssm_{obs} = (1/N)\sum_{i}^{N} ssm_{obs_i}$$

$$= ssm_{true} + \varepsilon_{MSE}$$

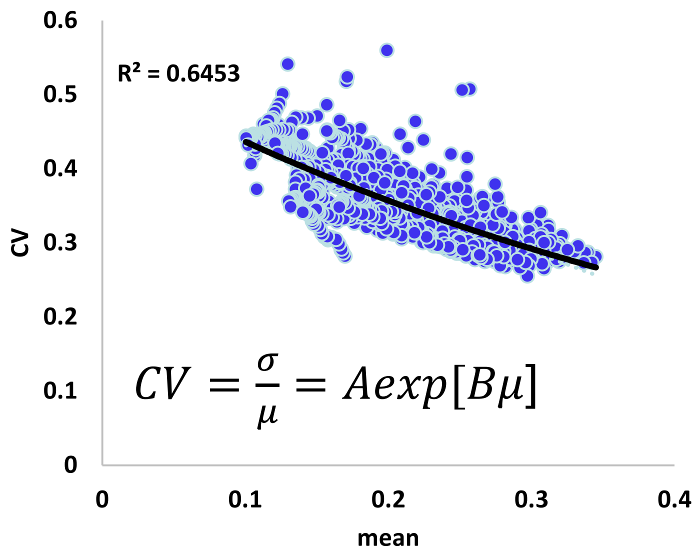
$$\langle(\varepsilon_{MSE})^2\rangle \xrightarrow{N \to \infty} 0$$

Example of temporal behaviour of backscatter at field scale (20ha) and SSM measured by ground station



Example of Coefficient of Variation (CV) at TxSON site

(Famiglietti et al., 2008, Crow et al., 2012..)



$$CV = \frac{\sigma}{\mu} = Aexp[B\mu]$$

$$MSE = z_{\alpha/2}\,\sigma(\mu)/\sqrt{N}$$

$z_{\alpha/2}$ is the standard normal variable at the chosen significance level α

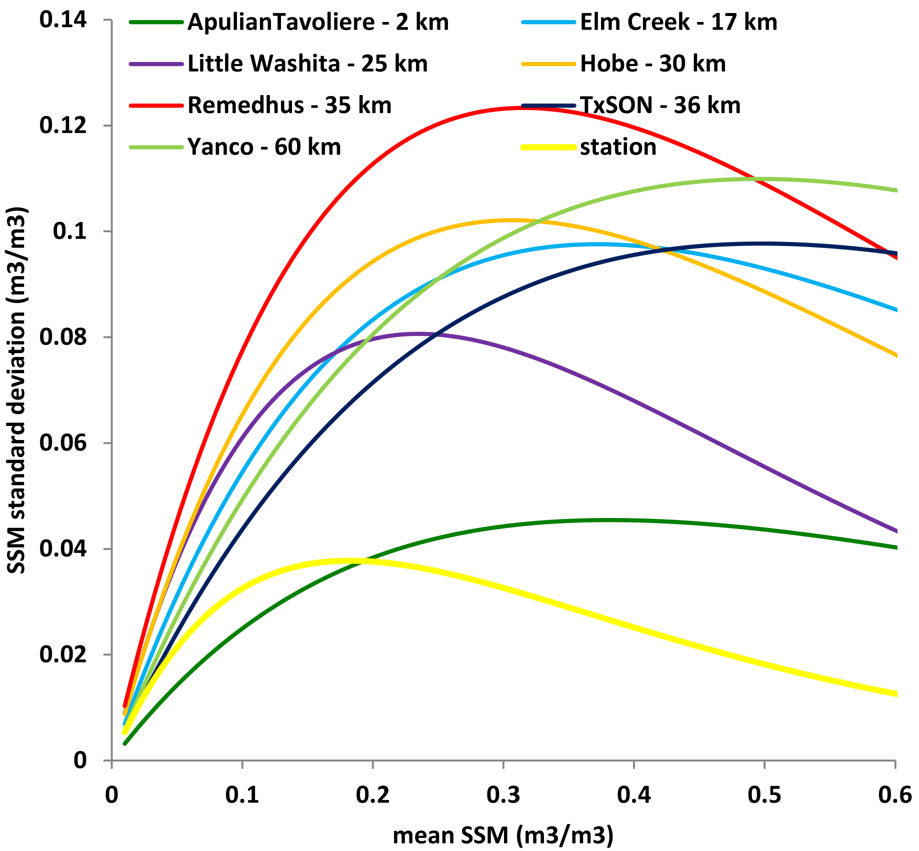
MSE provides an estimate (with a CL) of $\langle(\varepsilon_{MSE})^2\rangle$

CV allows to quantify the number of samples in relation to the mean SM and for a CL to achive a specified error

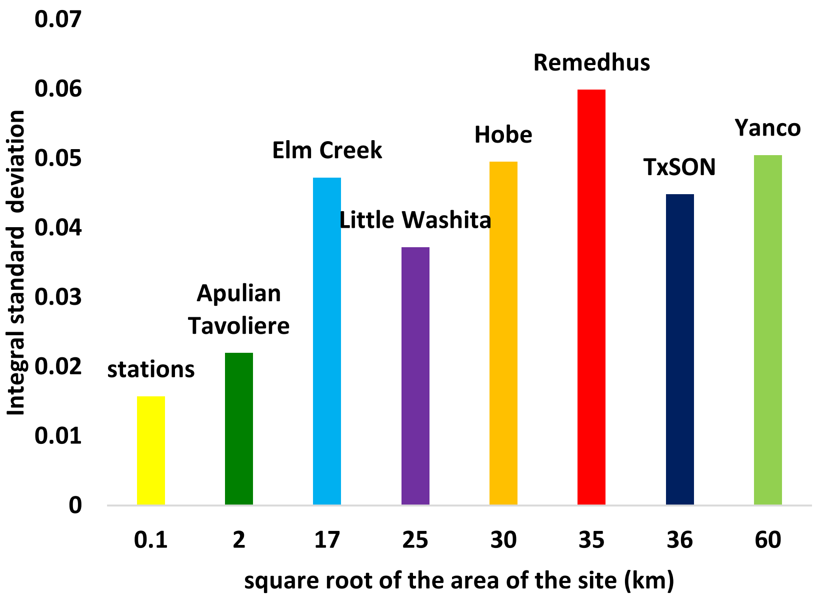
SSM variability across scales

Using SSM observations per date & site (ground stations)

$$\sigma = k_1 \cdot \mu \cdot e^{-k_2 \mu}$$



$$\Delta\sigma = k_1 \int_0^{\tau} \mu \cdot e^{-k_2 \mu} d\mu$$



Multi-scale geophysical parameters

$$\begin{cases} mean_{max} = \dfrac{1}{k_2} = \alpha_1 \, L^{\beta_1} \\ stdev_{max} = \dfrac{k_1}{k_2} = \alpha_2 \, L^{\beta_2} \end{cases}$$

modelling the "representativeness" error across scales

Famiglietti et. al, 2008

Table 3. Parameters of CV Versus Mea
and RMSE

Scale	k_1	k_2	$1/k_2$
2.5 m	0.7803	9.0607	0.1104
16 m	0.7287	7.3796	0.1355
100 m	0.8941	8.0774	0.1238
800 m	0.8840	5.8070	0.1722
1.6 km	1.2070	7.1128	0.1406
50 km	1.0429	5.2212	0.1915

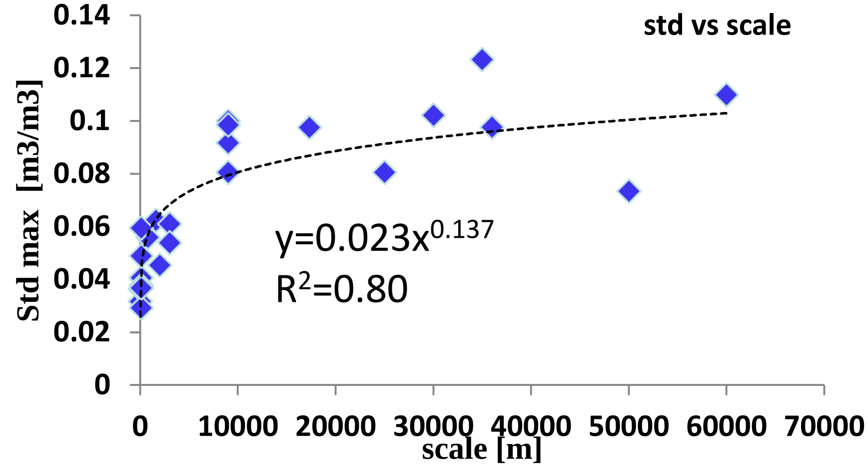
Jacobs et. al, 2008

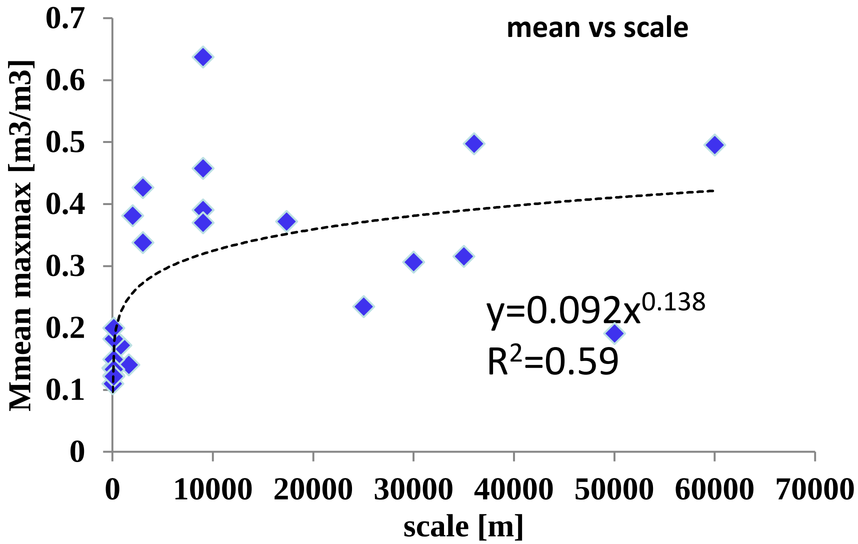
Table 2
Regression relationship between the coefficient of variation (CV) and the mean soil moisture content ($\bar{\theta}$) where $CV = A\ \exp[B\bar{\theta}]$

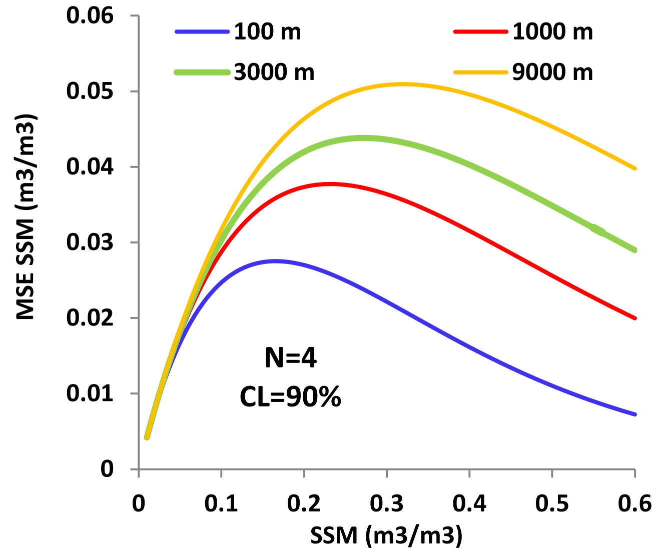
Field	A	B	R^2
WC11	0.890	−0.067	0.91
WC12	0.597	−0.075	0.75
WC13	1.327	−0.082	0.92
WC14	0.500	−0.050	0.89

SITE	SCALE (km)
Segezia	0.1, 2
Little Washita	25
Hobe	30
Remedhus	35
TxSON	3, 9, 36
Yanco	3, 9, 60

Total N=24







Retrieval Model error budget

Soil **MO**isture retrieval from multi-temporal **SAR** data (SMOSAR) prototype based on a **Short Term Change Detection Retrieval Model**

Approximations:

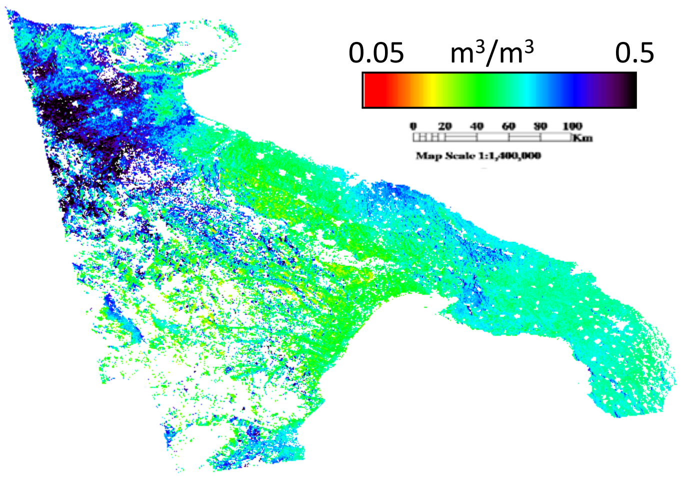
$$\sigma_0 \approx \tau^2 \cdot \sigma_0{}^{soil} \approx \tau^2 \cdot \left|\alpha_{pp}(\vartheta,\epsilon)\right|^2 \cdot F(\Omega,\vartheta,\lambda)$$

SAR response dominated by soil attenuated scattering

$$\frac{(\sigma_0)_{DoY(i+1)}}{(\sigma_0)_{DoY(i)}} \approx \frac{\left|\alpha_{pp}(\vartheta,\epsilon)\right|^2{}_{DoY(i+1)}}{\left|\alpha_{pp}(\vartheta,\epsilon)\right|^2{}_{DoY(i)}} \approx \frac{SSM_{DoY(i+1)}}{SSM_{DoY(i)}}$$

Backscatter temporal change depends solely on SSM

Example of 1km SSM product



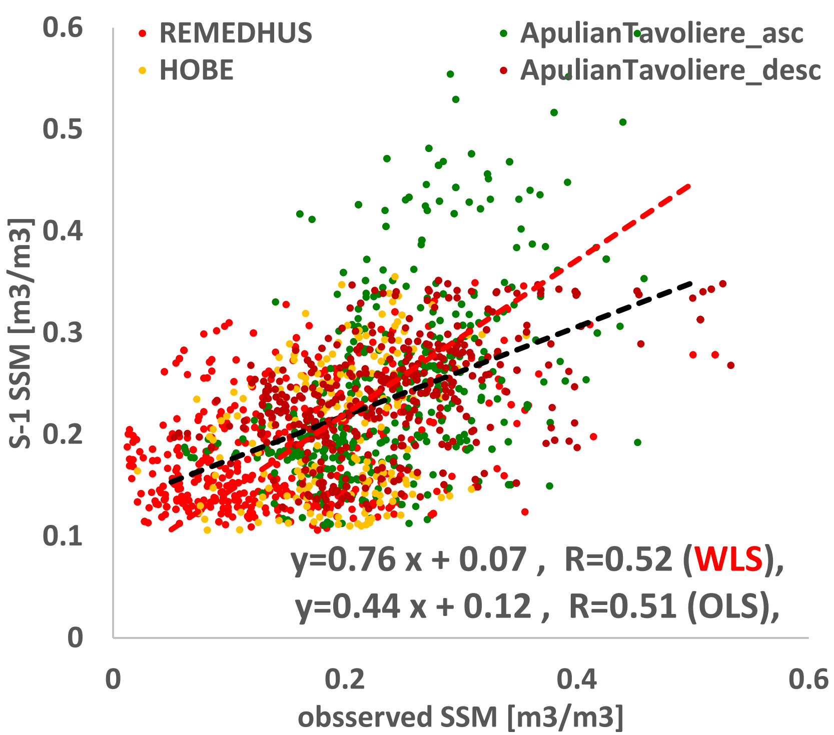
$$SSM_{retr} = SSM_{obs} + \varepsilon_{stat} + \varepsilon_{model} = SSM_{true} + \varepsilon_{MSE} + \varepsilon_{stat} + \varepsilon_{model}$$

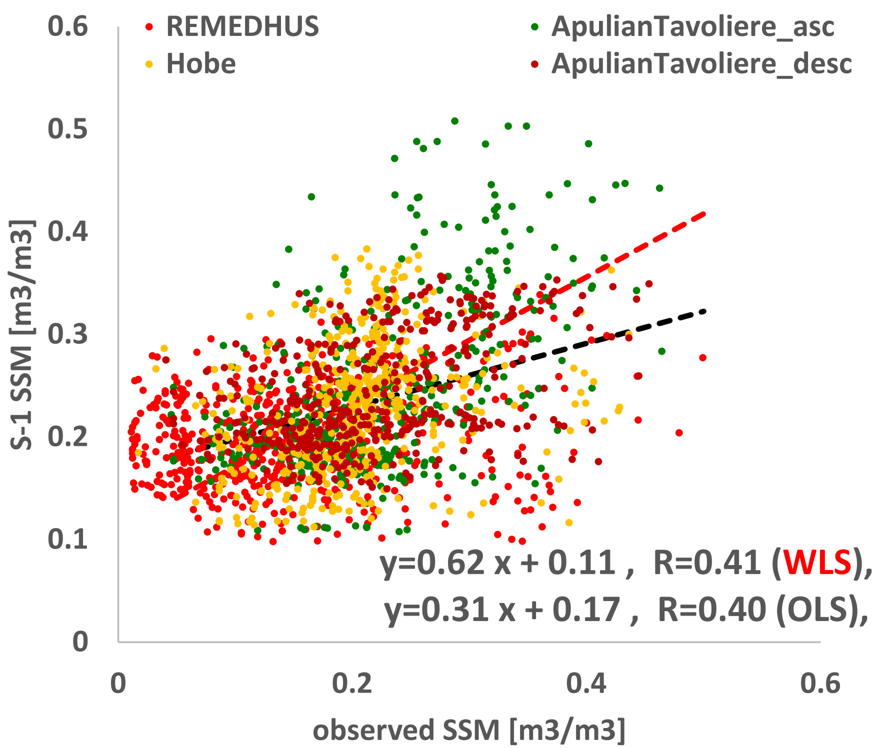
$$\boldsymbol{rmse^2} = <(SSM_{retr} - SSM_{obs})^2> =<(SSM_{retr} - SSM_{true})^2> + <\varepsilon_{MSE}{}^2> =$$

$$=<\varepsilon_{stat}{}^2> + <\varepsilon_{model}{}^2> + \boxed{<\varepsilon_{MSE}{}^2>}$$

$<\varepsilon_{stat}{}^2> \leq 0.03 m^3/m^3$ at 1km resolution with Number of looks >>100

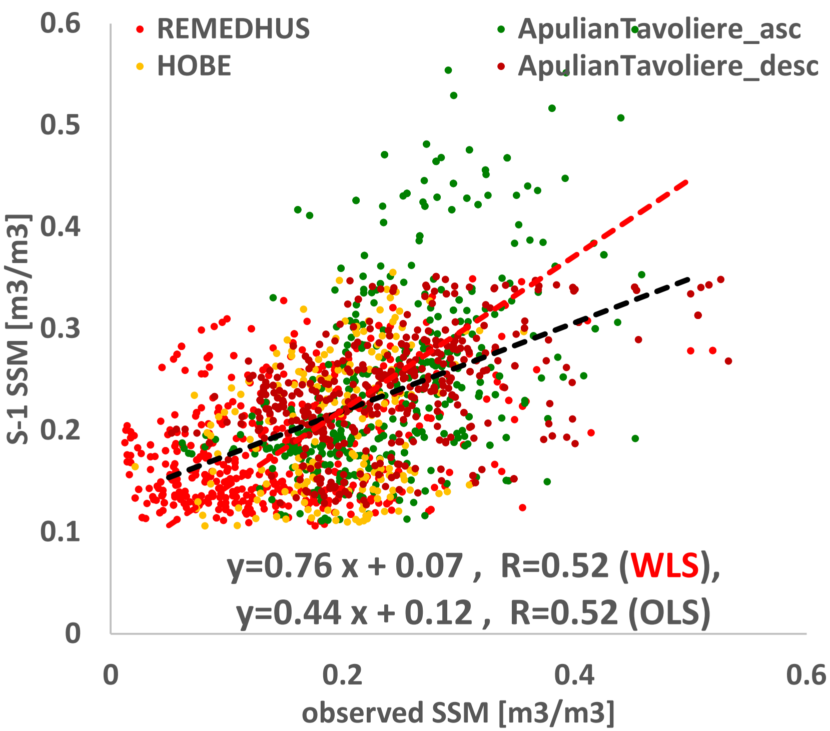
Retrieval error vs revisit time (6-days EU vs 12-days EU)

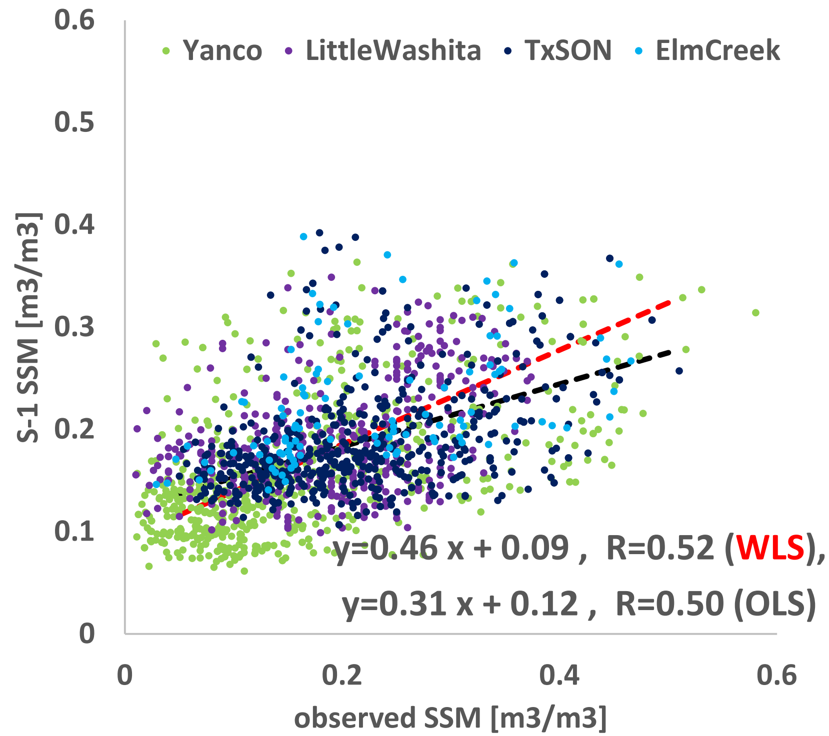




	N (random selected)	rmse (obs)	<obs. SSM>	<ret. SSM>	bias	MSE(90%conf.lev)	ubrmse (intrinsic)	rmse (intrinsic)	Outliers (3s)
		m3/m3	m3/m3	m3/m3	m3/m3	m3/m3	m3/m3	m3/m3	
6-days	1800	0.0838	0.2079	0.22223	-0.0144	0.0690	0.0453	**0.0475**	18
12-days	1800	0.0891	0.1970	0.2279	-0.030	0.0687	0.0477	**0.0568**	15

retrieval error vs revisit time (6-days vs 12-days no EU)



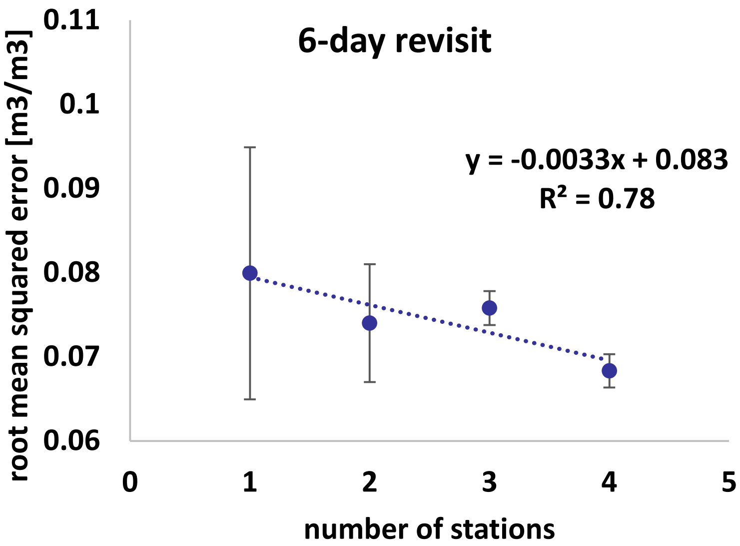


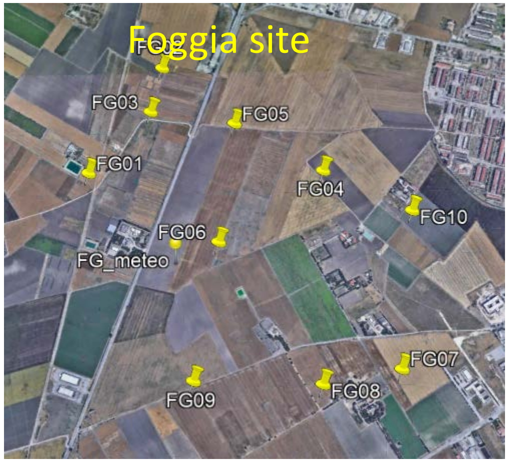
	N(random selected)	rmse (obs)	<obs. SSM>	<ret. SSM>	bias	MSE(90%conf.lev)	ubrmse (intrinsic)	rmse (intrinsic)	Outliers (3s)
		m3/m3	m3/m3	m3/m3	m3/m3	m3/m3	m3/m3	m3/m3	
6-days	1800	0.0838	0.2079	0.22223	-0.0144	0.0690	0.0453	**0.0475**	18
12-days	1800	0.888	0.1915	0.1806	-0.0376	**0.0659**	0.0585	**0.0595**	34

Apulian Tavoliere: impact of “representativeness” error at 1 km resol.

- 11 stations & spacing between stations ~500 m;
- in a 1 km res. cell up to 4 stations (example for S-1 A146, stations with longer temporal (**from 92 to 104 dates**) series were only considered

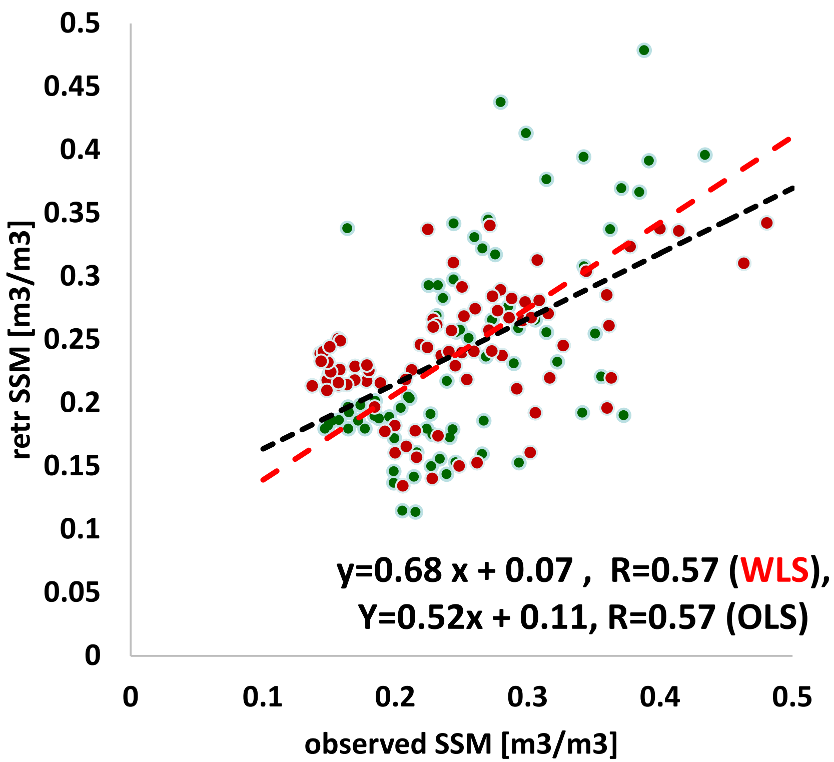
rmse (from S-1 A146 and D124) at 1 km vs number of stations (N)

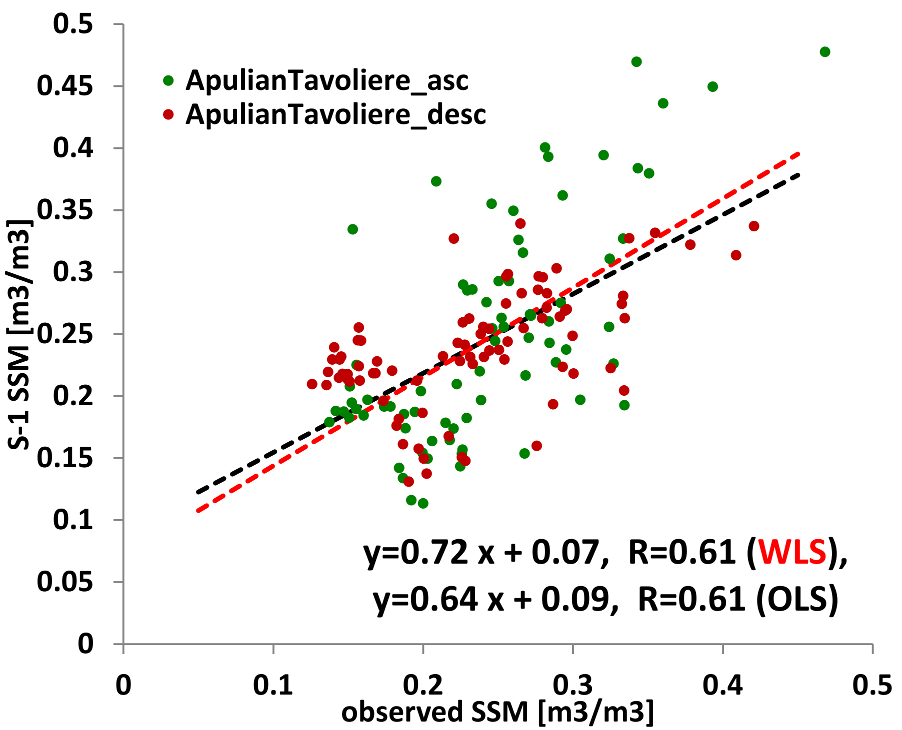




Additional 10 stations in 3km^2:
Calibration in progress (see poster section)

Validation at $1km^2$ (4 stations) and $2km^2$ (11 stations)





	N	rmse (obs)	<obs. SSM>	<ret. SSM>	bias	MSE(90%conf.lev)	ubrmse (intrinsic)	rmse (intrinsic)	Outliers
		m3/m3	m3/m3	m3/m3	m3/m3	m3/m3	m3/m3	m3/m3	
4 stations	158	0.0683	0.2508	0.2413	-0.001	0.036	0.0572	0.0564	4
11 stations	158	0.0605	0.2383	0.2429	-0.005	0.0169	0.0579	0.0581	1

Summary

- the impact of the measurement sampling error (MSE) across scales has been modelled & observed at 1 km scale
- In the validation of S-1 SSM products at 1 km the impact of MSE for N=1 cannot be disregarded → Core sites with at least 4 stations in 1km resolution are required (CL 90% and < 0.04 m^3/m^3 SSM uncertainty)
- Status of validation for S-1 SSM product at 1 km derived by a Short Term Change Detection SSM retrieval algorithm has been illustrated: the estimated retrieval error on long time series (6-days revisit) is approximately 0.06m^3/m^3