FRM4SM – Fiducial Reference Measurements for Soil Moisture

QA4SM Service Impact Assessment Report
DT3-2
Version 2.0

01 June 2023

Prepared by

Angewandte Wissenschaft, Software und Technologie GmbH

in cooperation with

Vienna University of Technology

CESBIO

Document Revision History:

Issue	Date	Details	Editor
0.1	2021-12-03	Draft document structure	A. Boresch
1.0	2022-05-03	Full version 1 for mid-term review	A. Boresch, W. Preimesberger, P. Stradiotti, M. Tercjak
2.0	2023-06-01	Updated version 2 for final review	A. Boresch, W. Preimesberger, P. Stradiotti, M. Tercjak

For any clarifications, please contact support@qa4sm.eu.

Table of Contents

List of Figures

List of Tables

Acronyms

API	Application Programming Interface
AWST	Angewandte Wissenschaft Software und Technologie GmbH
CCI	Climate Change Initiative
CDRs	Climate Data Records
CEOS	Committee on Earth Observation Satellites
CESBIO	Centre d'Etudes Spatiales de la Biosphère
CMUG	Climate Modelling User Group
DOI	Digital Object Identifier
ECMWF	European Centre for Medium Range Weather Forecasting
ECV	Essential Climate Variable
EODC	Earth Observation Data Centre
ESA	European Space Agency
FAQ	Frequently Asked Questions
FRM	Fiducial Reference Measurements
FRM4SM	Fiducial Reference Measurements for Soil Moisture
GCOS	Global Climate Observing System
GEO	Group on Earth Observation
GEWEX	Global Energy and Water Cycle Experiment
GUI	Graphical user interface
GTOS	Global Terrestrial Observing System
ISMN	International Soil Moisture Network
NASA	National Aeronautics and Space Administration
netCDF	Network Common Data Format
NRT	Near Real Time
QA	Quality Assurance
QA4SM	Quality Assurance for Soil Moisture
RFI	Radio Frequency Interference
SAG	Scientific Advisory Group
SAR	Synthetic Aperture Radar
SM	Soil Moisture
SMAP	Soil Moisture Active Passive
SMOS	Soil Moisture and Ocean Salinity
TC	Triple Collocation
TOPC	Terrestrial Observation Panel on Climate
TUW	TU Wien

1 Introduction

1.1 Purpose and scope of document

The QA4SM Impact Assessment Report describes how the QA4SM service is used, how it impacts the validation studies and activities in its predominantly scientific user community and which additional requirements and needs have been identified by the user community. The QA4SM Impact Assessment Report constitutes deliverable DT3.2 of the FRM4SM project.

While this report includes a brief overview of the current QA4SM service, more comprehensive information of all QA4SM features, supported validation types, options, datasets and use cases is described in the QA4SM Service User Manual, QA4SM_SUM v1.1.

1.2 Document overview

After the introduction to this document in Section 1, Section 2 provides a brief overview of the current QA4SM service and the development that has led to the current status. Section 3 assesses the technical and scientific capabilities of the QA4SM service and Section 4 discusses the scientific outcomes which have been enabled by QA4SM so far. In Section 5 we present statistics on QA4SM usage and service performance. Section 6 summarises the feedback received from the QA4SM user workshop in June 2022, from the SAG and from QA4SM users in general, along with identified gaps and recommendations for service improvement.

1.3 Target audience

This document is addressed to stakeholders at ESA and in the user community, specifically the Scientific Advisory Group (SAG) of the FRM4SM project. It also serves the development team to plan the future evolution of QA4SM with respect to identified community needs.

2 Overview of QA4SM service

Earth Observation (EO) from space allows us to create global Soil Moisture (SM) records to understand the effect of changes in global water availability on the environment. Many different EO satellites and retrieval models - for example ESA's SMOS mission was designed to measure SM over land - collect large amounts of data every day. Therefore, the need for rigorous, automated quality assessment procedures stands out among requirements by both the developers and users of EO SM data. While validation standards have been agreed upon in several scientific publications, their application often varies between independent studies; for instance, in terms of the defined reference measurements, validation metrics and in the presentation of the results. Together with the complexity of processing large data volumes for global validation efforts, this calls for unified tools to perform this task and provide standardised quality assessments. The Quality Assurance for Soil Moisture (QA4SM, https://qa4sm.eu) online service has been developed to bridge the gap between Analysis Ready Data (ARD) production and validation.

QA4SM is built on a powerful computing environment hosted at TU Wien, providing a virtual space where users can freely perform validations of the SM data included in the database. Available data range from single satellite missions (e.g., SMOS, SMAP) or state-of-the-art multi-sensor products (e.g., the European Space Agency Climate Change Initiative, ESA CCI SM or CGLS) to model or reanalysis data sets (e.g., ERA5). In addition, in-situ SM measurements provided by the International Soil Moisture Network (ISMN, ismn.earth) are included and regularly updated to allow global comparisons against up to 5 decades of continuous ground measurements. Every validation is personalized by the user through a web interface to apply spatial and temporal constraints or make use of advanced validation techniques, such as random error characterization of a data set through Triple Colocation Analysis (TCA). All methods included in the core algorithm of QA4SM are based on the best practices and requirements agreed upon by the Global Climate Observing System and the Committee on Earth Observation Satellites. The outcome of each validation, including graphical outputs and validation metric scores, can be stored, further processed or included in scientific studies and reports thanks to the use of traceable Digital Object Identifiers (DOIs).

The QA4SM service has originally been developed with funding from the Austrian Research Promotion Agency's Space Applications Programme (ASAP, FFG project numbers 866004, 878929). This original development established the base service and was then enhanced for the validation of high-resolution soil moisture data from the Copernicus Global Land Service (CGLS).

In the ESA FRM4SM project, the focus has moved to the validation of satellite data with respect to fiducial reference measurements (FRM) from in-situ stations. While the identification of FRM and the development of a pertinent validation protocol is an ongoing process, QA4SM provides the connection between satellite and in-situ data. To this end, the

integration between QA4SM and ISMN has further been deepened in the recent evolution of the QA4SM service, and ISMN metadata, which are relevant in the FRM4SM context have been integrated into QA4SM. Furthermore, the QA4SM features to assess and intercompare validation results have been enhanced and several SMOS and SMAP level 2 and 3 datasets have been integrated to support the validation of SMOS and SMAP soil moisture products in QA4SM.

Furthermore, a Scientific Advisory Group (SAG) with 10 representative members of the global scientific FRM and soil moisture data validation community has been established for the FRM4SM project. The SAG and other members of the QA4SM user community continue to be engaged to assess the QA4SM service and its potential impact on their scientific work in the individual contexts and to provide feedback on their requirements and needs beyond the current service capabilities. This feedback helps to steer the further evolution of QA4SM in order to maximise its impact in the user community, to successfully establish a broadly adopted standard for soil moisture data validation and to help making validation studies easily reproducible. A QA4SM user workshop to initiate this engagement was organised in June 2022 in colocation with the 6th Satellite Soil Moisture Validation and Application Workshop in Perugia, Italy.

Comprehensive information on the features, user options and use cases of the current QA4SM service is available in the QA4SM Service User Manual (QA4SM_SUM v1.1).

3 Assessment of QA4SM capabilities

The goal of QA4SM is to provide a standardised (following community agreed best practices) framework to soil moisture data users and producers to perform basic evaluations of different (satellite) data sets for a certain study respectively study area and period. QA4SM is a centralized platform to simplify the process of data evaluation for users by taking away the need to download data and develop or install software to manually calculate performance metrics and create visualization of results. QA4SM provides traceable validation results in terms of input data, the validation process itself and its output.

A detailed description of technical features in QA4SM, how to use them, and a description of their scientific background is available in the QA4SM Software User Manual (QA4SM SUM). This includes the input data pre-processing, filtering and matching, methods for scaling (bias removal) and anomaly calculation, as well as an overview of validation metrics that are computed in the service. We therefore only provide a condensed overview in this section and assess where QA4SM is currently standing with regard to the above-described goal.

Server-side processing: QA4SM solves the need of client-side software installation by providing a server based online application. Maintaining traceability in terms of software and input data requires an isolated processing environment where the service provider is in control of any amendments. Only in this way it is possible to keep track of and avoid potential changes that would break the traceability and reproducibility of validation results. In addition, a server-side application is easy to use and minimizes prerequisites on the user side (supported operating systems, software updates and dependencies), respectively the entry barrier for potential new users in general.

Online data archive: In line with the previous point, it is required to provide data to users of the online service. The current selection of soil moisture data in QA4SM (overview in Table 1) is a set of popular and therefore most likely relevant data sets for new users (based on past soil moisture validation studies) and based on the specific requirements by users of QA4SM in the past. However, as the number of newly created datasets is growing, so does the need for more flexibility in terms of available up-to-date data in the service. This can be addressed by rigorous data management and further automation and standardisation of the data import process (also see the next paragraph in this regard). This point applies in particular to datasets which are uploaded by users for validation against other QA4SM datasets.

Standardised pre-processing: Soil moisture data is available in various formats. General data standards are developed foremost by the netCDF Climate and Forecast (CF) Metadata Conventions (Eaton et al., 2021) and specifically adapted by the soil moisture community. Still, in available datasets, there are often deviations from these target standards, which raises the need for QA4SM to implement data specific pre-processing routines to harmonize products before they can be imported into the validation framework. This has been done for the

currently implemented datasets. The complexity of this task varies for each case and is currently minimal for datasets that follow the above-described CF conventions. For datasets that fully follow CF standards, automation of the inclusion of updated versions is a next planned step.

Dataset Name	Dataset Version
ESA CCI Soil Moisture (COMBINED)	v04.4, v04.5, v04.7, v05.2, v06.1
ESA CCI Soil Moisture (ACTIVE)	v04.5, v05.2, v06.1
ESA CCI Soil Moisture (PASSIVE)	v04.5, v05.2, v06.1
C3S Soil Moisture (COMBINED)	v201706, v201812, v201912, v202012
ERA5	accessed: 2019-06-13
ERA5-Land	accessed: 2019-09-04
GLDAS-Noah	v2.1
HSAF ASCAT SSM	H113
ISMN	Accessed: 2019-12-11, 2021-01-31, 2023-01-10
SMAP SPL3SMP	v5 (PM/asc.), v6 (PM/asc.), v5 (AM/desc.), v6 (AM/desc.)
SMAP SPL2SMP	v8
SMOS-IC	V.105 (asc.)
SMOS L3	v339 Descending, v339 Ascending
SMOS L2	v700
CGLS SSM 1km	V1.1
CGLS SWI 1km	V1.0

Table 1: Overview datasets and versions in QA4SM

User Data Upload: Starting with QA4SM version 2.2.0, users can now upload and validate their own data. To ensure that data is properly formatted for this purpose, users can access the relevant guidelines for data preparation https://qa4sm.eu/ui/user-data-guidelines. As underlined in the guidelines, files are accepted in .zip, .csv or netCDF format. When user data has been uploaded, users can select their datasets for validation in the same manner as they select other data available in QA4SM. Currently, user data is accessible only to the owners of the data, i.e., the users who uploaded the data. While sharing results of a validation that uses

user uploaded data is possible, other users cannot use the data for their own validations in QA4SM. Furthermore, validations based on user uploaded data cannot be published with the QA4SM publication service. Users can remove their uploaded data again, as long as the data has not yet been used in a validation. In this case, the validation first needs to be deleted.

Standardised metrics: A set of "standard" validation metrics is agreed upon by the soil moisture community and described in the best-practice guidelines (Gruber et al., 2020). QA4SM implements them (see section 4.1 in the QA4SM Software User Manual) in a way that makes it relatively easy to add new relevant metrics that the community may agree upon in the future. QA4SM enables the intercomparison of 2 to 6 datasets with pairwise and triple-wise metrics with confidence intervals. The implemented intercomparison approach uses only common time stamps after temporal matching between all datasets for the metric calculation, making the comparison as fair as possible and reducing the impact of, e.g., differently stringent levels of flagging or data coverage between the compared data sets.

Standardised outputs and postprocessing: QA4SM creates a validation summary (summarizing and storing the settings used in a validation run) and provides validation results in netCDF format for all selected datasets. This enables the subsequent automatic computation of statistics and the creation of result visualizations, respectively the comparison of validation results between different validations performed with QA4SM. By downloading the netCDF files generated in a validation run, users can create and apply their own post-processing routines to further extract specific validation results for their study, for example by sub-setting based on more complex shapes or removing outliers.

Results management and publishing: QA4SM provides a wide range of usability features in order to manage the potentially large number of validation results assigned to a user account. This includes options to browse and filter results, archive (to exclude them from automatic clean-up), delete and share them with other users (by URL), copy and track results from other users (i.e., add them to the user's own collection) and publish them. When a validation result is published, results are copied to an external service (https://zenodo.org/), where they are archived under the user's name and are both downloadable and citable through a unique DOI without any access restrictions. For traceability reasons, validations involving user uploaded data are currently excluded from publication, as users can always remove their uploaded data from QA4SM again.

Open source development: The development of QA4SM is openly accessible in the sense that the source code is public and all development changes are fully tracked in a public repository: https://github.com/awst-austria/qa4sm. Anyone can examine, fork and contribute to the repository. The pytesmo backend for metric calculation is publicly available in https://github.com/TUW-GEO/pytesmo.

Community Feedback: The solid technical basis of QA4SM is recognized by the SM community as this quote from Bayat et al. (2021) shows:

> *"**QA4SM is perhaps the most advanced online validation platform** compared to the other ECVs investigated in this study. However, there is certainly room for improvement in QA4SM. Particularly, there is a **need to include more satellite SM products**, especially those with global coverage and longer time-series [...]. It **should also have the ability to implement user-generated SM products** in the near future. This can provide end-users with an opportunity to compare a local or regional product of their own [...]. Of course, it should be noted that **all new additions to the platform, either from satellites or in situ, should pass the CEOS LPV SM protocol requirements.**"*

This statement generally matches the feedback received from users, namely, that the primary user interest is to work with additional data in QA4SM, both additional featured datasets and their own data. This has already been addressed to a certain degree in QA4SM release 2, which was deployed in March 2023, where additional SMOS and SMAP level 2 datasets have been integrated and, most prominently, support for user-uploaded datasets has been added.

Users can now upload up to 5 GB of data in a variety of formats, including .zip, .csv, and netCDF, and validate these data against all featured datasets in QA4SM. We continuously update our list of available datasets and welcome suggestions from users on which datasets to add next. And online request and feedback form to ask for additional datasets to be added and to provide general feedback would certainly be helpful to facilitate the request process for interested users. A further plan is to update and extend existing QA4SM datasets in regular intervals and to automate such updates via automated download, preprocessing and QC validation processes where download APIs are available from the data providers.

4 Scientific outcomes enabled by QA4SM

QA4SM is openly and freely accessible and is therefore at the disposal of a wide range of users and applications. The citation rules of QA4SM are so far fairly loose as no official QA4SM publication exists. Its use can be generally be referenced with a direct link to the online website or, in the case of specific validation results which were published with QA4SM, with the DOI's that are issued for these published results. A stricter citation policy could help keeping track of the employment and evaluation of QA4SM in scientific studies.

Although the specific aim of the service is to perform validations of soil moisture data sets, the context of applications varies greatly. Keeping track of the different subject areas where QA4SM is used is fundamental to meeting the needs of the user community. Table 2 gives an overview of the publications and documents where QA4SM was cited[1]. Figure 1 classifies the QA4SM citations by subject area, where the following four areas are identified based on the current citations pool:

- *Data set Paper*: publication that presents a novel soil moisture data set (e.g., an *in-situ* Network) or processing method.
- *Hydrology and Land Surface Modeling:* publication that presents methods in modeling and assessing the water cycle (e.g., data assimilation, drought analysis).
- *Validation Practices*: publication that presents or evaluates methods for satellite or model soil moisture data sets validation.
- *High-Resolution Applications*: publication that presents novel methods or data sets assessment of high-resolution soil moisture products.

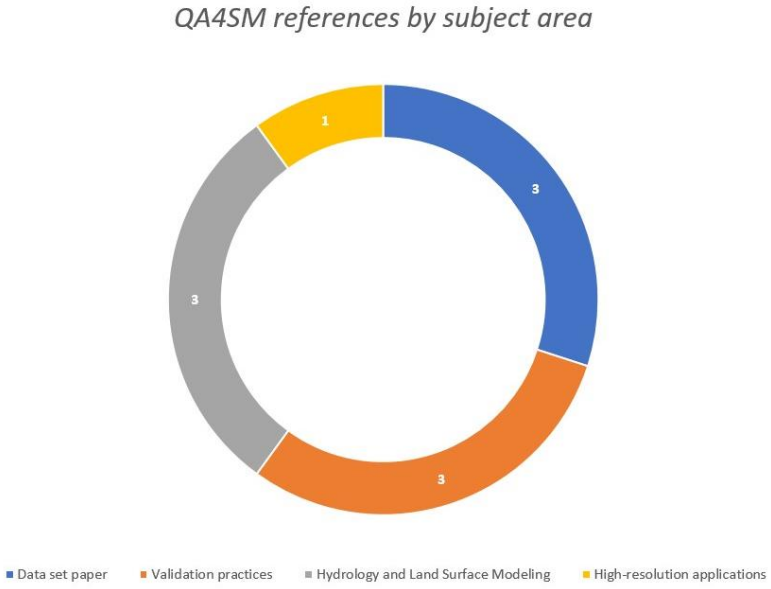

Figure 1: Classification of publications and documents citing QA4SM based on subject area.

[1] generally through the website link, originally https://qa4sm.eodc.eu , since mid-2022 https://qa4sm.eu.

Used	Subject Area	Datasets used	Authors	Title	Journal	DOI	Year
YES	Data set paper	ISMN, ERA5, ESA CCI SM Combined	Wouter Dorigo et al.	The International Soil Moisture Network: serving Earth system science for over a decade	Hydrology and Earth System Sciences	n.a.	2021
NO	Validation practices	n.a.	Gruber et al.	Validation practices for satellite soil moisture retrievals: What are (the) errors?	Remote Sensing of Environment	n.a.	2019
NO	Hydrology and Land Surface Modeling	n.a.	Zhao, H.; Montzka, C.; Baatz, R.; Vereecken, H.; Franssen, H.-J.H.	The Importance of Subsurface Processes in Land Surface Modeling over a Temperate Region: An Analysis with SMAP, Cosmic Ray Neutron Sensing and Triple Collocation Analysis	remote sensing	https://doi.org/10.3390/rs13163068	2021
NO	Validation practices	n.a.	B. Bayat, F. Camacho, J. Nickeson, M. Cosh, J. Bolten, H. Vereecken, C. Montzka,	Toward operational validation systems for global satellite-based terrestrial essential climate variables	International Journal of Applied Earth Observation and Geoinformation	https://doi.org/10.1016/j.jag.2020.102240	2021
NO	High-resolution applications	n.a.	Jian Peng, et al.	A roadmap for high-resolution satellite soil moisture applications – confronting product characteristics with user requirements	Remote Sensing of Environment	n.a.	2021
NO	Data set paper	n.a.	Cosh, Michael H., et al.	Developing a strategy for the national coordinated soil moisture monitoring network	Vadose Zone Journal	n.a.	2021
YES	Hydrology and Land Surface Modeling	GLDAS, ESA CCI SM Combined	De Santis, D., et al.	Assimilation of Satellite Soil Moisture Products for River Flow Prediction: An Extensive Experiment in Over 700 Catchments Throughout Europe	Water Resources Research	https://doi.org/10.1029/2021WR029643	2021
NO	Validation practices	n.a.	Montzka, Carsten, et al	Soil moisture product validation good practices protocol	n.a.	n.a.	2021
YES	Data set paper	n.a.	W Preimesberger et al.	Homogenization of Structural Breaks in the Global ESA CCI Soil Moisture Multisatellite Climate Data Record	IEEE Transactions on Geoscience and Remote Sensing	10.1109/TGRS.2020.3012896	2021
YES	Hydrology and Land Surface Modeling	ESA CCI SM Combined, ISMN	Chevuturi, A., et al.	Improving global hydrological simulations through bias-correction and multi-model blending	Journal of Hydrology	doi.org/10.1016/j.jhydrol.2023.129607	2023

Table 2: References to QA4SM in publications and documents

A total of ten references to QA4SM were found in online research (keywords: QA4SM, QA4SM validation) on various search engines. Two of these references (Gruber et al., 2019 and Cosh et al., 2020) specify the validation standards which are implemented in QA4SM and cite QA4SM as the recipient of the validation good practices studies and protocols. In addition to these, a third paper, citing QA4SM in a comparison of validation protocols, finds that QA4SM is more suitable than other alternatives that were considered (Bayat et al., 2021). Only two of the considered papers so far made direct use of the QA4SM platform, in one case in the context of a data set presentation (Dorigo et al., 2021), in the other case while evaluating assimilation methods (De Santis et al., 2021). Judging from the reference to QA4SM in the diverse application contexts and by different investigation groups, the platform is fairly accepted in the community and regarded as an instrumental standard validation tool. However, the ambition for QA4SM to become a transversal reference tool across the community is not reflected in the current scientific usage. This can be asserted by considering that the total number of publications in satellite SM validation is more than 2800 from 2023 only (research on Google Scholar on 19th May, 2023 with keywords: satellite soil moisture validation). Hence, the actual use and integration of QA4SM results in publications should be further encouraged in the project.

Beyond formally published scientific articles, QA4SM is already used more routinely during data products development and in project reports. It should be noted that tracing and quantifying this use is made difficult by the fact that these validations are not necessarily connected to a DOI or publicly released but often are part of internal assessments. For instance, QA4SM is used on a regular basis for the generation of reports for the Copernicus Climate Change Service[2] (C3S) and the ESA Climate Change Initiative[3] (ESA CCI), respectively to evaluate the stability of Interim Climate Data Record in an operative fashion and to systematically and regularly evaluate different product versions in a standardised way.

Figure 2 gives an overview of the QA4SM validations that have been published and associated to a DOI in Zenodo (based on Appendix A). The two most referenced data sets are ISMN and ESA CCI Combined. Both have been object of extensive validation in the analysis carried out by Dorigo et al. (2021). ESA CCI Combined validations have also been published trough a Zenodo DOI in the internal project validations, as well as C3S and ESA CCI SM Active. ERA5 (Land) and ISMN are also regarded as standards in the satellite soil moisture validation, hence they stand out among the other data sets. It should be noted that 2021 and 2022 saw a decrease of the DOI issuing compared to 2020; the reason can be traced to the activity carried out in 2020 in the context of the production of Dorigo et al. (2021). It should be noted that

[2] Public Product Quality Assessment reports at: https://cds.climate.copernicus.eu/cdsapp#!/dataset/satellite-soil-moisture?tab=doc

[3] Public Product Validation and Inter-Comparison reports at: https://climate.esa.int/en/projects/soil-moisture/key-documents/

none of the DOIs issued through Zenodo refer to the studies in Table 2, since the publication of validation results produced within studies is not mandatory. Similarly, as for the citation of the QA4SM service, a stricter referencing policy with respect to the DOI issuing could reflect in an increase of the numbers presented here.

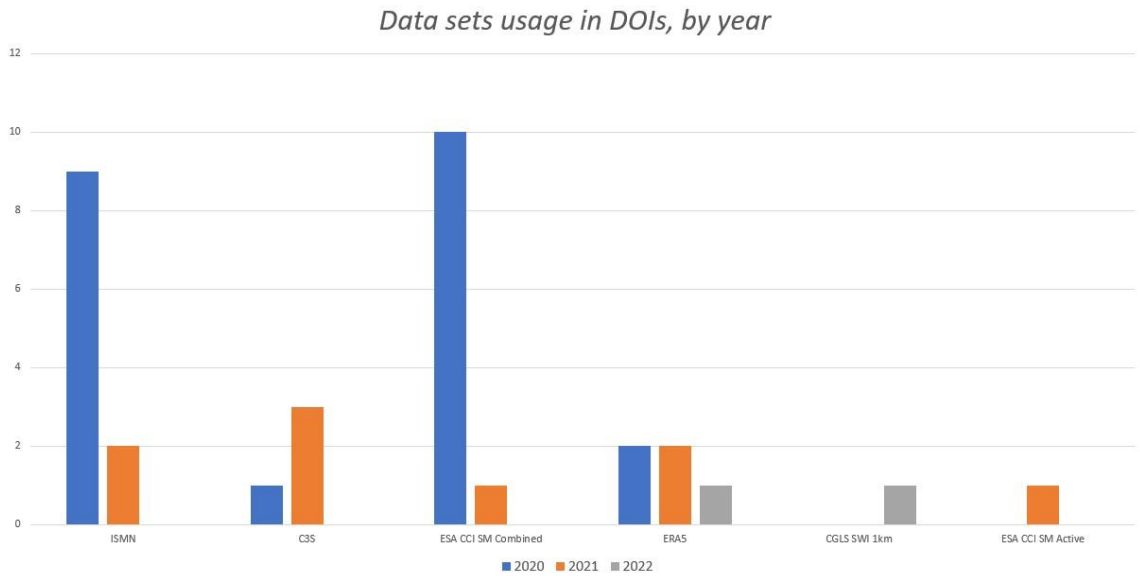

Figure 2: DOI references in Zenodo by year and data set, based on a query made on May 19th, 2023 (Appendix A).

5 QA4SM usage and performance assessment

5.1 QA4SM users

By the time of writing, 176 external users from 32 countries were registered in the QA4SM service. Nearly all users are from scientific institutions or international organisations including DLR, ECMWF, ESA and EUMETSAT. Table 3 shows the distribution of users by country. The aggregated numbers by geographic continent are shown in Table 4. Perhaps not surprisingly, most QA4SM users currently come from Europe, but second is a strong user group from Asia, followed by a lower number from America and a small but noticeable user group from Africa.

Country	Number of Users	Country (continued)	Number of Users (continued)
Argentina	1	Japan	2
Armenia	1	Luxembourg	1
Australia	1	Netherlands	4
Austria	19	Nigeria	1
Brazil	1	Poland	3
China	19	Romania	1
Czechia	1	Spain	5
Egypt	1	Sudan	1
Ethiopia	2	Sweden	2
Finland	1	Switzerland	4
France	11	Taiwan	1
Germany	9	Uganda	1
Ghana	1	Ukraine	1
Greece	4	United Kingdom	6
India	7	United States of America	15
Iran	4	N/A	40
Italy	5	**Total**	**176**

Table 3: QA4SM users by country

Continent	Number of countries with registered users	Number of Users
Africa	6	7
America	3	17
Australia	1	1
Asia	6	34
Europe	16	77
N/A (country not specified)		40
Total	**32**	**176**

Table 4: QA4SM users by geographic continent

5.2 Usage of datasets in QA4SM validations

A total of 753 validations of soil moisture datasets have been run in the two-year period of the FRM4SM. 13 validations were run for user uploaded data after the upload feature had been added in release 2 in March 2023. While most users have typically run a few individual validations, a user in China has used QA4SM for more than 130 validations of various datasets.

Figure 3 shows how frequently each available dataset has been used in individual validations so far. Note that each validation includes two or more datasets which are selected by the user. In this way, QA4SM supports the validation of one selected dataset against a reference dataset as well as the intercomparison of several datasets among each other. Shown on the horizontal axis are the 16 integrated datasets which are currently available in the public QA4SM service. Different versions of datasets are indicated by different colours in the diagram. Additionally, all (currently six) user uploaded datasets are shown in a single bar on the right side of the diagram. Overall, ISMN is the most used dataset, as it is the default reference dataset for QA4SM validations and includes the FRM data, which are identified in the FRM Protocols and Procedures for Soil Moisture Products (FPP_SM) document.

Figure 4 shows the validations run per calendar month. There was an increase after the first few months, followed by a phase of fairly constant usage and lower user activity. The peak usage in December 2021 is partially influenced by increased internal testing activity related to the ongoing service evolution. The validation activities performed for release 1 and 2, though, have been performed on a separate test services instance of QA4SM, hence they are not reflected in the diagram. A second activity peak occurred around the QA4SM user workshop, which was conducted in June 2022, followed again by lower user activity during evolution towards QA4SM release 2. At the time of writing, SAG members and key users are engaged to try out QA4SM with their own uploaded datasets, so we expect the user activity to increase again with the new release 2 features being available, as can already be seen for May 2023.

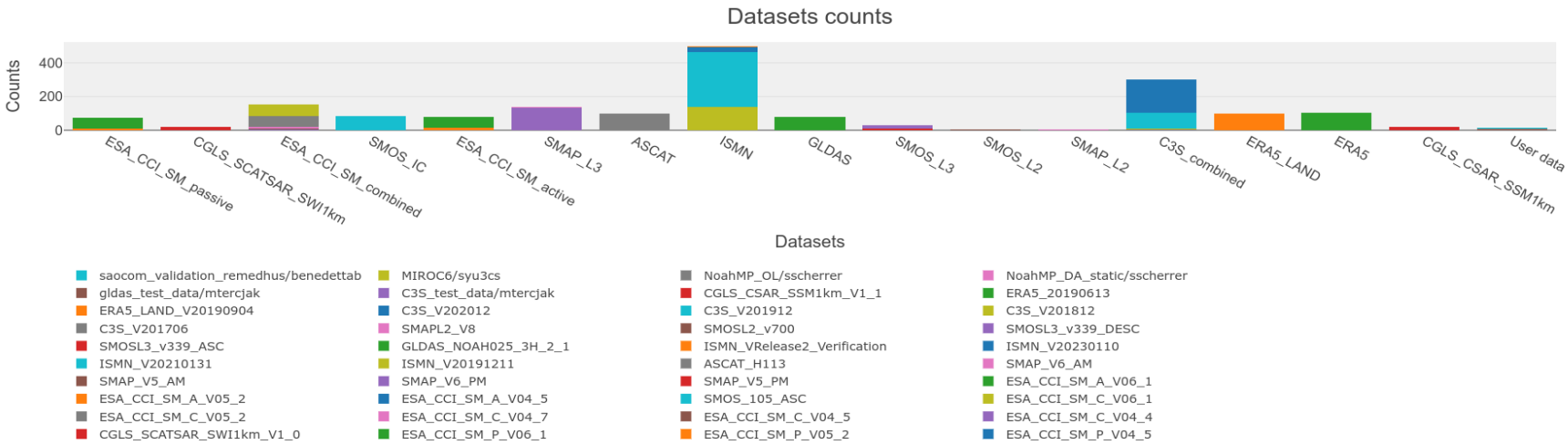

Figure 3: Distribution of datasets used in validations performed by QA4SM users

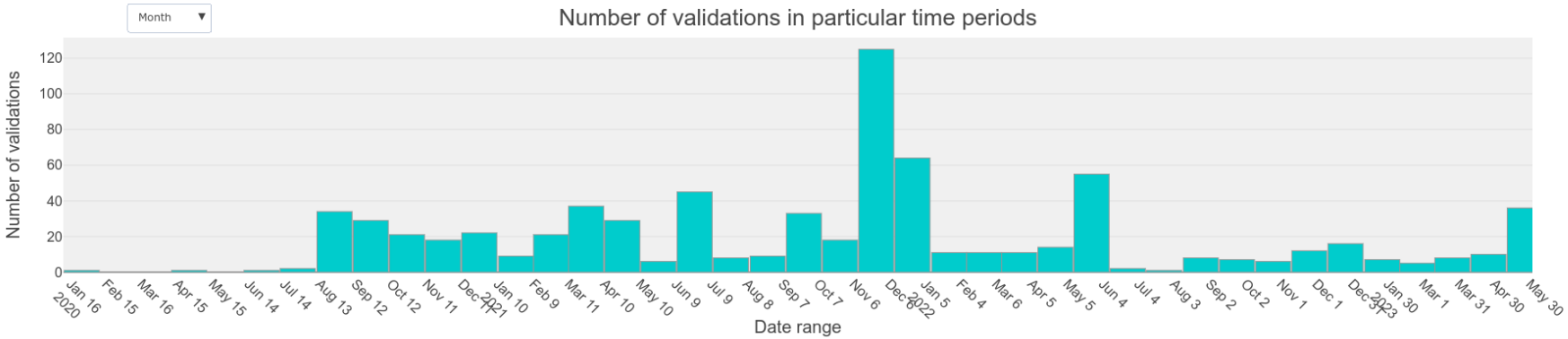

Figure 4: Numbers of QA4SM validations run per calendar month

5.3 QA4SM service availability and key performance indicators

Three key performance indicators have been defined to monitor the performance of the QA4SM service in the FRM4SM project. These include indicators of the annual service availability, the daily down time and the response time for user requests. In particular:

- The QA4SM service shall be available for a cumulated time of 99% per year.
- The daily down time, due to software or hardware failures shall not exceed four hours.
- Initial answers to users accessing the email help desk at support@qa4sm.eu shall be provided within three working days.

5.3.1 QA4SM availability

Figure 5 shows the monthly availability of the QA4SM service in the first year of the FRM4SM project. Until September 2021, QA4SM was hosted in a cloud infrastructure provided by EODC, which had a relevant infrastructure outage at the beginning of the project. In October 2021, QA4SM was migrated to a new and dedicated server at TU Wien, which has reliably been available for the remaining project period. The service availability has been above 99% in most months. The service migration was done seamlessly, with a brief downtime of one hour, and did not impact the service availability. The cumulated service availability was 99,4% in the first year and 99,6% in the second year of the project.

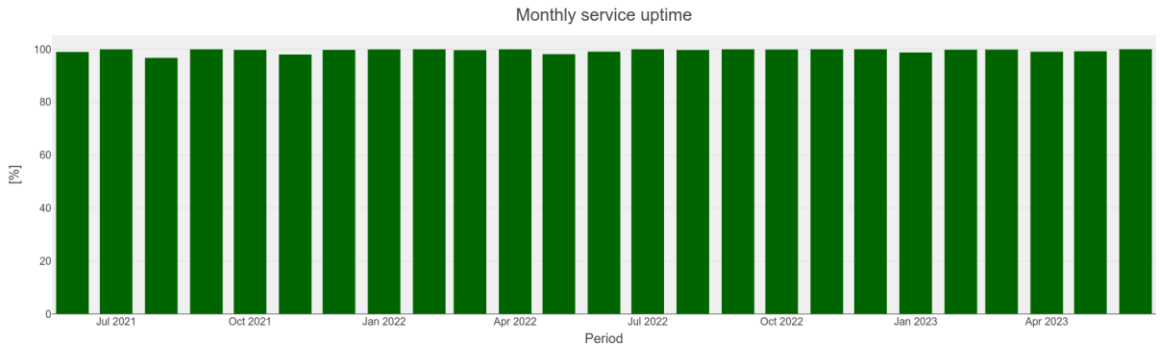

Figure 5: Monthly availability of QA4SM service

5.3.2 QA4SM service downtime

Figure 6 shows the cumulated monthly downtimes of the QA4SM service per calendar months. The typical downtime was below one hour in most month, while the maximum monthly downtime was about 14 hours in November 2021.

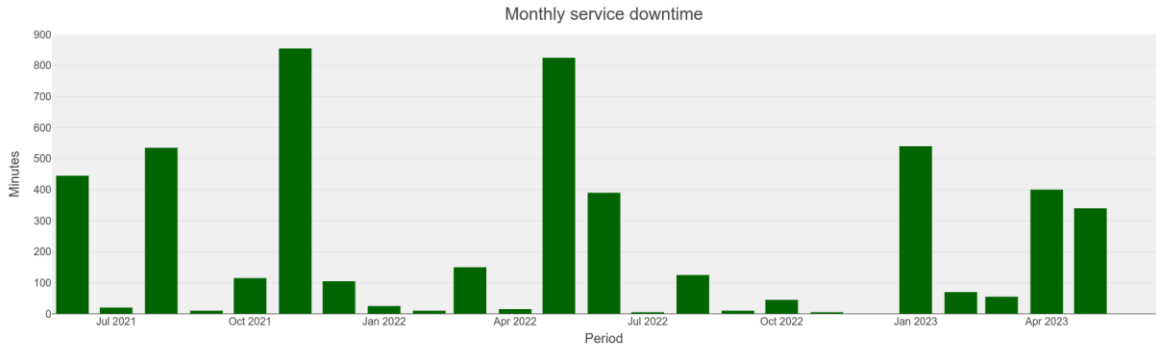

Figure 6: Monthly downtimes of QA4SM service

Figure 7 shows the daily downtimes. The four-hour threshold was exceeded on seven days in the two-year project period with a maximum downtime of about 12 hours on one day in May 2022. These longer outages were typically related to temporary non-availability, planned upgrades or technical issues of the underlying computer infrastructure. One case was attributed to an expired web certificate and another case was related to setup issues of software containers used to deploy the QA4SM service. In these cases, measures have been taken to avoid re-occurrence of the same issue again. E.g., notifications have been set up for the renewal of web certificates before their expiry date.

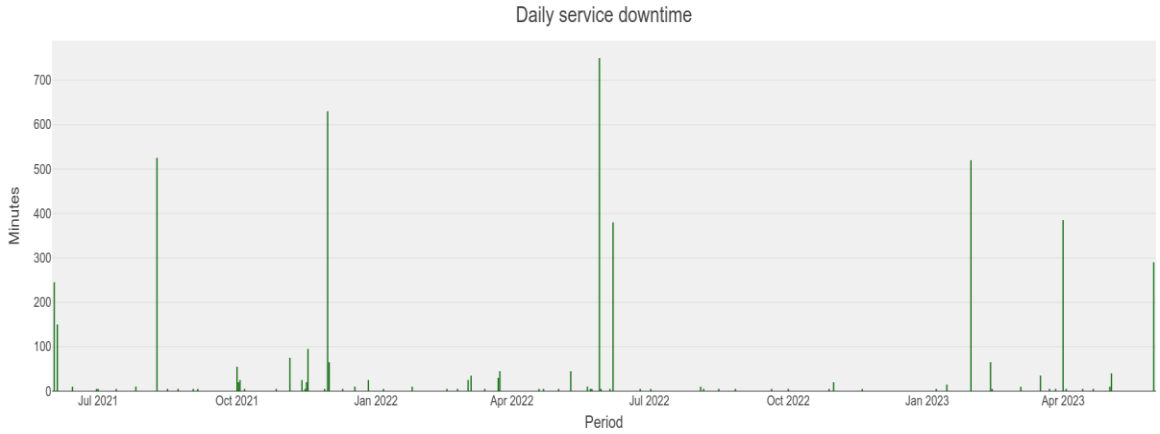

Figure 7: Daily maximum downtimes of QA4SM service in first project year

5.3.3 QA4SM support help desk

The email help desk has been set up and the user support address support@qa4sm.eu has continuously been available from the project start. This address is also used to send automatic service notifications to the development and maintenance team, including notifications of new user registrations, which are manually accepted or rejected by the team based on the provided registration information in each case. While new user registrations have been received on a continuous basis, first support requests have only been received in May 2023 in

relation to the user data upload feature, which was added in release 2. One contributing factor to this observation is that most of the more active QA4SM users are personally known by the project team and used other communication channels, like emails to individual team members, to clarify questions.

We expect that the number of users, the user activity and eventually also the rate of support requests will rise in the future with the increasing popularity of the data upload feature. We anticipate that this functionality will attract attention and also lead to an increase in support requests and request for enhancements, as the preparation of data for upload in QA4SM is more complex than other user interactions with QA4SM and some users may ask for relaxed compatibility constraints in order to validate less-standard datasets in different contexts.

6 User feedback and recommended platform improvements

The service developers gather user feedback through multiple channels. Users have the option to contact the team using the support email address (support@qa4sm.eu). Feedback is also collected from platform users reaching out to the developers via personal email or direct contacts, e.g., during conferences or workshops. Members of the Scientific Advisory Group (SAG) are pro-actively approached in regular intervals to provide feedback. Furthermore, a feedback questionnaire is available directly on the QA4SM service website. Initially established for the QA4SM workshop in 2022, the form remains accessible to this day. Certain questions within the questionnaire may, however, be outdated by now.

The most specific user feedback was received during and after the QA4SM user workshop in June 2022. A summary is provided below. After deployment of QA4SM release 2, the SAG and other key users were engaged to try the new service features. Several SAG members confirmed to assign team members on this activity and special interest was noted on the new upload and validation options for user data. At the time of writing this report, we do indeed see increasing user activity and feedback and will take this into account for further evolution of the service.

6.1 Summary of feedback collected at the QA4SM user workshop

During the past year, the most effective channel for receiving feedback was the QA4SM user workshop, which took place in combination with the 6th Satellite Soil Moisture Validation and Application Workshop in June 2022. At this workshop, the development team, the SAG, and QA4SM users gathered feedback to establish development objectives for the next QA4SM release.

The feedback session revolved mainly around the **upload feature for user data**, which has subsequently been integrated in QA4SM release 2. This feature was recognized as the most valuable (from a technical standpoint) enhancement for the service, and it was deemed particularly useful by the majority of participants for their respective applications. Various technical and scientific aspects were discussed, and the key feedback points are provided below.

- It was found that the netCDF file format is acceptable from a user perspective; the Climate Forecast (CF) metadata convention[4] is a good starting point. However, care should be taken in the formatting of irregularly gridded data sets and the accounting for the error that is introduced in the resampling. Especially with Level 2 data, this could pose a risk that should be dealt with. As a starting point 3-dimensional netCDF

[4] https://cfconventions.org/conventions.html

files should be the baseline for the service, however, in future versions there should be more flexibility to also allow the validation of L2 data (2 netCDF dimensions).

- A general discussion on preprocessing found that certain preprocessing steps are necessary in order to handle differences in the structures of data sets (e.g., spatial resampling). However, the service should be more transparent about these steps that could have a potential impact on the validation results.
- It was the general opinion in the discussion group that a few Gigabytes (up to ~10 GB) should be sufficient for the various purposes. A trade-off between the traceability of the data sets uploaded by the users and the hardware limitations related to permanent data storage must be found to guarantee sustainability. A possible way forward is to have a 'temporary' storage of data sets in an allocated space to each user, where test data and non-final versions can be uploaded and validated. Here, the clean-up will be performed with a higher frequency. Along with this, a more permanent space could be made available, where users can upload data that can be validated by other users and used to publish validation results. In this case, the clean-up will happen less often.
- The inclusion of a data filtering mechanism for user-uploaded data is necessary. It is advisable to implement an option that allows for the consideration of flags within the uploaded data, thereby eliminating the need for unnecessary pre-processing on the user's end before utilizing the service.
- In the future, it is essential to incorporate **stability metrics** to monitor sensor performance over time and identify potential degradation issues resulting from sensor drift, among other factors.
- At present, there is a lack of guidance regarding best practices for **"spatial" validation metrics**, which refer to metrics calculated on spatial samples and evolve at the time resolution of the satellite dataset. The service could benefit from incorporating **"autocorrelation metrics"** as well.
- There is ambiguity regarding the **temporal sampling of ERA5 data** in the service. Most users familiar with the dataset would expect the original, hourly sampling, but it is not explicitly indicated.
- While the platform was initially designed with a primary focus on satellite products, it is worth noting that model and in situ datasets are extensively utilized within the community. Therefore, the validation service is likely to be relevant not only for satellite developers but also for model developers and potentially for in situ data providers, as indicated by the workshop survey (8 participants, Figure 8).

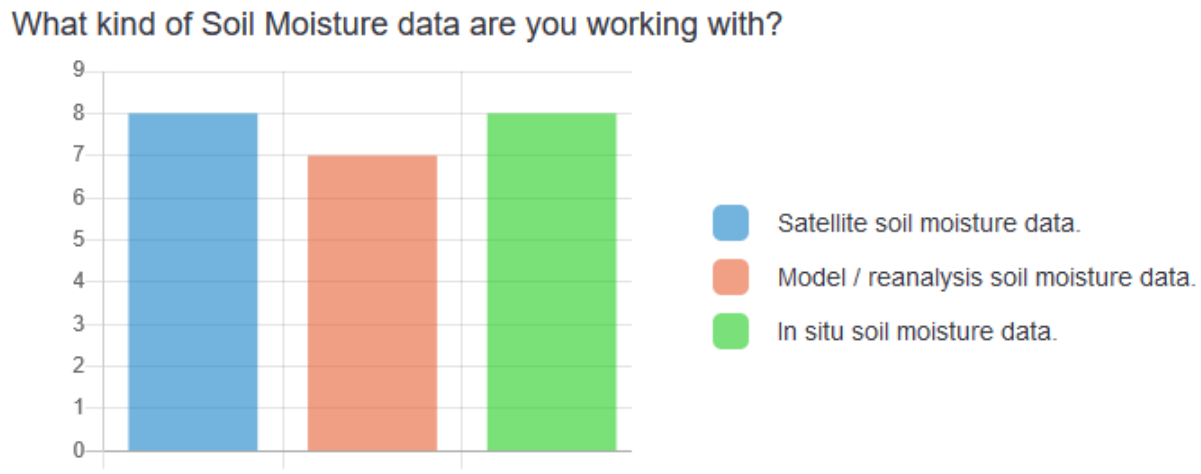

Figure 8: Survey results on "What kind of Soil Moisture data are you working with?"

- Regarding the necessary information accompanying FRM observations, which are presently incorporated into the QA4SM platform, a majority of workshop participants expressed a desire for **quality flags** to be included with the data (Figure 9). These flags would serve to indicate various aspects such as frozen soils or questionable patterns within the time series. Although flags already exist for all ISMN time series, it is acknowledged that they currently do not encompass all undesirable values. Hence, the potential development of supplementary flags has been suggested. **Uncertainty estimates for FRM observations** and **ancillary information**, such as on irrigation, are also considered important, but currently largely missing in the ISMN database.

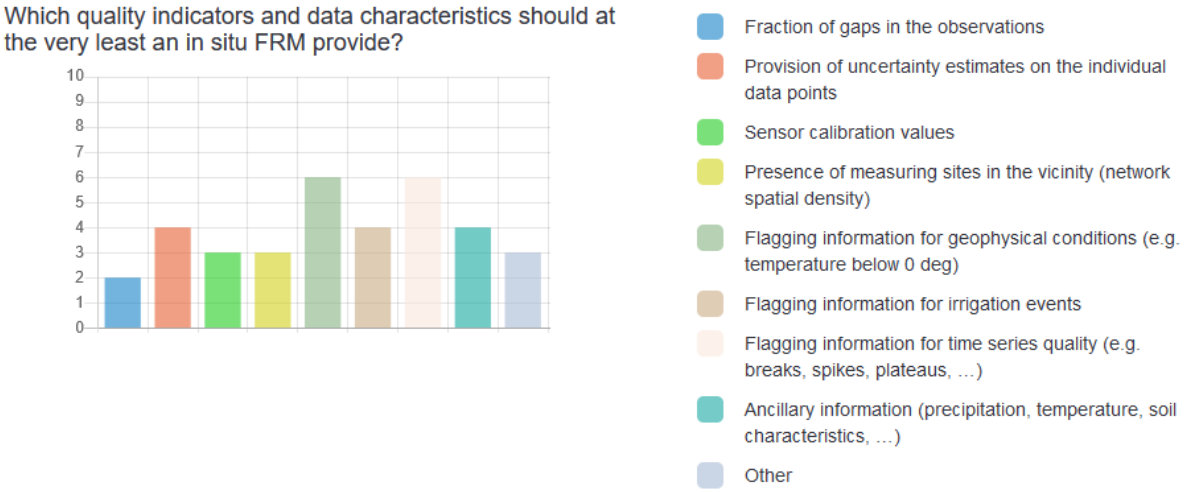

Figure 9: Survey results on "Which quality indicators and data characteristics should at the very least an in situ FRM provide?"

- In terms of the feature importance in the qualification of in situ measurements as FRMs, the **'time series length and temporal coverage'** of the measurements is agreed by the participants as one of the most important qualities in the measurements; the **'sensor calibration'** and **'number of stations in a satellite pixel'** come immediately

after (Figure 10). This indicates that there a high interest in the quality and quantity of the data production. According to the participants, all other features, including factors that are determined in the data analysis and post-production (e.g., 'committed area', 'spatial representativeness', 'scale mismatch') are less relevant.

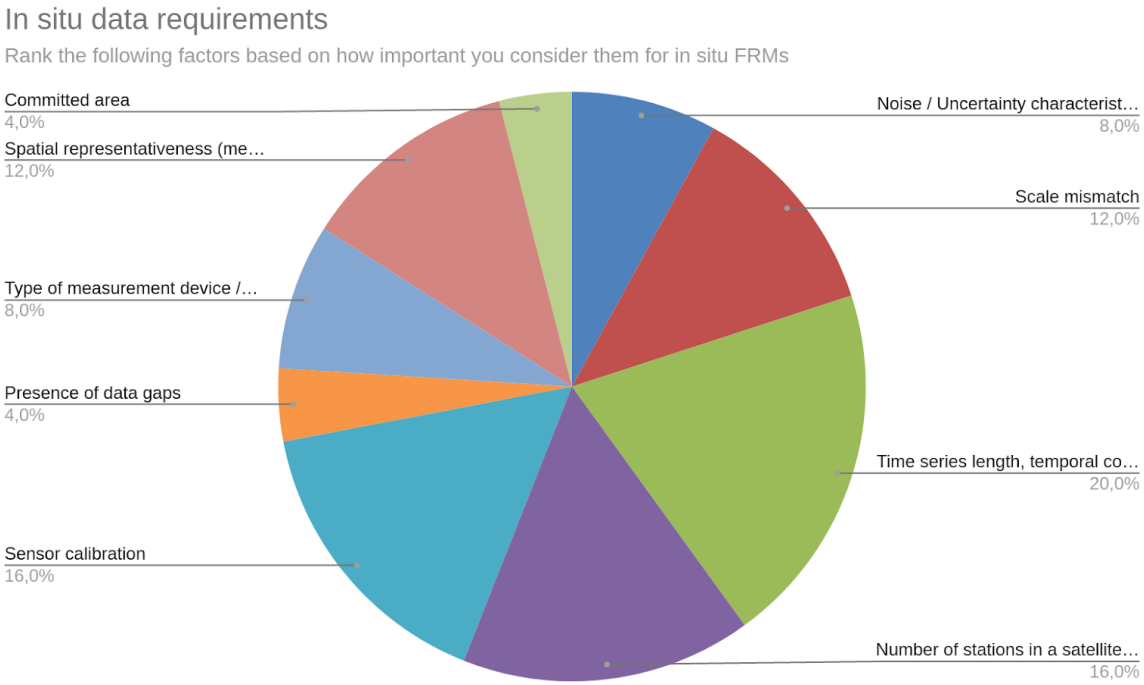

Figure 10: Survey results on "Rank the following factors based on how important you consider them for in situ FRMs"

6.2 Additional feedback collected from users directly

- The main request from users before QA4SM release 2 was to provide a **data upload form** to validate user specific data sets. There were multiple requests to support the validation of uploaded soil moisture data in netCDF and time series format (satellite, model and in situ data).
- With release 2, users can now upload up to 5 GB of either gridded (L3) stacked soil moisture images or location-specific time series as CSV files. However, additional formats are used in the community, which currently require pre-processing on the user side to bring the data into agreement with the (currently two) standards supported by QA4SM. Following recent user feedback, **additional help material** that describes the supported formats and provides guidelines on pre-processing data to fit the supported formats should be provided.
- To provide more specific information about the service capabilities to potential new users, it was suggested to give users access to the data set validation form without registration. This has been implemented in QA4SM release 2.

- It was suggested to rework the landing page of the service and provide a quick-start **introduction video or animation** for first time visitors to demonstrate the service without the need to create an account.
- Supported formats for user uploaded data are not clearly enough described. It was requested to provide additional help material on reformatting data correctly prior to uploading them to the service.
- It was noted, that the current limit of 5 GB is not sufficient for large data sets. However, it is recognized, that the provided storage space per user must be limited for a free service. It was also suggested to connect QA4SM to other cloud services such as Zenodo.
- The **support email address** (support@qa4sm.eu) is not placed prominently enough on the website **to be easily findable** for first time users.

7 References

Al-Yaari, A., S. Dayau, C. Chipeaux, C. Aluome, A. Kruszewski, D. Loustau, and J.-P. Wigneron 2018. "The AQUI Soil Moisture Network for Satellite Microwave Remote Sensing Validation in South-Western France" Remote Sensing 10, no. 11: 1839. https://doi.org/10.3390/rs10111839.

Bayat, Bagher, et al. "Toward operational validation systems for global satellite-based terrestrial essential climate variables." International Journal of Applied Earth Observation and Geoinformation 95 (2021): 102240.

Chevuturi, A., Tanguy, M., Facer-Childs, K., Martinez-de la Torre, A., Sarkar, S., Thober, S., ... & Blyth, E. (2023). Improving global hydrological simulations through bias-correction and multi-model blending. Journal of Hydrology, 129607.

Cosh, Michael, C. Montzka, A. Al Bitar, A. Berg, R. Bindlish, H. Bogena, F. Cabot, T. Caldwell, S. Chan, A. Colliander, W. Crow, N. Das, G. de Lannoy, E. Lopez-Baeza, W. Dorigo, A. Gruber, S. Hahn, T. Jagdhuber, S. Jones, Y. Kerr, S. Kim, C. Koyama, M. Kurum, F. Mattia, K. McColl, S. Mecklenburg, B. Mohanty, P. O´Neill, T. Pellarin, G. Petropoulos, M. Piles, R. Reichle, N. Rodriguez-Fernandez, C. Rüdiger, T. Scanlon, D. Spengler, P. Srivastava, S. Suman, R. van der Schalie, W. Wagner, U. Wegmüller, F. Camacho, and J. Nickeson (2020): Soil Moisture Product Validation Best Practice Protocol. Version 1.0.

Cosh, Michael H., et al. "Developing a strategy for the national coordinated soil moisture monitoring network." Vadose Zone Journal 20.4 (2021).

De Santis, D., et al. "Assimilation of satellite soil moisture products for river flow prediction: An extensive experiment in over 700 catchments throughout Europe." Water Resources Research 57.6 (2021): e2021WR029643.

Dorigo, W., Himmelbauer, I., Aberer, D., Schremmer, L., Petrakovic, I., Zappa, L., Preimesberger, W. and others (2021). The International Soil Moisture Network: serving Earth system science for over a decade. Hydrol. Earth Syst. Sci., 25, 5749–5804. https://doi.org/10.5194/hess-25-5749-2021.

Dorigo, W., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., … Lecomte, P. (2017a). ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. Remote Sensing of Environment. https://doi.org/10.1016/j.rse.2017.07.001.

Fernandez-Moran, R., Al-Yaari, A. , Mialon, A. , Mahmoodi, A , Al Bitar, A. ,De Lannoy, G., Lopez-Baeza, E., Kerr, Y., Wigneron, J.-P. (2017): SMOS-IC: An alternative SMOS soil moisture and vegetation optical depth product, Remote Sens., 9, 457; doi:10.3390/rs9050457.

Gruber, A., Scanlon, T., van der Schalie, R., Wagner, W., and Dorigo, W.: Evolution of the ESA CCI Soil Moisture climate data records and their underlying merging methodology, Earth Syst. Sci. Data, 11, 717–739, https://doi.org/10.5194/essd-11-717-2019, 2019.

Gruber, A., De Lannoy, G., Albergel, C., Al-Yaari, A., Brocca, L., Calvet, J.C., Colliander, A., Cosh, M., Crow, W., Dorigo, W., and others 2020. Validation practices for satellite soil moisture retrievals: What are (the) errors?. Remote Sensing of Environment, 244, p.111806.

Mousa, B. G., & Shu, H. (2020). Spatial evaluation and assimilation of SMAP, SMOS, and ASCAT satellite soil moisture products over Africa using statistical techniques. Earth and Space Science, 7, e2019EA000841. https://doi.org/10.1029/2019EA000841.

Peng, Jian, et al. "A roadmap for high-resolution satellite soil moisture applications–confronting product characteristics with user requirements." Remote Sensing of Environment 252 (2021): 112162.

Preimesberger, W., Scanlon, T., Su, C. H., Gruber, A., & Dorigo, W. (2020). Homogenization of structural breaks in the global ESA CCI soil moisture multisatellite climate data record. IEEE Transactions on Geoscience and Remote Sensing, 59(4), 2845-2862.

Wilkinson, Mark D.; Dumontier, Michel; Aalbersberg, IJsbrand Jan; Appleton, Gabrielle; et al. (15 March 2016). *"The FAIR Guiding Principles for scientific data management and stewardship"*. Scientific Data. 3: 160018. https://doi.org/10.1038/sdata.2016.18.

Zhao, H.; Montzka, C.; Baatz, R.; Vereecken, H.; Franssen, H.-J.H. The Importance of Subsurface Processes in Land Surface Modeling over a Temperate Region: An Analysis with SMAP, Cosmic Ray Neutron Sensing and Triple Collocation Analysis. Remote Sens.2021, 13, 3068. https://doi.org/10.3390/rs13163068

Appendix A. Uses and citations of QA4SM

Table 5 gives an overview of the QA4SM validations that have been published with a Digital Object Identifier in Zenodo, using the dedicated function in QA4SM.

Table 5: List of QA4SM validations that have been published in Zenodo

Title	Data sets used	Author	Date published
Validation of C3S v201912 vs ISMN 20191211 global	ISMN, C3S	Preimesberger, Wolfgang	July 7, 2020
Validation of ERA5 v20190613 vs CGLS SWI 1km V1.0	ERA5, CGLS SWI 1km	Stradiotti, Pietro	October 6, 2022
Validation of C3S v202012 vs C3S v201912 vs ERA5 v20190613	C3S, ERA5	Preimesberger, Wolfgang	May 4, 2021
Validation of ESA CCI SM active v06.1 vs ISMN 20210131 global	ESA CCI SM Active, ISMN	Srivastava, Harsh	July 19, 2021
Validation of C3S v202012 vs C3S v201912 vs ERA5-Land v20190904	C3S, ERA5-Land	Preimesberger, Wolfgang	May 1, 2021
Validation of ESA CCI SM combined v05.2 vs ISMN 20191211 global - without anomalies and without ISMN flags	ESA CCI SM Combined, ISMN	Aberer, Daniel	November 24, 2020
Validation of ESA CCI SM combined v05.2 vs ISMN 20191211 global - Anomalies and ISMN flagged	ESA CCI SM Combined, ISMN	Aberer, Daniel	November 24, 2020
Validation of C3S v202012 vs C3S v201912 vs ISMN 20210131 global	ISMN, C3S	Preimesberger, Wolfgang	May 4, 2021
Validation of ESA CCI SM combined v05.2 vs ISMN 20191211 global	ESA CCI SM Combined, ISMN	Preimesberger, Wolfgang	September 9, 2020
Validation of ESA CCI SM combined v05.2 vs ISMN 20191211 global - Anomalies and no ISMN flags	ESA CCI SM Combined, ISMN	Aberer, Daniel	November 24, 2020
Validation of ESA CCI SM combined v05.2 vs ISMN 20191211 global - without Anomalies and ISMN flagged	ESA CCI SM Combined, ISMN	Aberer, Daniel	November 24, 2020

Validation of ESA CCI SM combined v05.2 vs ERA5 v20190613	ESA CCI SM Combined, ERA5	Scanlon, Tracy	February 18, 2020
Validation of ESA CCI SM combined v06.1 vs ESA CCI SM combined v05.2 vs ERA5 v20190613	ESA CCI SM Combined, ERA5	Scanlon, Tracy	March 31, 2021
Validation of ESA CCI SM combined v05.2 vs ESA CCI SM combined v04.7 vs ERA5 v20190613 - Whole period	ESA CCI SM Combined, ERA5	Scanlon, Tracy	October 23, 2020
Validation of ESA CCI SM combined v04.7 vs ESA CCI SM combined v05.2 vs ISMN 20191211 global	ESA CCI SM Combined, ISMN	Preimesberger, Wolfgang	October 22, 2020
Validation of ESA CCI SM combined v04.7 vs ESA CCI SM combined v05.2 vs ISMN 20191211 global	ESA CCI SM Combined, ISMN	Scanlon, Tracy	October 23, 2020
Validation of ESA CCI SM combined v04.7 vs ESA CCI SM combined v05.2 vs ISMN 20191211 global	ESA CCI SM Combined, ISMN	Scanlon, Tracy	October 23, 2020