



**TECHNISCHE
UNIVERSITÄT
WIEN**
Vienna University of Technology

DIPLOMARBEIT

Automated Quality Control Procedures for the International Soil Moisture Network

Ausgeführt am
Department für Geodäsie und Geoinformation
der Technischen Universität Wien

unter der Anleitung von
Univ. Prof. Dipl.-Ing. Dr.techn. Wolfgang Wagner

Und der Betreuung durch
MSc. Dr.rer.nat. Wouter A. Dorigo

durch

Angelika Xaver
Schlöglgasse 45/1/9,
1120 Wien

Wien, im Oktober 2015

Abstract

Since its establishment in 2009, the International Soil Moisture Network (ISMN) provides harmonized ground-based soil moisture measurements, originating from a variety of operating networks. The quality of the collected data is highly variable, caused by differences in the sensing technique as well as by different local conditions at the measuring sites. In situ soil moisture observations are essential to evaluate and calibrate modelled and remotely sensed soil moisture products. Thus, the importance of meaningful quality measures for in situ soil moisture measurements is evident.

This study presents sophisticated automated quality control procedures to detect spikes, jumps and plateaus based on analyzing the shape of the soil moisture time series. Several conditions will be defined to identify these erroneous events through investigating the first and second derivatives, derived by the widely-known Savitzky-Golay filter.

The performance of the introduced quality control procedures will be evaluated by comparing the automated flagging results to manually flagged data of 40 selected soil moisture datasets from the ISMN.

Finally, a flagging statistic based on all soil moisture measurements contained in the ISMN will be presented.

Kurzfassung

Seit seiner Gründung im Jahr 2009, werden durch das International Soil Moisture Network (ISMN) harmonisierte in situ Bodenfeuchtemessungen bereitgestellt, die von einer Vielzahl von operativen Netzwerken stammen. Die Qualität der gesammelten Daten variiert sehr stark, aufgrund von Unterschieden der angewandten Sensortechnik ebenso wie aufgrund verschiedener lokaler Bedingungen bei den Messstationen. In situ Bodenfeuchtebeobachtungen sind essentiell, um modellierte und fernerkundungsbasierte Bodenfeuchteprodukte zu evaluieren und zu kalibrieren. Die Relevanz von aussagekräftigen Qualitätsmaßen für in situ Bodenfeuchtemessungen ist daher offensichtlich.

Diese Arbeit präsentiert komplexe automatisierte Verfahren der Qualitätskontrolle, um Spikes, Sprünge und Plateaus basierend auf der Form von Bodenfeuchte-Zeitreihen zu detektieren. Mehrere Bedingungen werden definiert, um diese fehlerhaften Ereignisse durch Untersuchung der ersten und zweiten Ableitungen, berechnet mithilfe des bekannten Savitzky-Golay Filters, zu identifizieren.

Die Performance der vorgestellten Verfahren der Qualitätskontrolle wird durch einen Vergleich der automatisiert geflaggt Resultate mit manuell geflaggt Daten, basierend auf einer Auswahl von 40 Bodenfeuchtedatensätzen des ISMN, evaluiert.

Abschließend wird eine Flagging-Statistik über alle im ISMN enthaltenen Bodenfeuchtedaten präsentiert.

Acknowledgements

I would like to thank my supervisors Wouter Dorigo and Wolfgang Wagner for their support and their patience.

Furthermore, I thank all data providers that kindly shared their data with the ISMN:

Jeffrey Walker (AACES), Thierry Pellarin (AMMA-CATCH), Jim Mather (ARM), Natalie Umphlett (AWDN), Loredana Marsica and Luca Brocca (CALABRIA), Matteo Gentilella and Giovanni Battista Chirico (CAMPANIA), Jonas Ardö (CARBOAFRICA), Konstantin Vinnikov and Thomas Collow (CHINA, IOWA, MONGOLIA, RUSWET- AGRO, RUSWET-GRASS, and RUSWET-VALDAI), Marek Zreda (COSMOS), Zhao Long and Kun Yang (CTP_SMTMN), Håkan Torbern Tagesson; Rasmus Fensholt (DAHRA), Dennis D. Baldocchi (FLUXNET-AMERIFLUX), Matias Takala, Hanne Suokanerva, Jouni Pullianen, Jarkko Koskinen (FMI), Pekka Hänninen and Raimo Sutinen (GTK), Simone Bircher (HOBE), Alessia Flammini and Renato Morbidelli (HYDROL-NET_PERUGIA), Minha Choi and Yeon Gil Lee (HYU_CHEONGMICHEON, HSC_SELMACHEON), Bob Scott (ICN), Pankaj Kumar Rai and Shivam Tripathi (ITT_KANPUR), Cristian Mattar Bader (LAB-net), Laura Dente, Bob Su, and J. Wen (MAQU), Nazzareno Diodato (MetEROBS), Udo Rummel (MOL-RAO), (ORACLE), Jeffrey Walker and Christoph Rüdiger (OZNET), Kristine Larson (PBO-H2O), José Martínez Fernández (REMEDHUS), Andrei Diamandi (RSMN), Garry Schaeffer, Michael Strobel, Maggie Dunklie, Michael Cosh, and Robert Parry (SCAN), Jean-Christophe Calvet (SMOSMANIA), Michael Strobel and Garry Schaeffer (SNOTEL), Bogusław Usowicz, Jerzy Usowicz, and Wojciech Marczewski (SWEX_Poland), Heye Bogena (TERENO), Florian Schlenz, Johanna dall'Amico, Alexander Loew, Wolfram Mauser (UDC_SMOS), Luca Brocca and Nicola Berni (UMBRIA), Vittorio Marletto and Marco Bittelli (UMSUOL), Michael Palecki (USCRN), Michael Cosh and Thomas Jackson (USDA-ARS), Ernesto Lopez-Baeza (VAS), Gottfried Kirchengast (WEGENERNET), George Petropoulos and Gareth Ireland (WSMN).

Without their valuable contributions this initiative would not have been possible in the first place.

The ISMN has been funded through the SMOS Soil Moisture Network Study (ESA ESTEC Contract No.22954/09) and the SMOS Soil Moisture Network Study–Operational Phase (ESA ESTEC Contract No.4000102722/10).

Table of Contents

List of Figures	ix
List of Tables.....	xiii
1 Introduction	1
2 Data	2
2.1 ISMN.....	2
2.2 Soil Moisture characteristics	7
2.3 Potential erroneous events within soil moisture readings	12
2.3.1 Jumps.....	13
2.3.2 Spikes.....	15
2.3.3 Constant values	15
3 Methods	18
3.1 Savitzky-Golay Filter	18
3.2 Algorithms for Detection.....	22
3.2.1 Spikes.....	24
3.2.2 Jumps.....	26
3.2.3 Plateaus	28
4 Results and Discussion	32
4.1 Results	32
4.2 Discussion.....	35
4.2.1 Spikes.....	35
4.2.2 Positive Jumps	40
4.2.3 Negative Jumps	43
4.2.4 Low Level Plateaus	47
4.2.5 Saturated Plateaus	49
4.3 Results for Entire ISMN	52
5 Conclusions	54
Bibliography	56

List of Acronyms

AACES	Australian Airborne Cal/val Experiment for SMOS
AMMA	See AMMA-CATCH
AMMA-CATCH	African Monsoon Multidisciplinary Analysis – Coupling the Tropical Atmosphere and the Hydrological Cycle
ARM	Atmospheric Radiation Measurement Climate Research Facility
AWDN	Automated Weather Data Network (name of in situ network in the USA)
CALABRIA	Name of in situ network in Italy
CAMPANIA	Name of in situ network in Italy
CARBOAFRICA	Name of in situ network in Sudan
CHINA	Name of in situ network in China
COSMOS	COsmic-ray Soil Moisture Observing System (network in USA)
CRN	Climate Reference Network
CTP_SMTMN	Central Tibetan Plateau Soil Moisture and Temperature Monitoring Network (name of in situ network in China)
DAHRA	Name of in situ network in Senegal
ESA	European Space Agency
FLUXNET_AMERIFLUX	Name of in situ network in USA
FMI	Finnish Meteorological Institute (name of in situ network in Finland)
GTK	Geological Survey of Finland (name of in situ network in Finland)
HOBE	Name of in situ network in Denmark
HSC_SELMACHEON	Name of in situ network in South Korea
HYDROL-NET_PERUGIA	Name of in situ network in Italy
HYU_CHEONGMICHEON	Name of in situ network in South Korea
ICN	Illinois Climate Network (name of in situ network in USA)
IIT_KANPUR	Indian Institute of Technology Kanpur (name of in situ network in India)
IOWA	Name of in situ network in USA)
ISMN	International Soil Moisture Network
LAB-net	Laboratory for Analysis of the Biosphere Network (name of in situ network in Chile)
MAQU	Name of in situ network in China
METEROBS	Met European Research Observatory (name of in situ network in Italy)
MOL-RAO	Lindenberg Meteorological Observatory / Richard-Aßmann Observatory
MONGOLIA	Name of in situ network in Mongolia
NRT	Near Real Time
OARCLE	Name of in situ network in France

OzNet	Name of in situ network in Australia
PBO_H2O	Plate Boundary Observatory (Name of in situ network in USA)
QC	Quality Control
REMEDHUS	Name of in situ network in Spain
RUSWET-AGRO	Name of in situ network in Former Soviet Union
RUSWET-GRASS	Name of in situ network in Former Soviet Union
RUSWET-VALDAI	Name of in situ network in Former Soviet Union
SASMAS	Name of in situ network in Australia
SCAN	Soil Climate Analysis Network
SMOS	Soil Moisture and Ocean Salinity
SMOSMANIA	Name of in situ network in France
SNOTEL	SNOWpack TElemetry
SWEX_POLAND	Soil Water Experiment Poland (name of in situ network in Poland)
TERENO	Terrestrial Environmental Observatories (name of in situ network in Germany)
UDC_SMOS	Upper Danube Catchment SMOS (name of in situ network in Germany)
UMBRIA	Name of in situ network in Italy
UMSUOL	Umidita del Suolo (name of in situ network in Italy)
USCRN	US Climate Reference Network
USDA_ARS	Name of in situ network in USA
VAS	Valencia Anchor Station
WEGENERNET	Name of in situ network in Austria
WSMN	Wales Soil Moisture Network

List of Figures

Figure 1: Networks contained in the ISMN (Status September 2015)	3
Figure 2: Stations contained in the ISMN (Status September 2015)	3
Figure 3: Time period spanned by individual networks in ISMN (Status September 2015)	4
Figure 4: Example of typical soil moisture wetting and drying behaviour (Station Villeville of network SMOSMANIA)	8
Figure 5: Example of soil moisture during several consecutive precipitation events (Station Fitterizzi of network CALABRIA)	9
Figure 6: Example of a soil moisture time series containing random noise (Station CSS LAB of network SNOTEL)	10
Figure 7: Example of soil moisture measurements at different depths (Station Pezenas of network SMOSMANIA)	11
Figure 8: Example of temperature sensitivity of soil moisture at different depths. The diurnal cycle of temperature is visible within the soil moisture readings, well pronounced at the top layer and barely visible in the deepest layer. (Station L18 of network CTP_SMTMN)	12
Figure 9: Example of a soil moisture time series containing a positive jump. Also noticeable are missing values, their effect on the flagging result will be discussed in chapter 4. (Station Vaira Ranch of network FLUXNET-AMERIFLUX)	14
Figure 10: Example of a soil moisture time series containing a negative jump (Station Wildenrath of network TERENO)	14
Figure 11: Example of a soil moisture time series containing positive and negative spikes (Station INDEPENDENCE LAKE of network SNOTEL)	15
Figure 12: Example of a soil moisture time series containing a saturated plateau (Station Meeker of network ARM)	16
Figure 13: Example of a soil moisture time series containing a low level plateau (Station MelbexII of network VAS)	17
Figure 14: Shape of smoothed time series (top), first (middle) and second (bottom) derivative resulting from Savitzky-Golay filtering with parameters: symmetrical filter width of 3, 2 nd order polynomial (Station Earlsboro of network ARM, Water Matric Potential Sensor 229L_W, 0.15-0.15m depth)	20
Figure 15: Shape of smoothed time series (top), first (middle) and second (bottom) derivative resulting from Savitzky-Golay filtering with parameters: symmetrical filter width of	

19, 2 nd order polynomial (Station Earlsboro of network ARM, Water Matric Potential Sensor 229L_W, 0.15-0.15m depth)	20
Figure 16: Shape of smoothed time series (top), first (middle) and second (bottom) derivative resulting from Savitzky-Golay filtering with parameters: symmetrical filter width of 19, 3 rd order polynomial (Station Earlsboro of network ARM, Water Matric Potential Sensor 229L_W, 0.15-0.15m depth)	21
Figure 17: Shape of smoothed time series (top), first (middle) and second (bottom) derivative resulting from Savitzky-Golay filtering with parameters: symmetrical filter width of 19, 4 th order polynomial (Station Earlsboro of network ARM, Water Matric Potential Sensor 229L_W, 0.15-0.15m depth)	22
Figure 18: Shape of first (middle) and second derivative (bottom) for a spike within the soil moisture time series (Station Eulo of network OZNET, Stevens Hydra Probe, 0.00-0.05m depth)	25
Figure 19: Shape of first (middle) and second derivative (bottom) for a positive jump within the soil moisture time series (Station Kemole Gulch of network SCAN, Hydraprobe Analog (2.5 Volt), 0.50-0.50m depth)	27
Figure 20: Shape of first (middle) and second derivative (bottom) for a negative jump within the soil moisture time series (Station Kemole Gulch of network SCAN, Stevens Hydra Probe, 0.10-0.10m depth).....	28
Figure 21: Shape of first (middle) and second derivative (bottom) for a saturated plateau within the soil moisture time series (Station Meeker of network ARM, Water Matric Potential Sensor 229L_E, 0.15-0.15m depth)	30
Figure 22: Shape of first (middle) and second derivative (bottom) for a low level plateau within the soil moisture time series (Station Vaira Ranch of network FLUXNET-AMERIFLUX, ThetaProbe ML2X, 0.20-0.20m depth)	31
Figure 23: Example for flagging results of a spikes (Station Pawhuska of network ARM, Water Matric Potential Sensor 229L_E, 0.05-0.05m depth).....	36
Figure 24: Example for flagging results of spikes (Station Slatina of network RSMN, 5TM, 0.00-0.05m depth)	36
Figure 25: Example for flagging results of spikes (Station UAPB Dewitt of network SCAN, Hydraprobe Analog (2.5 Volt), 0.10-0.10m depth)	37
Figure 26: Example for flagging results of spikes (Station INDEPENDENCE LAKE of network SNOTEL, Hydraprobe Analog (2.5 Volt), 0.50-0.50m depth).....	37
Figure 27: Shape of first (middle) and second derivative (bottom) for a soil moisture time series with random noise (Station CSS LAB of network SNOTEL, Hydraprobe Analog (2.5 Volt), 0.20-0.20m depth).....	38

Figure 28: Instead of the spike in late August 2013 two correct but noisy measurements were flagged using the spike detection algorithm (Station CSS LAB of network SNOTEL, Hydraprobe Analog (2.5 Volt), 0.20-0.20m depth)	39
Figure 29: Due to noisy data not all of the spikes were correctly detected by the automatic quality control procedure (Station Kemole Gulch of network SCAN, Hydraprobe Analog (2.5 Volt), 0.05-0.05m depth) erroneous events (Station node828 of network SOILSCAPE, EC5, 0.30-0.30m depth)	39
Figure 30: Example for a flagging result of spikes for a soil moisture time series with consecutive erroneous events (Station Fairhope_3_NE of network USCRN, Stevens Hydraprobe II Sdi-12, 0.05-0.05m depth).....	40
Figure 31: Example for flagging results of a positive jump (Station Vaira Ranch of network FLUXNET-AMERIFLUX, ThetaProbe ML2X, 0.20-0.20m depth).....	41
Figure 32: Example for flagging results of a positive jump (Station Montrose_11_ENE of network USCRN, Hydraprobe II Sdi-12, 0.10-0.10m depth).....	41
Figure 33: Example for flagging results of a positive jump (Station Manhattan_6_SSW of network USCRN, Stevens Hydraprobe II Sdi-12, 0.10-0.10m depth).....	42
Figure 34: Example for flagging results of a positive jump (Station Slobozia of network RSMN, 5TM, 0.00-0.05m depth)	42
Figure 35: Example for erroneously flagged positive jumps (Station Kainaliu of network SCAN, Hydraprobe Analog (2.5 Volt), 0.05-0.05m depth)	43
Figure 36: Example for an erroneously flagged positive jump (Station UAPB Dewitt of network SCAN, Hydraprobe Analog (2.5 Volt), 0.10-0.10m depth)	43
Figure 37: Example for flagging results of a negative jump (Station Corugea of network RSMN, 5TM, 0.00-0.05m depth)	44
Figure 39: Example for flagging results of a negative jump (Station Kemole Gulch of network SCAN, Hydraprobe Analog (2.5 Volt), 0.05-0.05m depth)	45
Figure 40: Example for flagging results of a negative jump (Station node804 of network SOILSCAPE, EC5, 0.15-0.15m depth)	45
Figure 41: Example for flagging results of a negative jump (Station Slobozia of network RSMN, 5TM, 0.00-0.05m depth)	46
Figure 42: Example for a flagging result of negative jumps for a soil moisture time series with consecutive erroneous events (Station node828 of network SOILSCAPE, EC5, 0.30-0.30m depth)	46
Figure 43: Example for erroneously flagged negative jumps, catching some negative spikes (Station INDEPENDENCE LAKE of network SNOTEL, Hydraprobe Analog (2.5 Volt), 0.50-0.50m depth)	47

Figure 44: Example for flagging results of low level plateaus (Station Earlsboro of network ARM, Water Matric Potential Sensor 229L_W, 0.15-0.15m depth)	48
Figure 45: Example for flagging results of low level plateaus (Station RC of network USDA_ARS, Hydraprobe Analog (2.5 Volt), 0.00-0.05m depth).....	48
Figure 46: Example for flagging results of a low level plateau (Station MelbexII of network VAS, ThetaProbe ML2X, 0.00-0.05m depth)	49
Figure 47: Example for flagging results of a saturated plateau (Station Meeker of network ARM, Water Matric Potential Sensor 229L_W, 0.05-0.05m depth)	50
Figure 48: Example for flagging results of saturated plateaus (Station NST_04 of network MAQU, ECH20 EC-TM, 0.05-0.05m depth).....	50
Figure 49: Example for flagging results of saturated plateaus (Station Morris Farms of network SCAN, Hydraprobe Digital Sdi-12 (2.5 Volt), 0.20-0.20m depth)	51
Figure 50: Example for flagging results of saturated plateaus (Station 50 of network WEGENERNET, pF-Meter, 0.30-0.30m depth)	51
Figure 51: Example for flagging results of saturated plateaus (Station node412 of network SOILSCAPE, EC5, 0.20-0.20m depth)	52
Figure 52: Percentages of spectrum-based quality flags per network.....	53

List of Tables

Table 1: Networks contained in the ISMN (status September 2015).....	5
Table 2: NRT Networks.	7
Table 3: Datasets and periods selected for developing spectrum-based QC methods.	23
Table 4: Datasets and periods selected for evaluating spectrum-based QC methods.....	32
Table 5: Results of the flagging performance (all values given in percentage).	35

1 Introduction

This thesis is based on the journal paper '*Global Automated Quality Control of In situ Soil Moisture data from the International Soil Moisture Network*' (Dorigo, et al., 2013). It focuses in more detail on the theory behind the spectrum based quality flags presented in Dorigo et al. (2013), which were developed by the second author, i.e., the author of this thesis.. Furthermore, an extended evaluation of the proposed flagging scheme – carried out using recent soil moisture ground measurements – will be presented.

Soil moisture is defined as the water content in the root zone of the soil, reaching to a depth of 200 centimeters (Legates, et al., 2011). Although the overall amount of soil moisture represents only a fraction of the global water budget it plays an essential role within the hydrological cycle and the climate system.

At present, soil moisture can be determined through three different methods: in situ measurements, modeling and satellite retrieval. Spatially complete measurements of global soil moisture patterns cannot be obtained through in situ observations due to both economic and maintenance reasons. Satellite-derived soil moisture products overcome both of these issues and hence represent the most promising source of global and long-term soil moisture data availability. However, in situ soil moisture measurements are still crucial for evaluating and calibrating satellite-derived and model-based soil moisture products. Thus, knowledge about the quality of in situ soil moisture is of high importance for the reliability of any satellite product validation study, as well as modeled soil moisture products (e.g. Dorigo et al., 2015; Albergel et al., 2012).

In this thesis, quality control procedures will be elaborated for in situ soil moisture measurements from the International Soil Moisture Network (ISMN) (Dorigo et al., 2011a,b). Different strategies for quality assessments, such as simple threshold based methods or geophysically based checks are currently employed within the ISMN (Dorigo, et al., 2013). This thesis focuses on more advanced automated spectrum-based methods. In a nutshell, the plausibility of soil moisture measurements will be evaluated by examining the shape of a soil moisture time series by means of its first and second derivative derived by the widely-known Savitzky-Golay filter.

2 Data

2.1 ISMN

The International Soil Moisture Network (ISMN; <http://ismn.geo.tuwien.ac.at/>) was initiated by the European Space Agency (ESA) in 2009 and since its official launch in 2010 it has been operated by the Department of Geodesy and Geoinformation (GEO) of the TU Wien (Dorigo et al., 2011a,b). The ISMN acts as a data repository for ground based soil moisture measurements, intended to provide reference data for calibration and validation of remotely-sensed soil moisture missions (e.g. SMOS).

Soil moisture observations are collected from various networks distributed all over the globe. The gathered data differ for example in format, structure, naming conventions and meta data. The observations themselves are resulting from different measurement techniques and therefore differ in measurement depths, sampling intervals, units, accuracy and also in certain characteristics. Within the fully automated processing chain of the ISMN, the observations are harmonized in terms of unit, temporal resolution and data format. After running through several quality control procedures, which will be described in the subsequent chapters, the processed data is stored in a database. Through a web portal the measurements become available to the public, where the data is visualized through a data viewer and made accessible for download to registered users.

Currently, the ISMN stores the data of 49 networks consisting of 2021 stations providing more than 8100 different soil moisture datasets (Figure 1, Table 1), scattered over 22 different countries. While the density of in situ stations is especially high in the US and at a moderate level in Europe, also many data sparse region exist, such as South America, Africa and Asia (Figure 2).

In addition to soil moisture other climate variables such as soil temperature, air temperature, surface temperature, precipitation, snow depth, snow water equivalent and soil suction are collected and provided through the ISMN. The data hosted by the ISMN covers a time period of 63 years, starting from 1952 until now (Figure 3). While most datasets are updated on an irregular basis, a few networks exist which provide their data in near real time (NRT), on a daily or weekly basis (Table 2). These datasets are retrieved, processed and written into the database automatically without the possibility of any manual quality control. Thus, automated quality control procedures within the ISMN become even more important,

particularly for such NRT datasets, in order to provide reliable soil moisture measurements to its users.



Figure 1: Networks contained in the ISMN (Status September 2015)

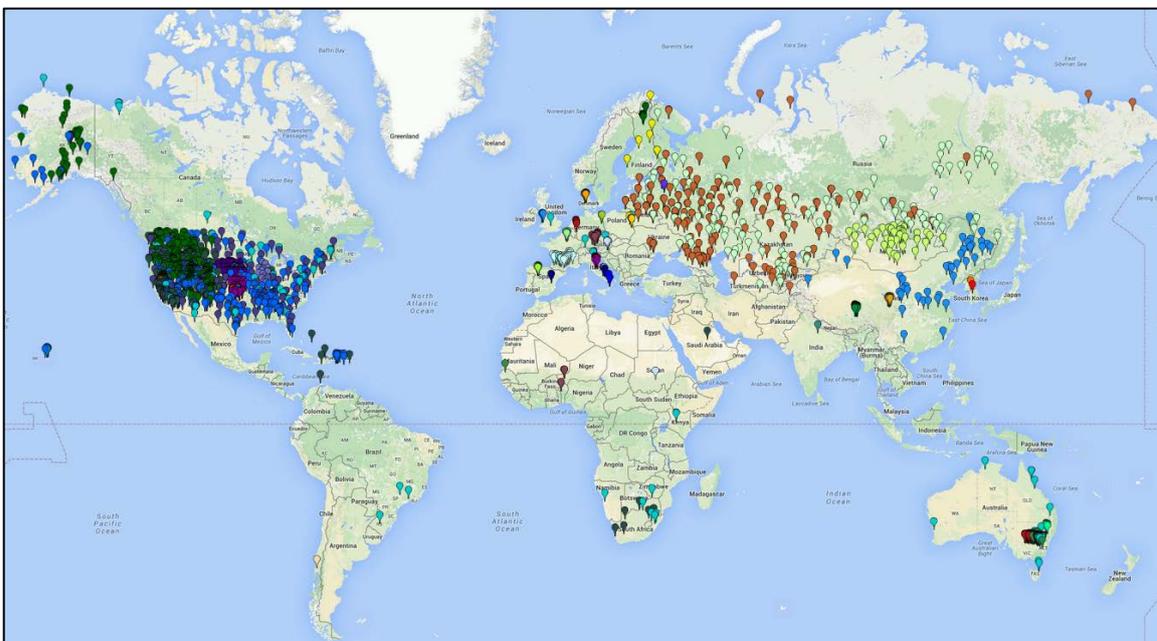


Figure 2: Stations contained in the ISMN (Status September 2015)

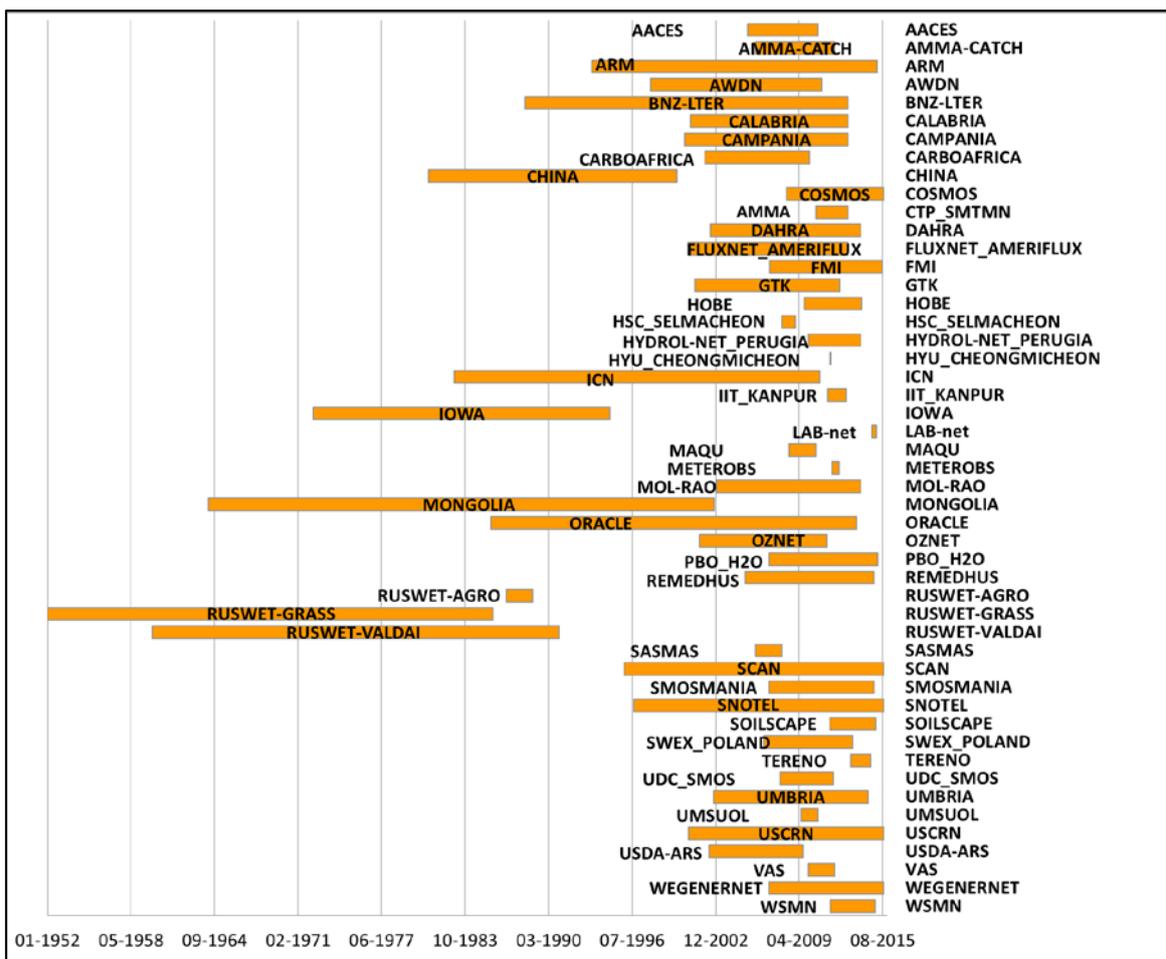


Figure 3: Time period spanned by individual networks in ISMN (Status September 2015)

As mentioned above, the provided in situ observations within the ISMN originate from different networks. Thus, the measurements are realized using a variety of different sensor types. Based on the underlying measurement principle, the following sensor types can be distinguished: coaxial impedance dielectric reflectometry sensors, capacitance sensors, frequency domain reflectometry (FDR) probes, time domain reflectometry (TDR) sensors, matric potential probes, neutron probes, gravimetric probes, and cosmic ray probes. Each measurement technique has its respective advantages and drawbacks, characteristics, and measurement accuracy (Dorigo et al., 2011b; Robinson et al., 2008). Furthermore, soil moisture observations are obtained at different measurement depths and/or depth intervals. Consequently, the soil moisture readings vary in their shape. The diversity of the network data is also noticeable in the data quality and the quality checks performed by the data providers. Not every network applies quality control procedures to their data or adds quality flags. Furthermore, if quality control mechanisms exist, they are usually not consistent, due to

the lack of standardized quality control methods. Thus, the importance of harmonized quality control procedures, which can be applied to all networks consistently, is evident.

Simple quality control procedures have been employed within the ISMN processing chain from the very beginning of its operation (Dorigo, et al., 2011b). Back then – and since - basic threshold-based approaches were applied to all implemented datasets. In spring 2014, a new sophisticated quality control procedure has been implemented into the automated processing chain. It consists of three different types of quality control methods: consistency test of geophysical range, geophysical consistency checks, spectrum-based approaches. While the checks for the consistency test of geophysical range represent basic threshold-based approaches, the geophysical consistency checks make use of additional in situ and model-based meteorological variables. Both methods are described and discussed in (Dorigo, et al., 2013). In the following, the spectrum-based quality control approach will be further explained in detail.

Table 1: Networks contained in the ISMN (status September 2015)

Name	Country	Stations	Variables (Number of depth Layers)	Period in database
AACES	Australia	49	SM (3), TS (5), P	2005-05-09 – 2010-09-24
AMMA-CATCH	Benin, Niger, Mali	6	SM (12)	2006-01-01 – 2011-12-31
ARM	USA	29	SM (10), TA (1), TS (10, P)	1993-06-29 – 2015-03-26
AWDN	USA	50	SM (4)	1997-12-31 – 2010-12-30
BNZ-LTER	Alaska	10	SM (4), TA (2), TS (9), TSF (1), P (1), SD (1), SWEQ (1)	1988-06-01 – 2013-01-01
CALABRIA	Italy	5	SM (3), TA (1), P	2001-01-01 – 2012-12-31
CAMPANIA	Italy	2	SM (1), TA (1), P	2000-07-27 – 2012-12-31
CARBOAFRICA	Sudan	1	SM (7), TA (1), TS (2), P	2002-02-08 – 2010-01-20
CHINA	China	40	SM (11)	1981-01-08 – 1999-12-28
COSMOS	USA, Germany, Switzerland, France, Brasil, Kenya	109	SM (41)	2008-04-28 – 2015-10-09
CTP_SMTMN	China	57	SM (4), TS (4)	2010-08-01 – 2013-01-01
DAHRA	Senegal	1	SM (5), TA (1), TS(5), P	2002-07-04 -- 2014-01-01
FLUXNET_AMERIFLUX	USA	2	SM (8), TA (1), TS (5), P	2000-10-22 – 2013-01-01

FMI	Finland	27	SM (7), TA (2), TS (7)	2007-01-25 - 2015-08-22
GTK	Finland	7	SM (5), TA (1), TS (6)	2001-05-16 - 2012-05-29
HOBE	Denmark	32	SM (3), TS (3)	2009-09-08 - 2014-02-03
HSC_SELMACHEON	Korea	1	SM (1)	2008-01-01 - 2009-01-01
HYDROL- NET_PERUGIA	Italy	2	SM (4), TA (1), TS (2), P	2010-01-01 - 2013-12-31
HYU_CHEONGMICHEON	Korea	1	SM (1)	2011-08-25 - 2011-09-20
ICN	USA	19	SM (11), TA (6, P)	1983-01-03 - 2010-11-21
IIT_KANPUR	India	1	SM (4)	2011-06-16 - 2012-11-22
IOWA	USA	6	SM (12)	1972-04-04 - 1994-11-15
LAB-net	Chile	1	SM (2), TA (1), TS (1), P	2014-10-25 - 2015-03-06
MAQU	China	20	SM (1), TS (1)	2008-06-30 - 2010-07-31
METEROBS	Italy	1	SM (5)	2011-10-10 - 2012-05-09
MOL-RAO	Germany	2	SM (9), TS (12), TA(2), P	2003-01-01 - 2014-01-01
MONGOLIA	Mongolia	44	SM (10layers)	1964-04-08 - 2002-10-28
ORACLE	France	6	SM (26), TA (2), TS (2), P	1985-10-18 - 2013-09-09
OZNET	Australia	38	P (2), SM (7), SU (5), TS (6)	2001-09-12 - 2011-05-31
PBO_H2O	USA	109	SM (2), TA (1), P	2007-07-07 - 2015-04-06
REMEDHUS	Spain	24	SM (1), TS (1)	2005-03-15 - 2015-01-01
RUSWET-AGRO	Former Soviet Union	212	SM (2)	1958-04-08 - 2002-06-28
RUSWET-GRASS	Former Soviet Union	122	SM (2)	1952-06-08 - 1985-12-28
RUSWET-VALDAI	Former Soviet Union	3	SM (3), TA (1), TS (3), P	1960-01-15 - 1990-12-15
SASMAS	Australia	14	SM (2), TS (1)	2005-12-31 - 2007-12-31
SCAN	USA	215	SM (25), TS (25), TA (1), P, SD, SWEQ	1996-01-01 - 2015-10-09
SMOSMANIA	France	21	SM (4), TS (4)	2007-01-01 - 2015-01-01
SNOTEL	USA	420	SM (16), TS (16), TA, SWEQ, P, SD	1996-09-11 - 2015-10-09
SOILSCAPE	USA	136	SM (28), TS (1)	2011-08-26 - 2015-02-17
SWEX_POLAND	Poland	6	SM (10), TS (10), P	2006-08-03 - 2013-05-06
TERENO	Germany	5	SM (3), TA (1), TS (3)	2013-04-01 - 2014-10-04

UDC_SMOS	Germany	11	SM (5)	2007-11-08 – 2011-11-18
UMBRIA	Italy	13	SM (5), TA (1), P	2002-10-09 – 2014-08-01
UMSUOL	Italy	1	SM (7)	2009-06-12 – 2010-09-30
USCRN	USA	115	SM (5), TA (1), TS (5), TSF (1), P	2000-11-15 – 2015-10-09
USDA-ARS	USA	4	SM (1), TS (1)	2002-06-01 – 2009-07-31
VAS	Spain	3	SM (1), TA (2), TS (10)	2010-01-01 – 2012-01-01
WEGENERNET	Austria	12	SM (2), TA (1), TS (2), P	2007-07-07 – 2015-05-03
WSMN	UK	7	SM (3), TS (3)	2011-09-02 – 2015-02-12
Total	49	2021		1952-06-08 – now

Table 2: NRT Networks.

Network	Number of stations updated in NRT	NRT status
ARM	29	Active
COSMOS	109	Active
FMI	27	Currently not active
LAB-net	1	Currently not active
PBO_H2O	109	Currently not active
SCAN	215	Active
SNOTEL	420	Active
SOILSCAPE	136	Currently not active
WEGENERNET	12	Active
USCRN	115	Active

2.2 Soil Moisture characteristics

The difficulty in defining quality principles for in situ soil moisture measurements lies in the characteristic shape of a soil moisture time series going through the phases of wetting and drying states (Hillel, 1998). Precipitation events may result in severe rise of soil moisture within only one or a few hours, while the drying process of soil proceeds slowly, resulting in an inverse exponential shape (Figure 4). These specific properties inhibit the use of typical

outlier detection algorithms as most of the natural rises of soil moisture caused by rain events would be flagged as erroneous data.

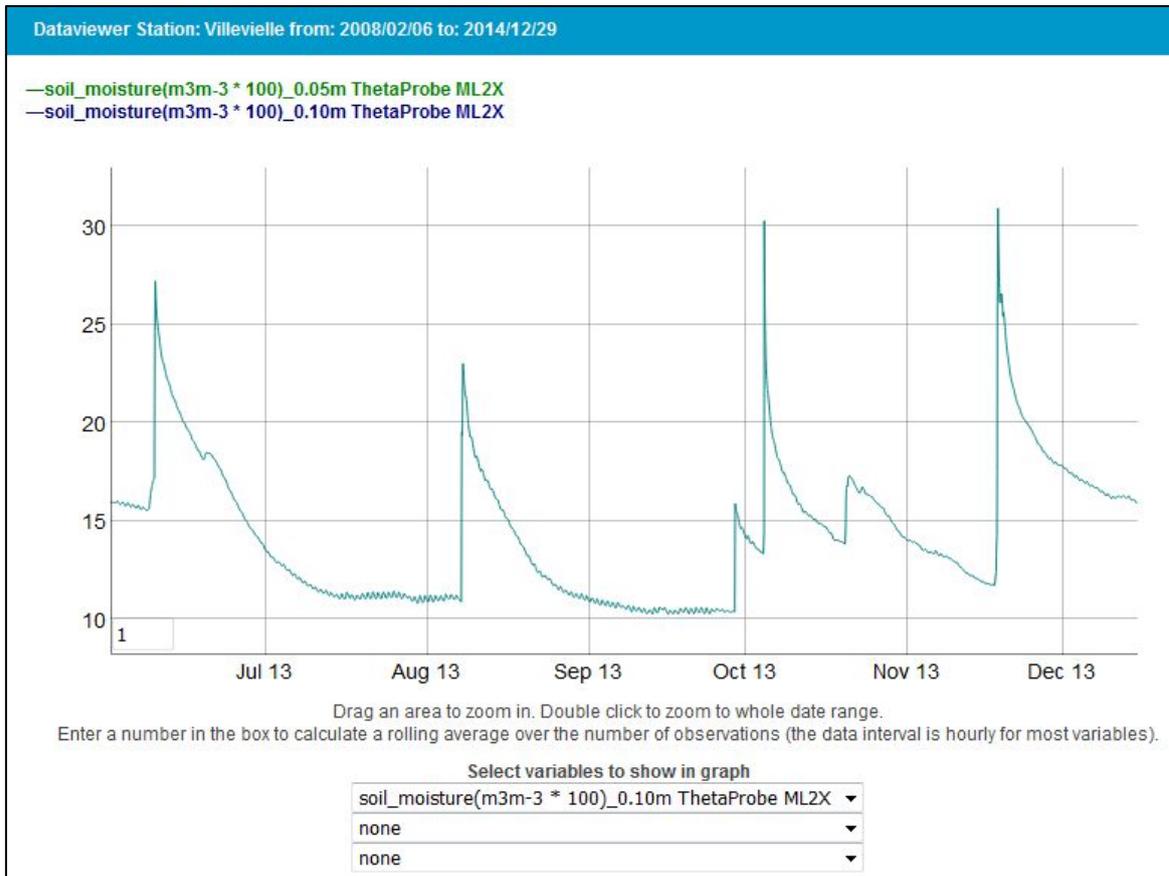


Figure 4: Example of typical soil moisture wetting and drying behaviour (Station Villeville of network SMOSMANIA)

If precipitation events recur faster than the soil needs to reach a stable dry level, the rises of soil moisture may overlap and may not be differentiable any more (Figure 5).

While soil moisture measured at top layers of the soil responds quite strongly to precipitation events, a smoother and less intensive reaction can be observed in deeper layers (Figure 7). Furthermore, it is known that also temperature dependent soil moisture variations exist (Figure 8; Dorigo et al., 2011b; Robinson et al., 2008). These fluctuations induced by daily temperature cycles again mainly affect the upmost soil moisture layer while their influence decreases with increasing depth. Furthermore, depending on the soil moisture sensor, random noise can be present within the soil moisture observations (Figure 6), which often does not exceed the sensor accuracy.

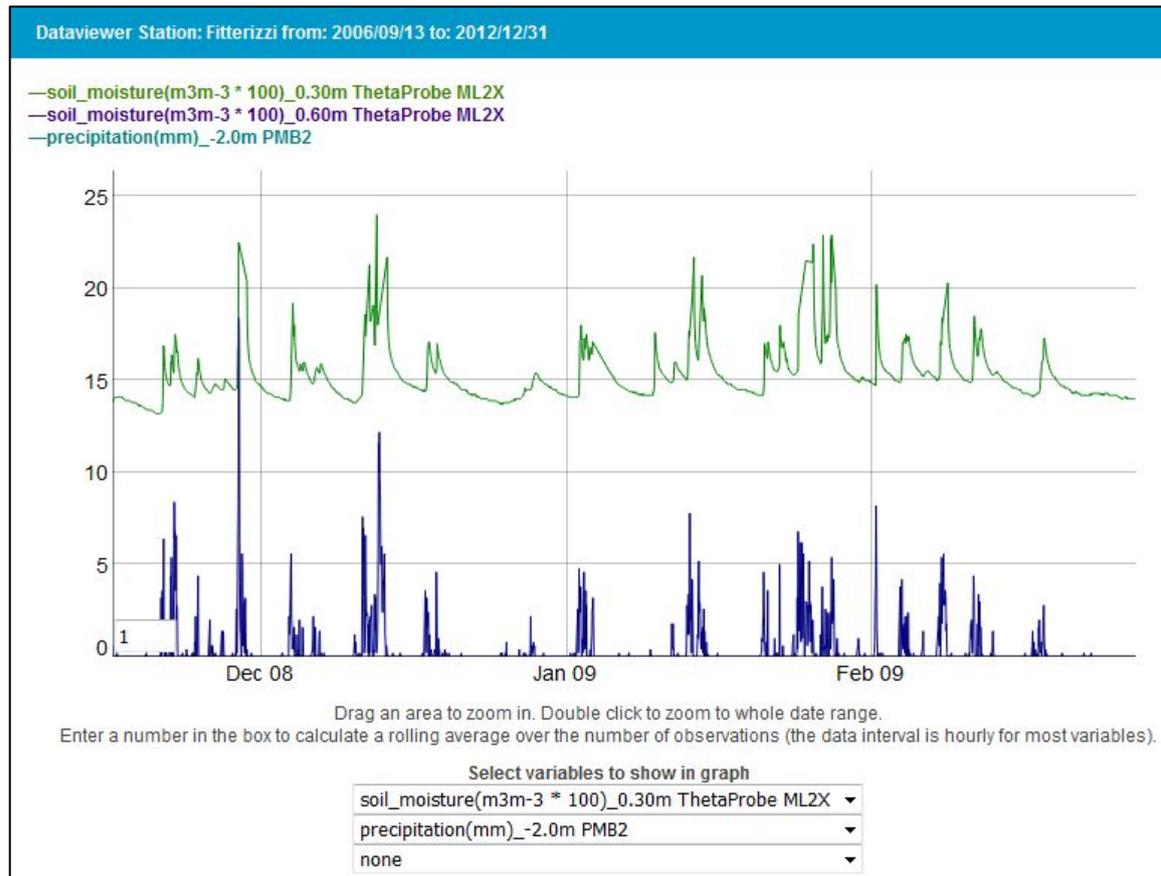


Figure 5: Example of soil moisture during several consecutive precipitation events (Station Fitterizzi of network CALABRIA)

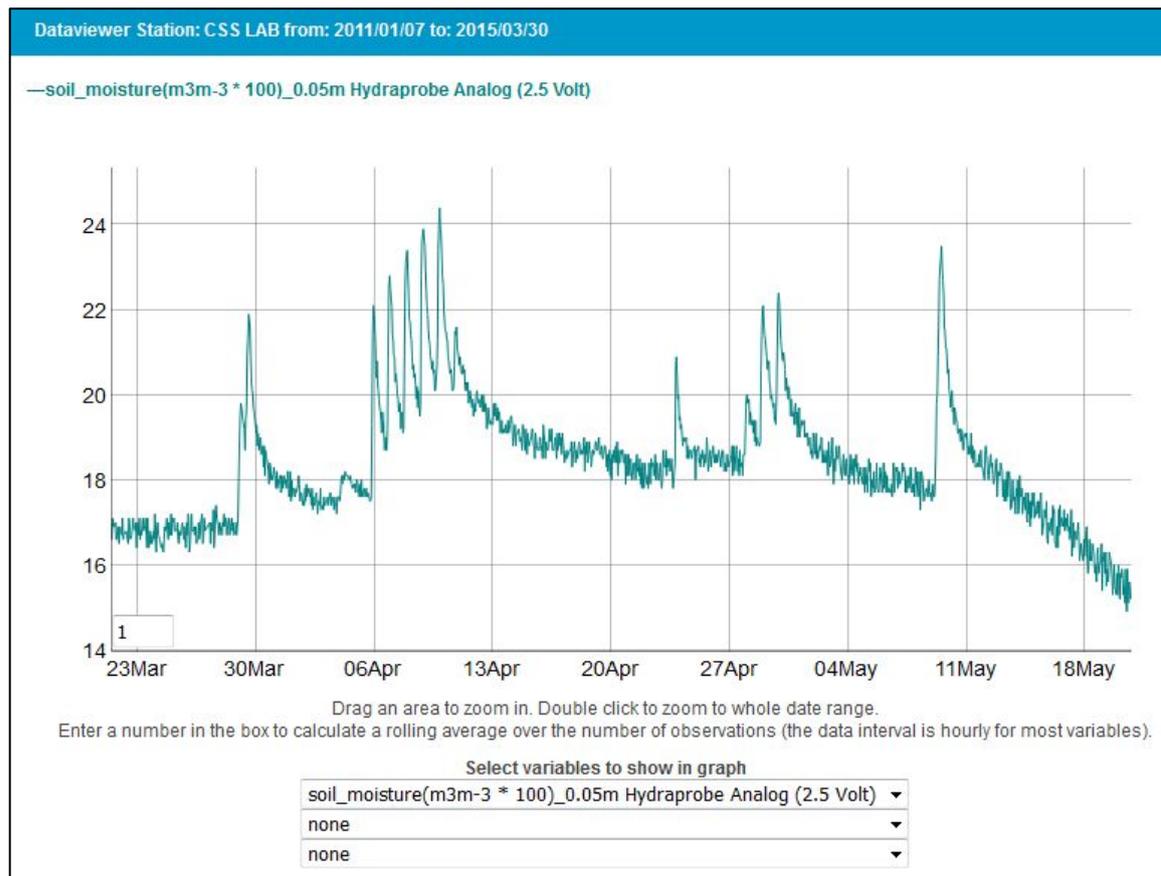


Figure 6: Example of a soil moisture time series containing random noise (Station CSS LAB of network SNOTEL)

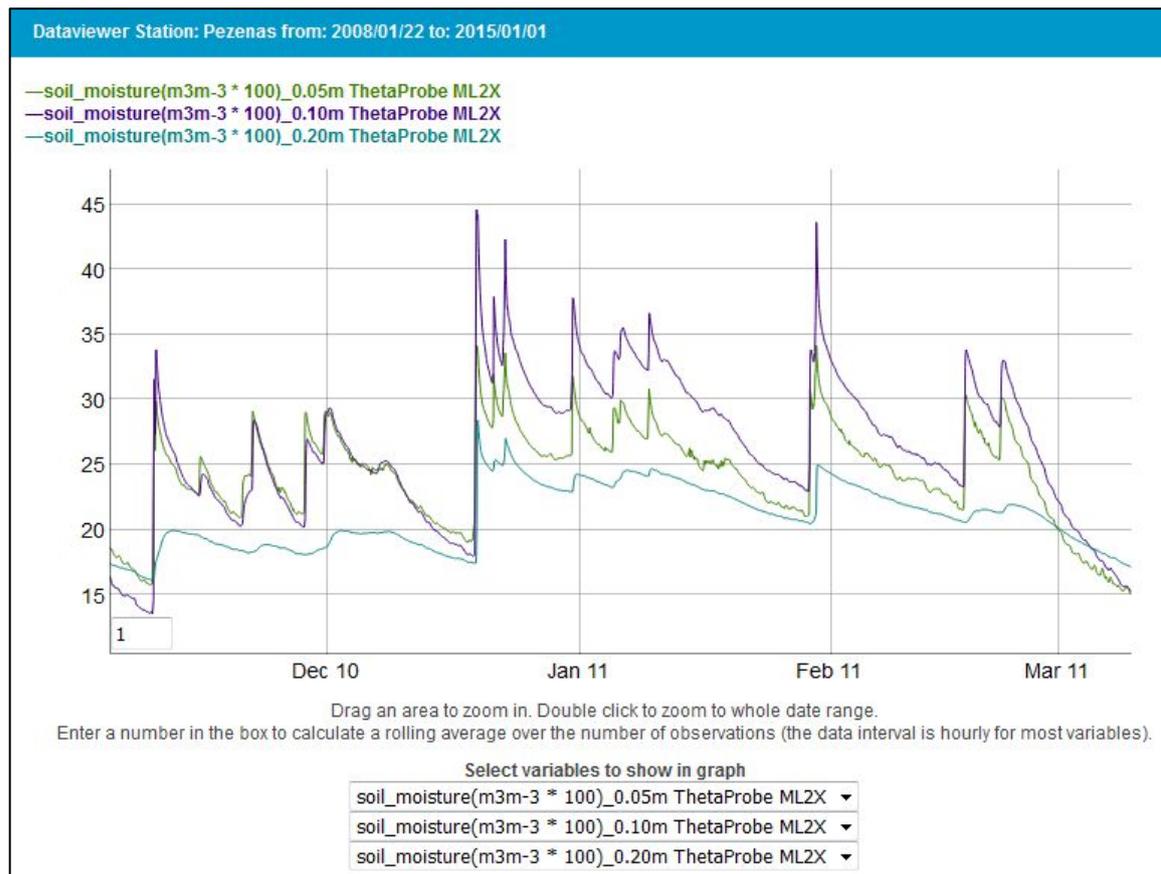


Figure 7: Example of soil moisture measurements at different depths (Station Pezenas of network SMOSMANIA)

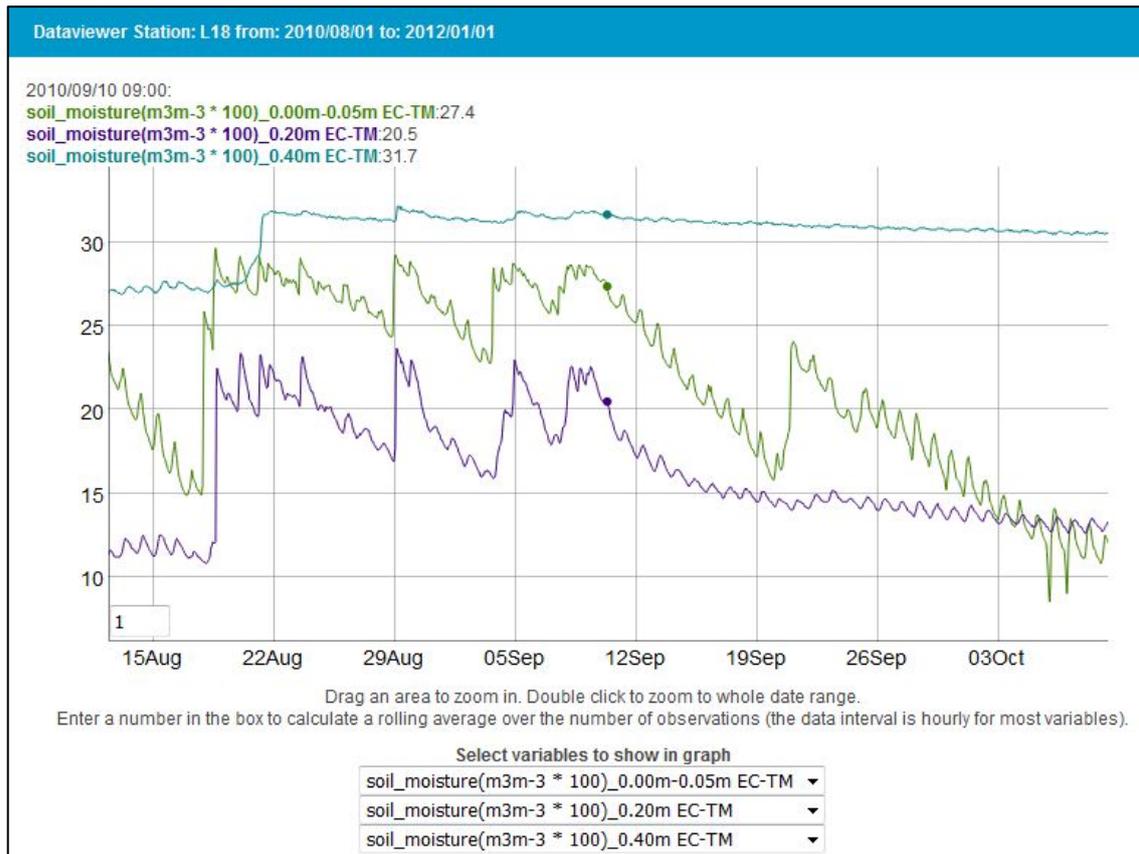


Figure 8: Example of temperature sensitivity of soil moisture at different depths. The diurnal cycle of temperature is visible within the soil moisture readings, well pronounced at the top layer and barely visible in the deepest layer. (Station L18 of network CTP_SMTMN)

Thus, it becomes very clear that soil moisture cannot be described by simply one typical shape or behavior, but that several characteristics are possible. In fact, soil moisture regimes may vary strongly depending on the prevalent climate, vegetation and soil composition. Ideally, quality control algorithms should be able to cope with these unavoidable natural and artificial phenomena. In reality, a trade-off has to be made to identify suspicious measurements without over-flagging natural events.

2.3 Potential erroneous events within soil moisture readings

Since the quality flags considered in this thesis are based only on the spectrum of a soil moisture time series, additional variables such as precipitation, air temperature or soil temperature are not used in this approach.

A variety of suspicious measurements can be observed caused by malfunction of the sensor, irresponsiveness of the sensor, a problem of energy supply, or a connection problem while writing on the data logger. Driven by these causes, several erroneous shapes may occur within the soil moisture time series, which can be generalized into three categories: jumps, spikes, and constant values.

2.3.1 Jumps

By the term “jump” a sudden, from one timestamp to the next, rise or fall in the measurements is meant, referred to as “positive jump” (Figure 9) or “negative jump” (Figure 10), respectively. As already described above (Chapter 2.2), the drying process of soil normally performs slowly and may take some hours or even days. Thus, a sudden drop of the level of soil moisture cannot occur naturally. A sudden rise in the amount of soil moisture, however, is normally the result of a precipitation event. Thus, an artificial “positive jump” can only be identified if the shape following the jump does not reflect the natural drying behaviour of soil moisture.

Problematic is the fact that jumps may lead to a general offset of the amount of soil moisture and it is not possible with the described spectrum-based approach to decide which period describes the actual soil moisture level more realistically: the period before or after the jump.

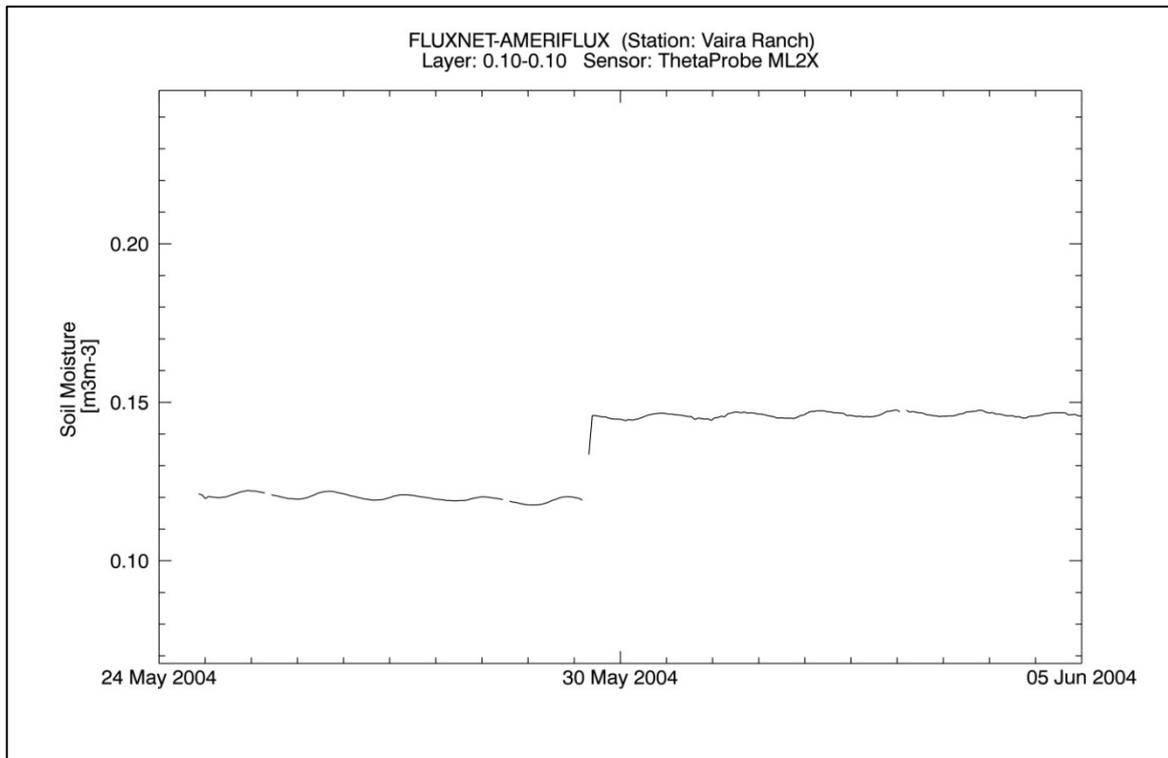


Figure 9: Example of a soil moisture time series containing a positive jump. Also noticeable are missing values, their effect on the flagging result will be discussed in chapter 4. (Station Vaira Ranch of network FLUXNET-AMERIFLUX)

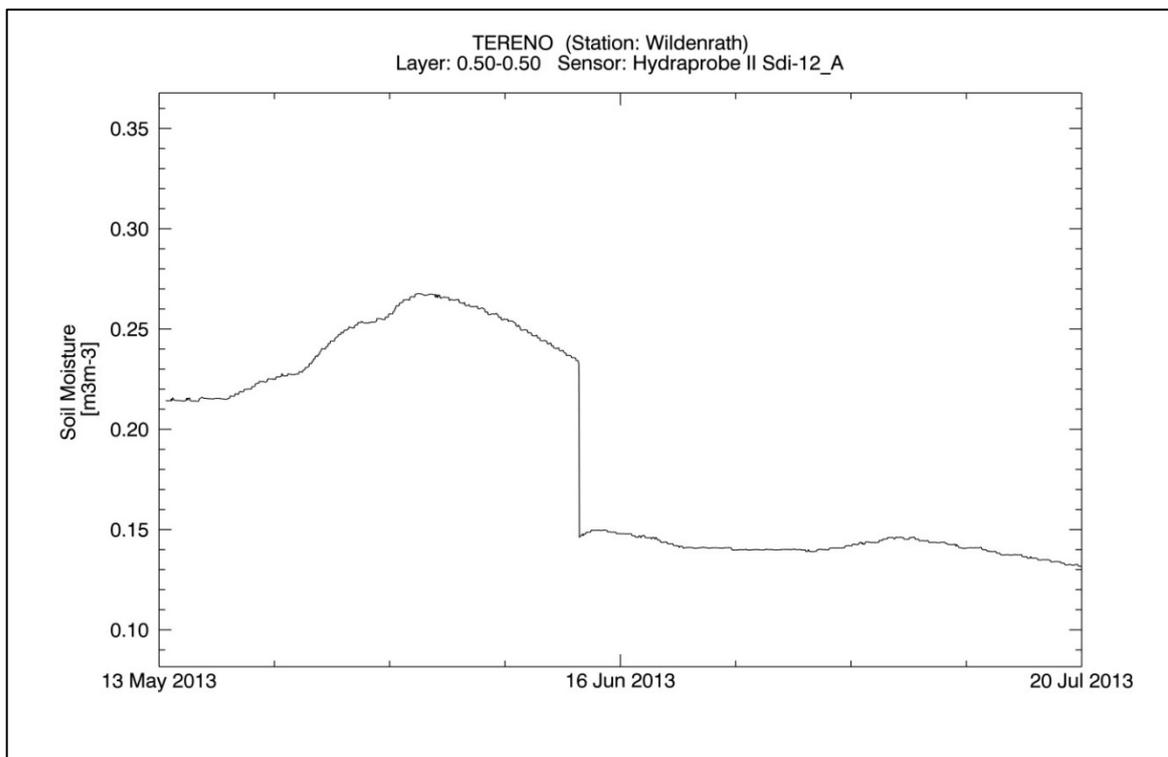


Figure 10: Example of a soil moisture time series containing a negative jump (Station Wildenrath of network TERENO)

2.3.2 Spikes

A “spike” can be understood as an outlier value, represented by a single measurement which lies significantly above or beneath the surrounding measured values of soil moisture (Figure 11). Thus, a “spike” is typically only a short-time event, lasting only for a single unit of time.

Spikes are typically caused by a temporary sensor drop out or energy supply shortage. They are unpredictable and appear in both positive and negative directions.

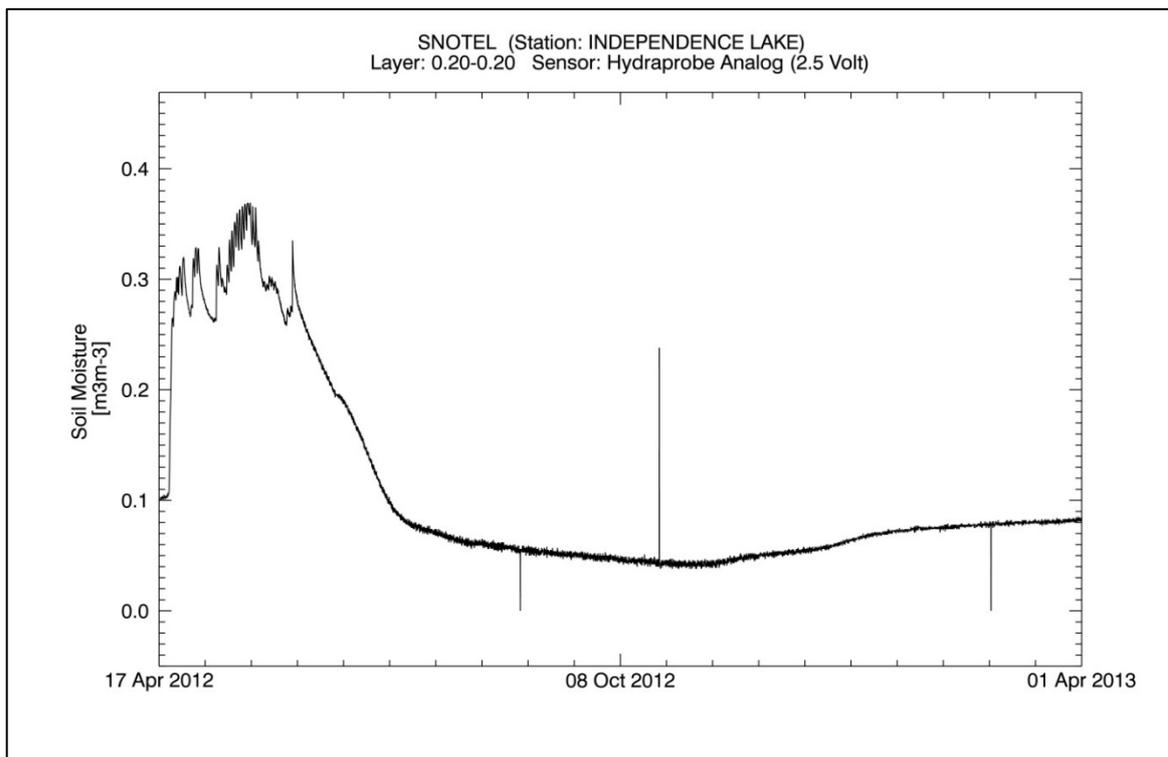


Figure 11: Example of a soil moisture time series containing positive and negative spikes (Station INDEPENDENCE LAKE of network SNOTEL)

2.3.3 Constant values

Two different kinds of constant values may occur.

First, constant values can occur if the soil is saturated due to consecutive precipitation events and the sensor is not able to represent any amount of water in the soil beyond that level. This behavior can be often observed within soil moisture time series obtained from capacitance probes (Mittelbach, et al., 2011) and this will be referred to as “saturated plateau” (Figure 12).

Second, constant values may occur after a sudden drop (“negative jump”) of soil moisture, for instance after an energy supply problem or when the soil is frozen. After such a negative jump the reading usually stays at an unreasonable low level until the energy supply is restored and soil moisture variations can be detected again by the sensor. This event will be referred to as “low level plateau” (Figure 13).

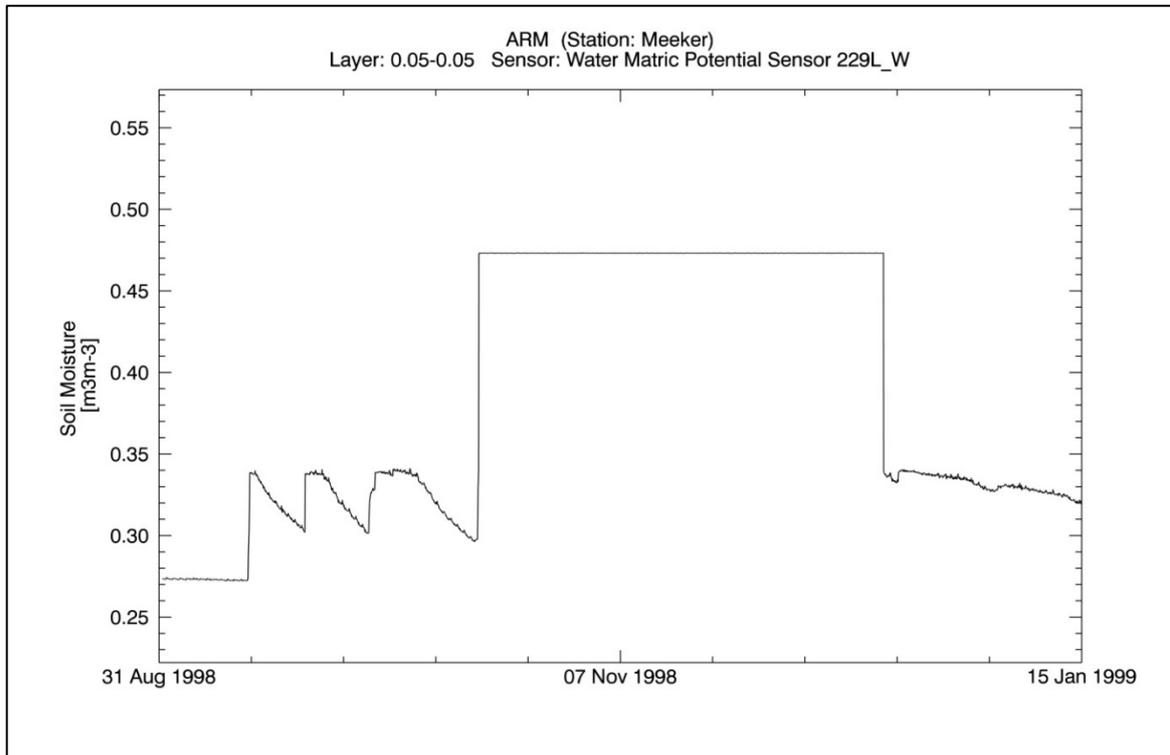


Figure 12: Example of a soil moisture time series containing a saturated plateau (Station Meeker of network ARM)

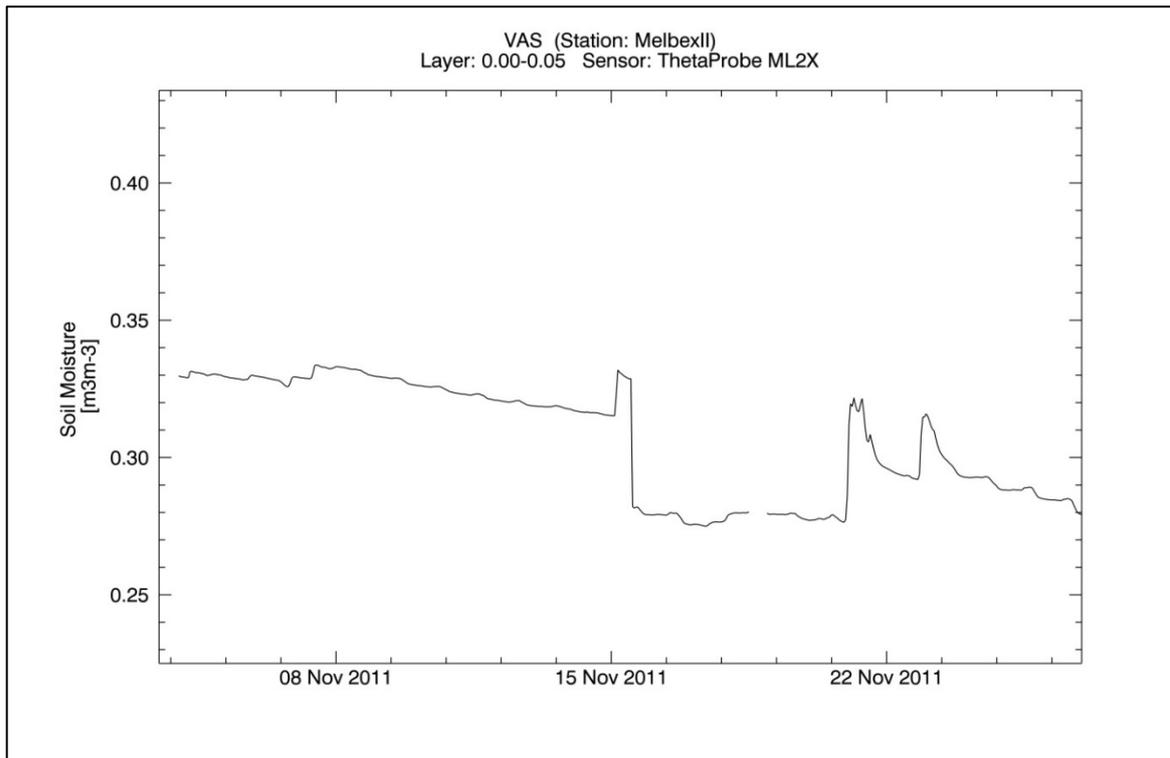


Figure 13: Example of a soil moisture time series containing a low level plateau (Station MelbexII of network VAS)

3 Methods

Since the quality control procedures should be valid for all kind of different sensors and for various measurement depths, the quality control algorithms have to work for different kind of data. Thus, the solution has to be a trade-off between a simple method, ensuring an easy implementation within the ISMN processing chain and a sophisticated method which is able to flag as much erroneous measurements as possible.

3.1 Savitzky-Golay Filter

The author's initial idea for the detection of certain dubious events within a soil moisture time series was to use the extreme slope and the occurrence of discontinuity points during such events. Since the derivative at a point represents the slope of the tangent of the underlying function, the usage of derivatives seemed natural. On the other hand, a smoothing of the underlying observations was desired to eliminate noise while preserving the natural variations of soil moisture. During testing of several algorithms, the widely-known Savitzky-Golay filter (Savitzky, A. and Golay, M.J.E., 1964) was found to fulfill the requested purpose of smoothing and numerical differentiation. The construction of the filter is based on a window-wise least squares fit by a polynomial of a fixed degree. The resulting smoothed output is taken at the center of the window (Persson & Strang, 2003). The advantage of this method is not only the ability to get a smoothing filter and derivatives with the same equation and only changed parameter values, but the preservation of higher moments (Flannery, Teukolsky, & Vetterling, 1992). In comparison to other typical filters, the smoothed signal from this filter has no delay, which means that peaks in the signal are not shifted. Short periods of missing values can be handled well by the Savitzky-Golay filter (Eilers, 2003). In addition, the computation is simple, as the 'smoothing coefficients' are applied to the data by a convolution, and therefore fast, i.e., suited for NRT processing.

One requirement for the application of the Savitzky-Golay filter is that the underlying data is available for equidistant time steps. As all datasets within the ISMN are harmonized to hourly data, this requirement is fulfilled.

Four parameters have to be defined when using the Savitzky-Golay filter:

- the number of points to the left and
- the number of points to the right of each data point to be included in the filter,

- the order of the derivative and
- the degree of the polynomial used for smoothing.

The smoothness of the resulting graph can be controlled with the number of included filter points. The larger the number of filter points, the smoother is the resulting graph, while a small number of filter points preserves peaks and spikes.

If the smoothed time series itself is of interest, the order of the derivative has to be set to zero.

The degree of the polynomial should be small for smoother results, while higher degrees may give a noisier result but will reduce the filter bias. Values from two to four are typical for applications. The degree must be less than the filter width, which is equal to the sum of the number of the used filter points and one for the data point.

In order to evaluate the feasibility of using one certain set of parameters for the detection of all three types of spurious measurements, a series of parameter values was tested during the occurrence of all these suspicious events. The impact of different parameter settings was tested for several different datasets and is shown for the example of a low level plateau, starting with a negative jump and ending with a positive jump, at the station "Earlsboro" of the network "ARM" (Figure 14 - Figure 17).

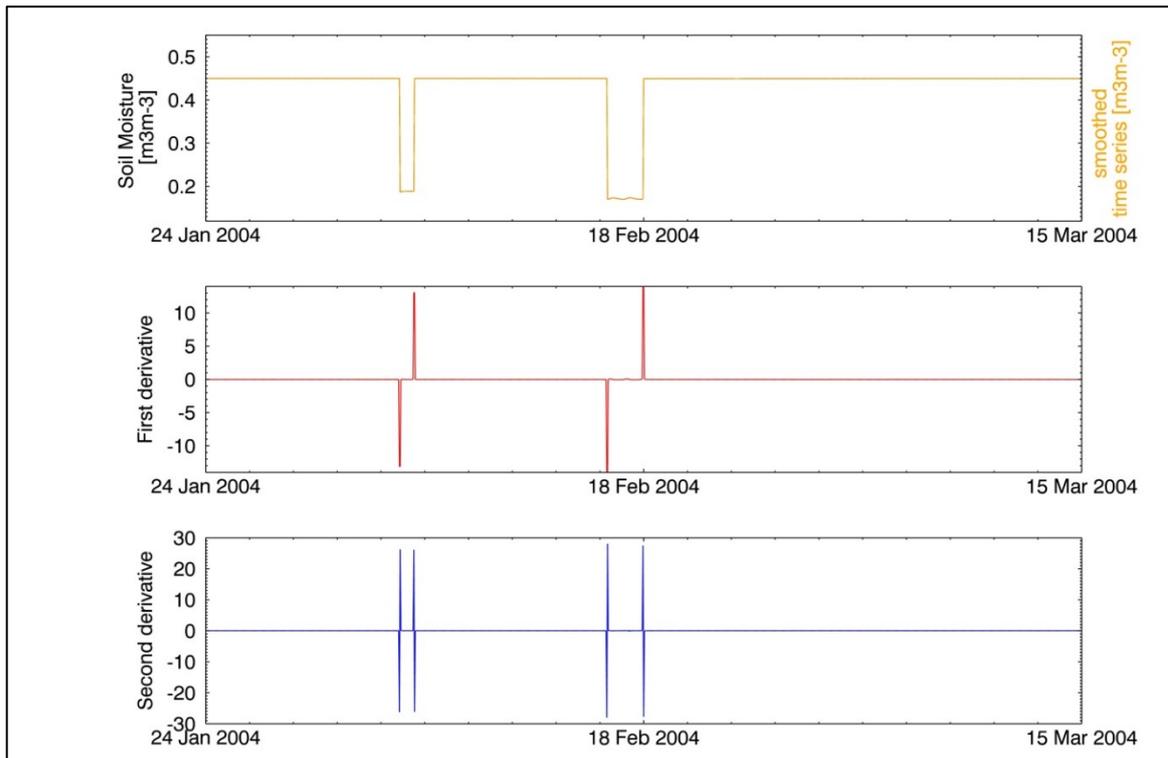


Figure 14: Shape of smoothed time series (top), first (middle) and second (bottom) derivative resulting from Savitzky-Golay filtering with parameters: symmetrical filter width of 3, 2nd order polynomial (Station Earlsboro of network ARM, Water Matric Potential Sensor 229L_W, 0.15-0.15m depth)

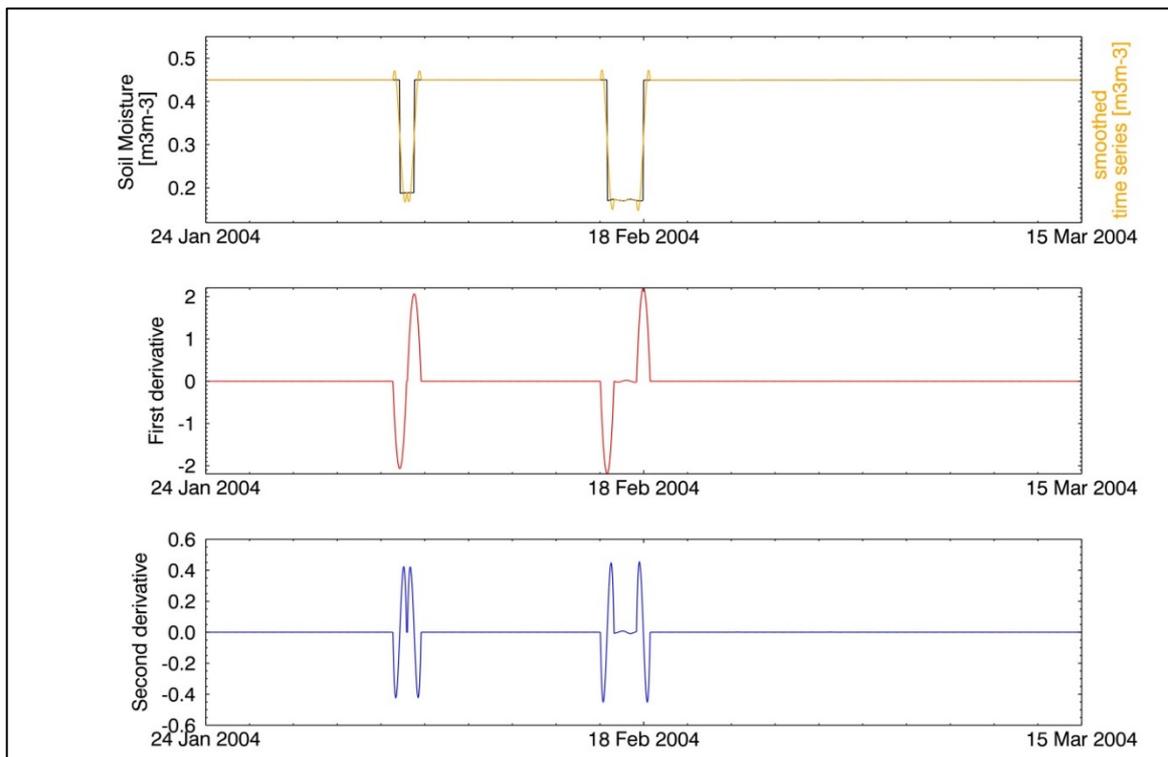


Figure 15: Shape of smoothed time series (top), first (middle) and second (bottom) derivative resulting from Savitzky-Golay filtering with parameters: symmetrical filter width of 19, 2nd order

polynomial (Station Earlsboro of network ARM, Water Matric Potential Sensor 229L_W, 0.15-0.15m depth)

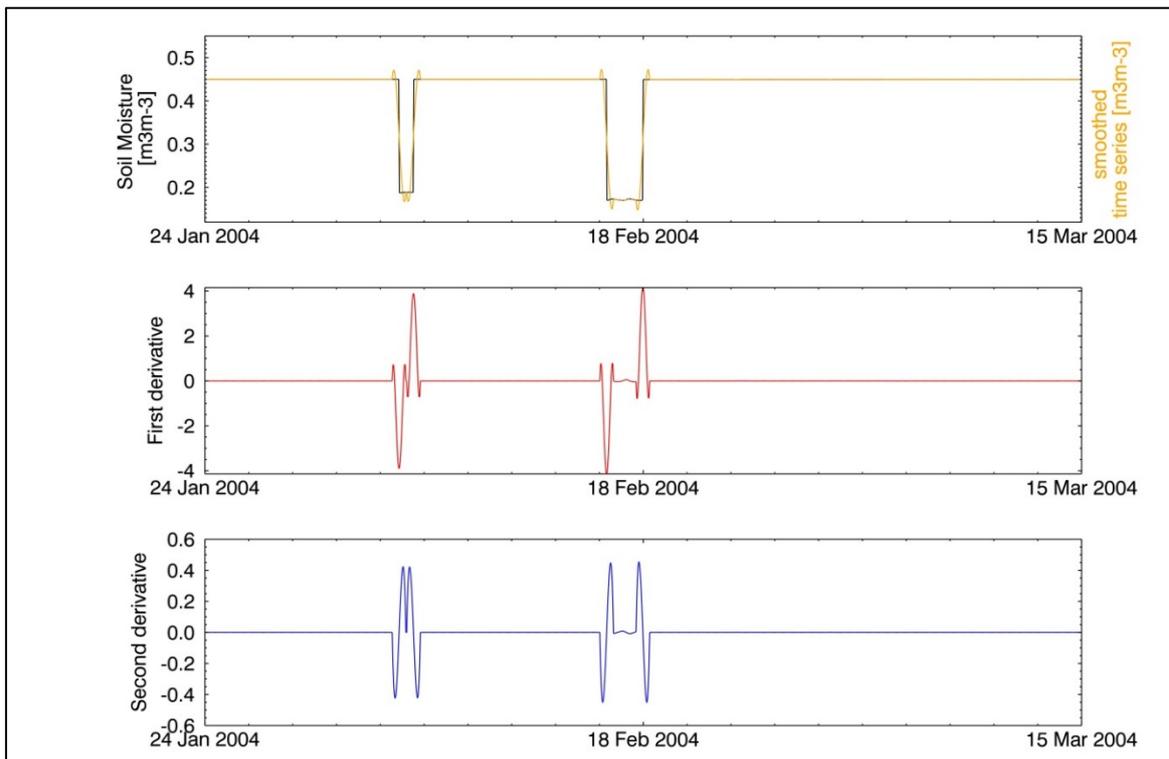


Figure 16: Shape of smoothed time series (top), first (middle) and second (bottom) derivative resulting from Savitzky-Golay filtering with parameters: symmetrical filter width of 19, 3rd order polynomial (Station Earlsboro of network ARM, Water Matric Potential Sensor 229L_W, 0.15-0.15m depth)

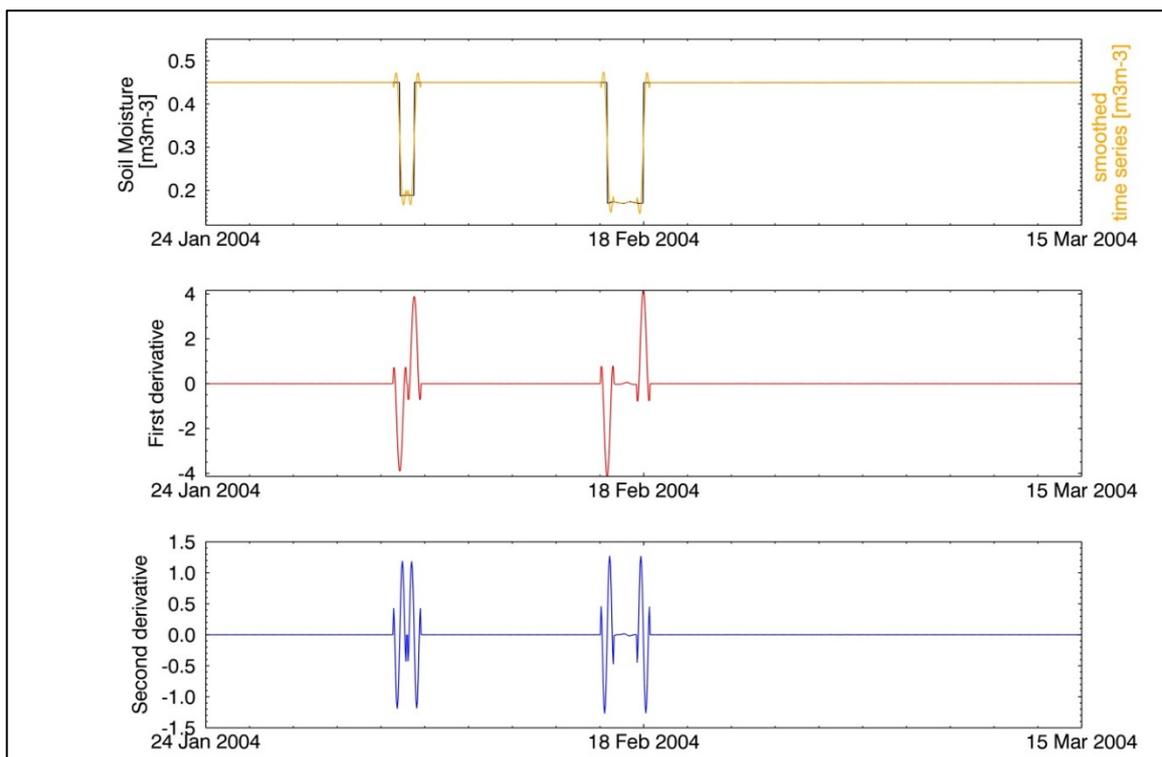


Figure 17: Shape of smoothed time series (top), first (middle) and second (bottom) derivative resulting from Savitzky-Golay filtering with parameters: symmetrical filter width of 19, 4th order polynomial (Station Earlsboro of network ARM, Water Matric Potential Sensor 229L_W, 0.15-0.15m depth)

3.2 Algorithms for Detection

Since the goal is to detect spurious events such as spikes, jumps and plateaus by examining the first and second derivative, a clear appearance of such events is desired in both derivatives. Based on the above demonstrated evaluation, the parameter set with a small symmetrical filter width of three data points and a second degree polynomial was chosen. Only for the detection of saturated plateaus the filter width was enlarged to 25 data points, which will be discussed later in this section.

In the following, the appearance of each type of suspicious event within the first and second derivative will be shown. Based on these characteristics, consecutive conditions have been developed through empirical investigation by using the datasets in Table 3 as calibration datasets.

Table 3: Datasets and periods selected for developing spectrum-based QC methods.

Network	Station	Sensor	Depth interval (m)	Period	Error types
FMI	Maws	Theta Probe ML2X	0.02-0.02	May 2011 – Jul 2011	Negative breaks, low level plateau
REMEDHUS	H13	Stevens Hydra Probe	0.00-0.05	Jul 2010 – Dec 2011	None
REMEDHUS	J14	Stevens Hydra Probe	0.00-0.05	Feb 2009 – Apr 2009	Spikes
OzNet	K11	Stevens Hydra Probe	0.00-0.05	Dec 2006 – Feb 2007	Breaks
SCAN	2052	Hydra Probe Analog	0.51-0.51	Jul 2008 – Nov 2009	Noisy data
SCAN	2054	Hydra Probe Analog	1.02-1.02	Feb 2002 – May 2012	Saturation plateaus
SCAN	2075	Hydra Probe Analog	0.20-0.20	May 2010 – Nov 2010	Spike
SMOSMANIA	PZN	ThetaProbe ML2x	0.10-0.10	Jun 2009 – Jul 2009	Breaks
SMOSMANIA	SVN	ThetaProbe ML2X	0.10-0.10	May 2010 – Dec 2010	Saturated plateaus, breaks, noisy data
SWEX_POLAND	P2	D-LOG-mpts_D	0.10-0.10	Nov 2008 – Dec 2008	Breaks, noisy data
SWEX_POLAND	P3	D-LOG-mpts_C	0.10-0.10	Aug 2009 – Sept 2009	Breaks, Spikes
UDC_SMOS	80	EC5 II	0.05-0.05	May 2009 – Mar 2010	Missing values, Spikes, Breaks
UDC_SMOS	501	EC5 I	0.05-0.05	Dec 2010 – Apr 2011	Saturation plateaus
UMBRIA	C.Rigone	EnviroSCAN	0.15-0.25	Apr 2008 – May 2008	saturation plateaus
USDA-ARS	WG	Hydraprobe Analog (2.5 V)	0.00-0.05	Sept 2006 – Jan 2008	Breaks, low level plateaus, spikes

3.2.1 Spikes

A spike is represented by a considerable change in the amount of soil moisture, positive or negative, lasting only for a single measurement. Thus, the first condition for spikes investigates the amount of change in soil moisture between one time step and the next. If soil moisture increases or decreases by at least 15%, which corresponds approximately three times a typical maximum sensor uncertainty, then this event is classified as spike.

$$\frac{x(t)}{x(t-1)} > 1.15 \quad \text{or} \quad \frac{x(t)}{x(t-1)} < 0.85, \quad [1]$$

where x represents the soil moisture observations and t the time step.

Obviously, this condition does not suffice to distinguish a spike from natural precipitation event. The second derivative, calculated by using the Savitzky-Golay filter with a filter width of three time steps and a second order polynomial, is the basis for the second condition. For a negative spike, a characteristic shape of a positive spike surrounded by smaller negative peaks can be observed in the second derivative (Figure 18, bottom). For a positive spike, a strong negative peak is visible in the second derivative.

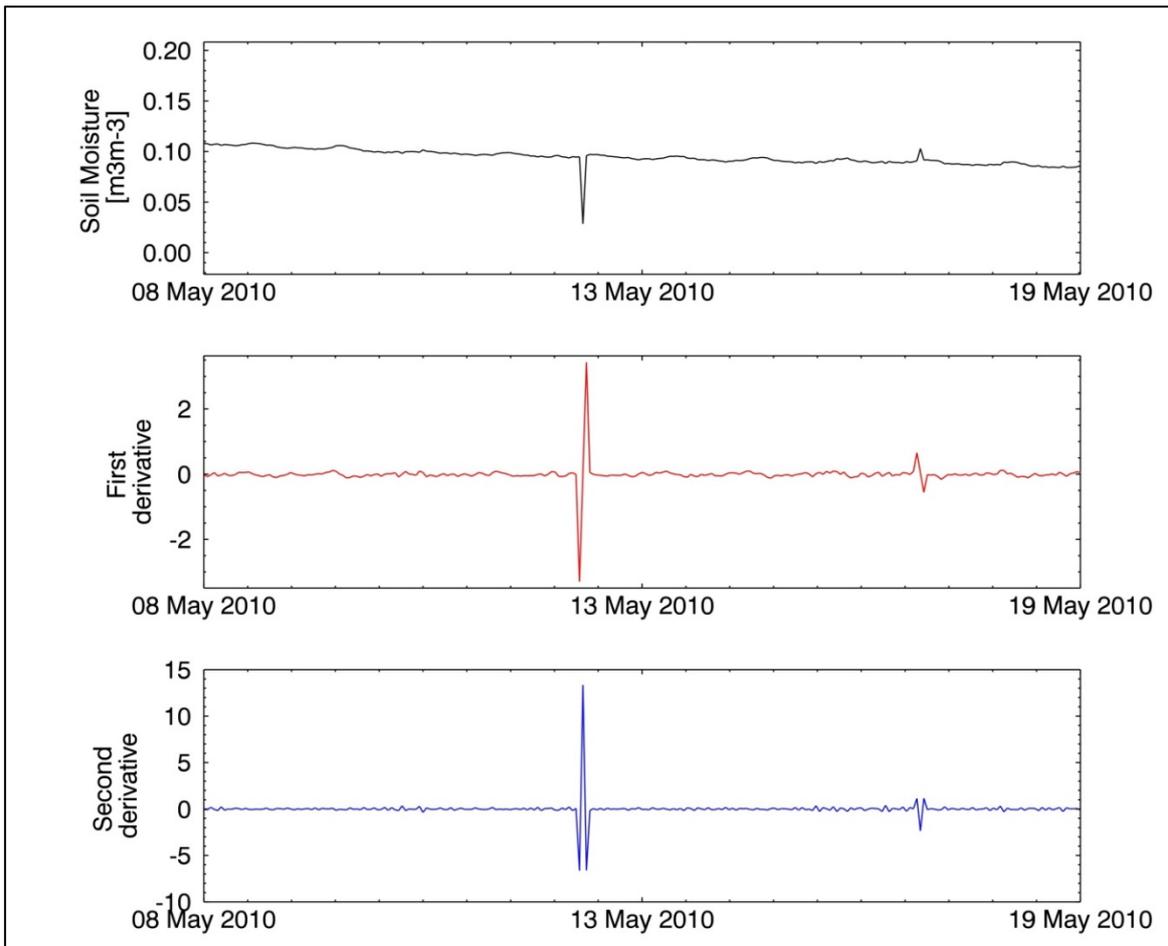


Figure 18: Shape of first (middle) and second derivative (bottom) for a spike within the soil moisture time series (Station Eulo of network OZNET, Stevens Hydra Probe, 0.00-0.05m depth)

The similarity of the actual size of the two lower peaks represents the assumption that the soil moisture value before the spike has almost the same amount as after the spike. Thus, the ratio between those two values is close to unity. Since soil moisture observations are never perfectly stable, this strict assumption needed to be loosened up. Through empirical tests with the calibration data sets (Dorigo, et al., 2013), the ratio was found to vary between approximately 0.8 and 1.2:

$$0.8 < \left| \frac{x''(t-1)}{x''(t+1)} \right| < 1.2 \quad \text{and} \quad x''(t+1) \neq 0 \quad [2]$$

where x'' represents the second derivative and t the time step.

During testing with the calibration data it was experienced that the criteria in [2] does not perform sufficiently for noisy data. Therefore, a third condition was added based on the coefficient of variation. The coefficient of variation, also known as relative standard deviation,

is defined as the ratio between the standard deviation σ and the mean μ . Applied to the soil moisture time series to an interval of 12 hours before and 12 hours after the potential spike, but excluding the measurement of the spike itself, the third condition reads as following:

$$\left| \frac{\sigma^2[x(t-12), x(t+12)]}{\mu[x(t-12), x(t+12)]} \right| < 1 \quad [3]$$

where σ^2 stands for the variance and μ the arithmetic mean of the interval $[x(t-12), \dots, x(t-1), x(t+1), \dots, x(t+12)]$. An observation at time step t is flagged as spike if and only if conditions [1] to [3] are fulfilled.

3.2.2 Jumps

Jumps are characterized by a sudden change of the amount of soil moisture from one time step to another. Unlike after a spike, the soil moisture level does not return to its initial state, but remains changed until a certain period of time or the entire following period.

Again, based on the derivatives calculated by the Savitzky-Golay filter, a sequence of conditions was defined to detect positive or negative jumps. Therefore, a measurement at time step t is flagged as a negative (positive) jump if and only if all three conditions are fulfilled:

$$1. \quad \frac{(x(t) - x(t-1))}{x(t)} > 0.1 \quad [4]$$

where $|x(t) - x(t-1)| > 0.01m^3m^{-3}$ and $x(t) \neq 0$.

This means, that the relative change of soil moisture in respect to the previous time step has to be at least 10%. To avoid over flagging the absolute amount of change between the current and the previous soil moisture value has to be at least $0.01m^3m^{-3}$.

$$2. \quad x'(t) > 10 \cdot \frac{1}{25} \cdot \sum_{k=-12}^{k=+12} x'(t+k) \quad [5]$$

This condition expresses the characteristic strong negative (positive) change within the first derivative x' (Figure 19, middle), when a negative (positive) jump appears in the soil moisture time series. This first derivative jump should be at least ten times smaller (larger) than the arithmetic mean of a 25 hour period of the first derivative centered at t .

$$3. \quad \left| \frac{x''(t-1)}{x''(t+1)} \right| = 1 \quad \text{and} \quad \left| \frac{x''(t+1)}{x''(t+2)} \right| > 10 \quad \text{with} \quad x''(t+1) \neq 0 \quad [6]$$

and $x''(t+2) \neq 0$.

As shown on the bottom of Figure 19, a negative (positive) jump leads to a pronounced negative (positive) second derivative value at the time step $t-1$, followed by large positive (negative) value at $t+1$. These two values have approximately the same amount, thus their ratio is about one. Since the second derivative returns to a value around zero at time step $t+2$, the ratio between the second derivatives at $t+1$ and $t+2$ is very high and was found to usually exceed at least a value of 10.

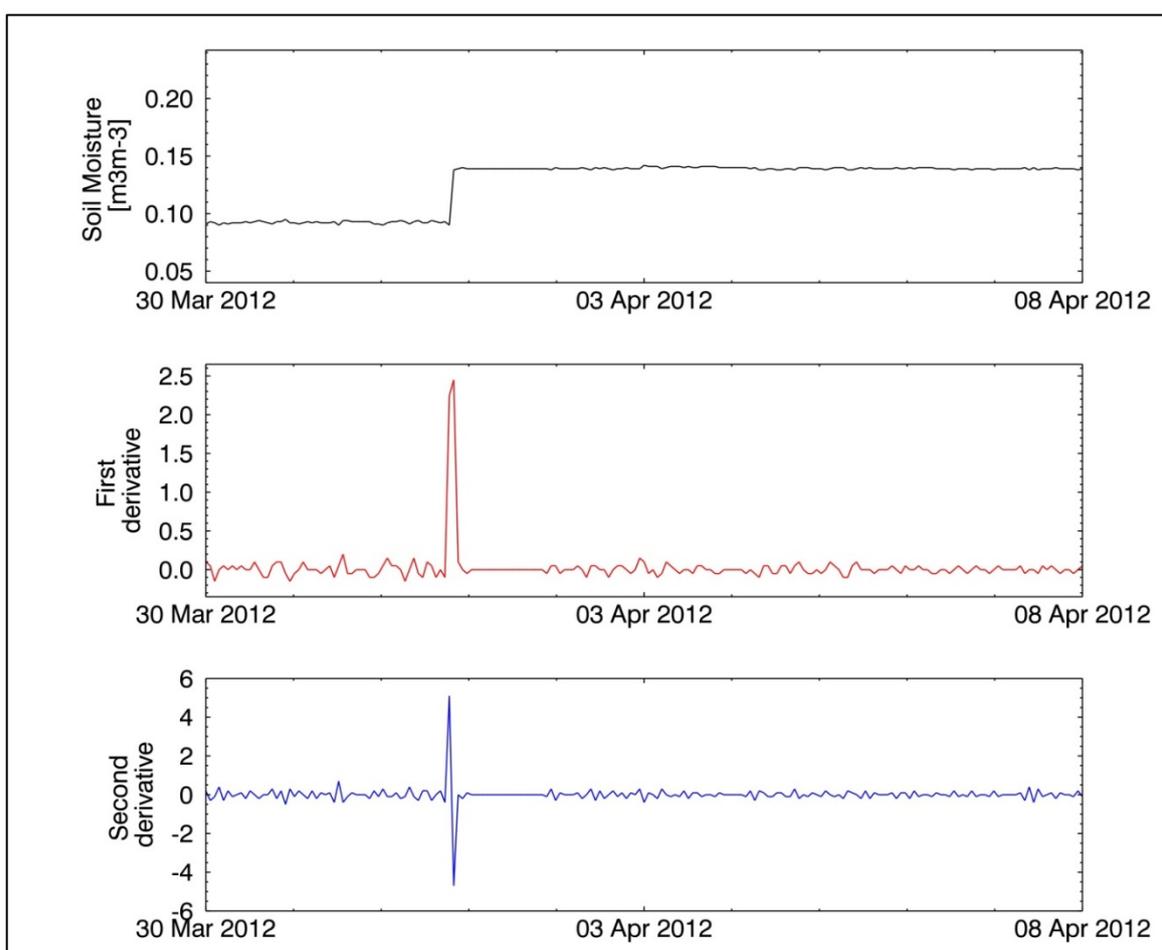


Figure 19: Shape of first (middle) and second derivative (bottom) for a positive jump within the soil moisture time series (Station Kemole Gulch of network SCAN, Hydraprobe Analog (2.5 Volt), 0.50-0.50m depth)

As mentioned above, it is impossible to decide which level of soil moisture represents the actual true soil moisture value better, the level before or after the jump. Therefore, the above described criteria are designed to flag only the jump itself.

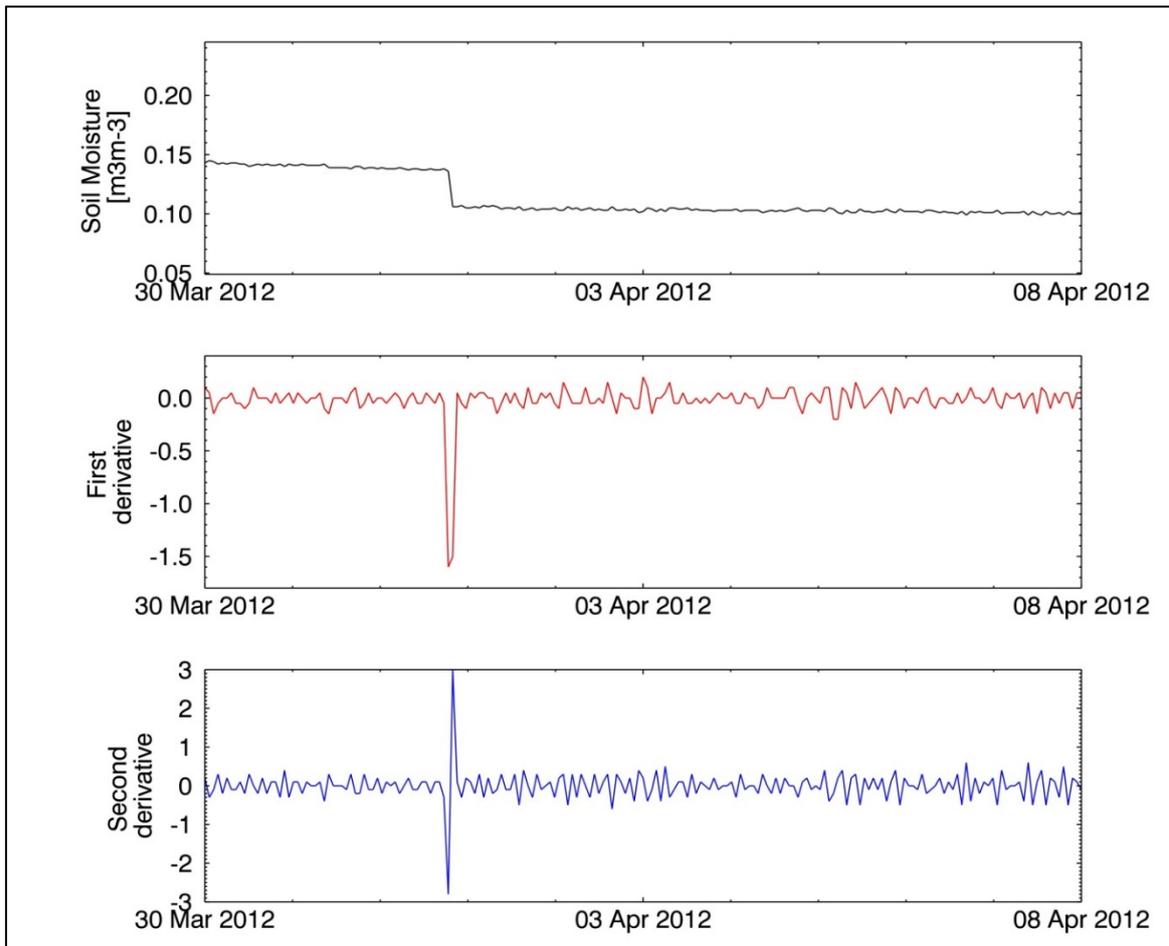


Figure 20: Shape of first (middle) and second derivative (bottom) for a negative jump within the soil moisture time series (Station Kemole Gulch of network SCAN, Stevens Hydra Probe, 0.10-0.10m depth)

3.2.3 Plateaus

Since saturated plateaus and low level plateaus have different characteristics, the quality control procedures for their detection have to be defined separately.

Soil moisture observations are flagged as saturated plateaus if and only if the following conditions are fulfilled:

1. $\sigma^2[x(t-n), x(t+n)] \leq 0.0005 \quad \forall t \text{ and } n \geq 6$ [7]

Equation [7] represents the character of saturated plateaus where a sensor does not respond to further soil moisture changes anymore and its readings remain almost constant. A minimum of allowed residual variation was defined as 1% of a typical average sensor uncertainty of $0.05\text{m}^3\text{m}^{-3}$ for at least a period of 12 hours. The whole data period is examined for this condition.

$$\begin{aligned}
2. \quad t = t_{pl\ start} &\leftrightarrow \exists \max([x'(t - n - 12), x'(t - n + 12)]) \geq 0.0025 \\
t = t_{pl\ end} &\leftrightarrow \exists \min([x'(t - n - 12), x'(t - n + 12)]) \leq 0
\end{aligned} \tag{8}$$

Since saturated plateaus usually occur after one severe or multiple consecutive precipitation events, they are characterized by a strong rise of soil moisture at the beginning and a drop at the end, where the sensor starts reacting to the lower soil moisture values again. These events are visible in the first derivative through a sharp positive peak at the beginning, and a strong negative peak at the end of the plateau (Figure 21, middle). To detect these peaks, threshold values of 0.0025 and 0 have been empirically derived from the calibration data sets.

$$3. \quad \mu([x(t_{pl\ start}), x(t_{pl\ end})]) > \max([x(t_0), x(t_{end})]) * 0.95 \tag{9}$$

By their nature, saturated plateaus consist of the highest recorded soil moisture values of a time series where the sensor is not capable to respond to further soil moisture increases anymore, typically because the sensor voltage reaches its maximum. Thus, equation <> describes the fact, that saturated plateaus only occur at a high soil moisture level, which was defined to be 95% of the maximum value measured during the whole measurement period.

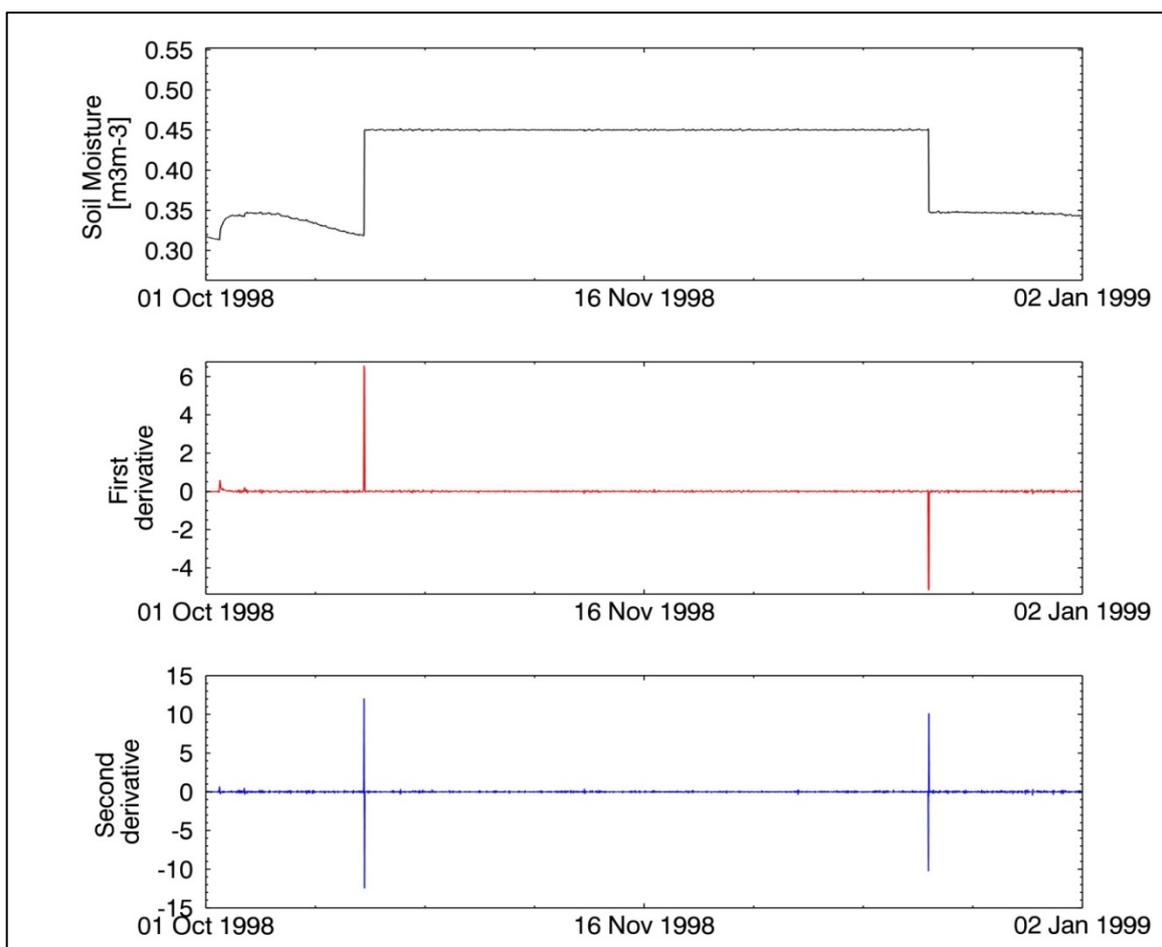


Figure 21: Shape of first (middle) and second derivative (bottom) for a saturated plateau within the soil moisture time series (Station Meeker of network ARM, Water Matric Potential Sensor 229L_E, 0.15-0.15m depth)

Low level plateaus are typically caused by two different events:

1. The soil freezes which results in a drop of apparent soil moisture and therefore very low sensor readings. When the temperature is rising again, fluctuating soil moisture can often be observed, which hampers the detection of a distinct end of the low level plateau.
2. A sensor failure caused for example by a temporary shortage of energy supply results in a negative jump and a subsequent period of abnormally low soil moisture readings. When the energy supply is recovered, the soil moisture measurements return to a normal level, ending the low level plateau with a positive jump (Figure 22).

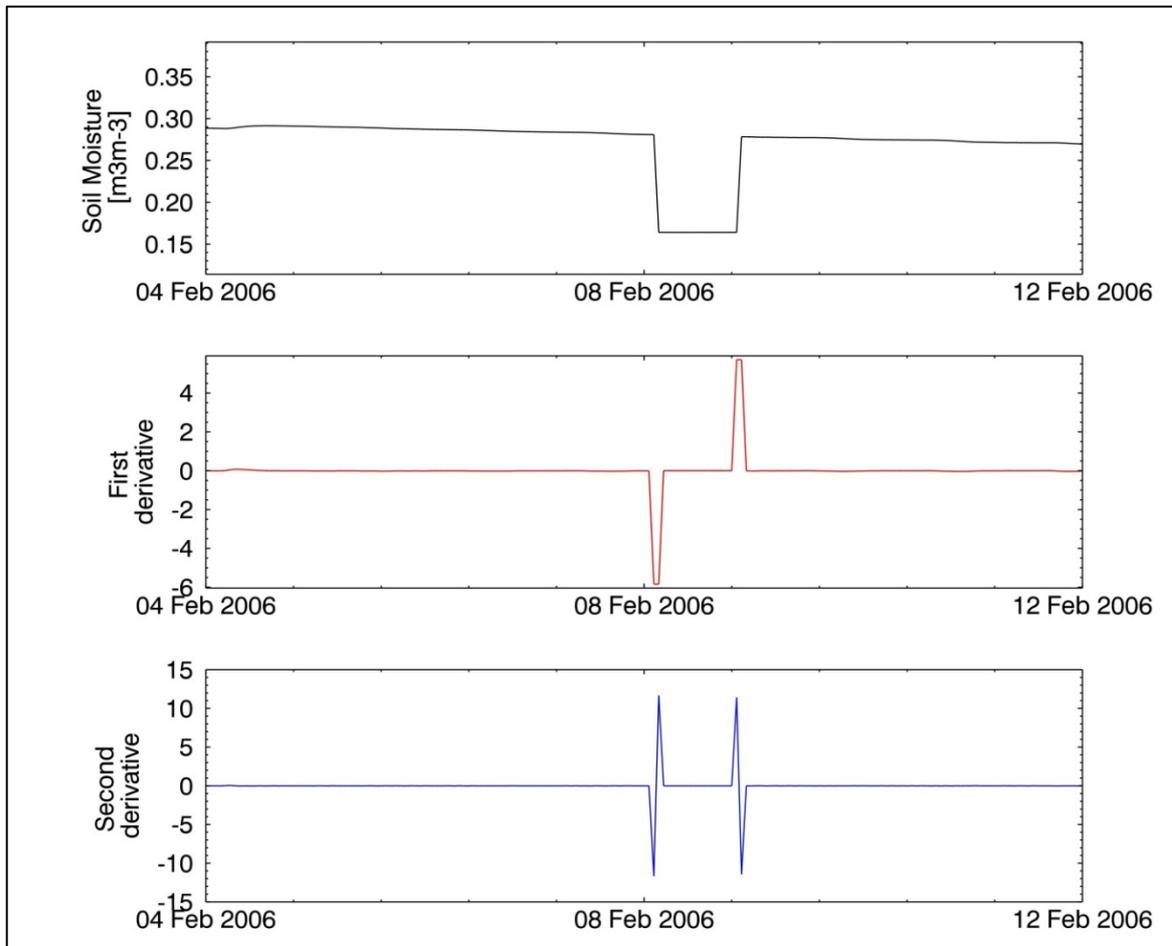


Figure 22: Shape of first (middle) and second derivative (bottom) for a low level plateau within the soil moisture time series (Station Vaira Ranch of network FLUXNET-AMERIFLUX, ThetaProbe ML2X, 0.20-0.20m depth)

Soil moisture observations are flagged as low level plateau if and only if the following conditions are fulfilled:

1. A negative jump, as described above, is detected at time step t .

Since both types of low level plateaus show a preceding negative jump, the existence of such jump is the first condition for the detection of low level plateaus. Subsequent measurements are then flagged as low level plateaus if they satisfy the following condition.

2.
$$\left| \frac{\sigma^2[x(t),x(t+n)]}{\mu[x(t),x(t+n)]} \right| < 0.01 \text{ with } n \geq 12 \quad [10]$$

The coefficient of variation is again used to identify a period of n measurements starting at the negative jump at the time t with a relatively low variation.

4 Results and Discussion

4.1 Results

An evaluation of the performance of the described spectrum based quality checks was already performed in Dorigo et al., (2013) for a small number of selected validation datasets. Therefore, the flags obtained from the proposed quality control procedures were compared to a manual classification of occurring erroneous measurements, presented through a classical error matrix (Congalton & Green, 2009; Hubbard et al., 2005). Since the development of the quality check algorithms a number of new networks were added to the ISMN and existing networks and datasets were updated. Thus, a variety of new datasets exists for an up-to-date performance test to identify if the developed spectrum based quality control procedures still perform sufficiently.

Again, a selection of datasets were drawn from the ISMN and inspected manually to identify true occurrences of positive and negative jumps, spikes, saturated and low level plateaus. 40 datasets from 19 different networks were selected, including different depth layers and sensors types. Also, datasets with a high random noise level as well as data sets without erroneous measurements were selected, in order to further test the developed for errors in omission and commission. The automated quality control procedures were applied to these datasets and the results were compared to the visually flagged observations. A detailed list of the used datasets can be found in Table 3. Preference was given to newly implemented networks and datasets, which did not exist within the ISMN when the quality control procedures were developed. Thus, a more objective evaluation of the performance of the automated quality control algorithms is possible.

Table 4: Datasets and periods selected for evaluating spectrum-based QC methods.

Network	Station	Sensor	Depth interval (m)	Period	Error types
ARM	E12	Water Matric Potential Sensor	0.05-0.05	Aug 1998 – June 1999	Spikes, Jumps, Saturated Plateau, Low Level Plateau
ARM	E20	Water Matric Potential Sensor	0.05-0.05	Sept 1998 – Jan 1999	Saturated plateau, Jumps

ARM	E3	Water Matric Potential Sensor	0.05-0.05	Jul 2006 – Aug 2006	Spikes, Jumps
CTP_SMTMN	M09	5TM	0.40-0.40	Feb 2012 – Dec 2012	None
DAHRA	DAHRA	ThetaProbe ML2X	0.05-0.05	Sep 2012 – Oct 2012	None
FLUXNET- AMERIFLUX	US-Var	ThetaProbe ML2X	0.20-0.20	Jan 2006 – Feb 2006	Jumps, Low Level Plateau
HOBE	3.01	Decagon 5TE_B	0.20-0.25	Dec 2012 – Feb 2013	Saturated Plateaus
HOBE	3.08	Decagon 5TE_A	0.20-0.25	Jan 2013 – Feb 2013	Saturated Plateaus, Jump
HYDROL- NET_PERUGIA	WEEF 2	TDR-SME Corp. TRASE-BE	0.25-0.25	Jul 2012 – Aug 2012	Negative Jump
MAQU	NST_04	ECH20 EC-TM	0.05-0.05	Aug 2009 – Sep 2009	Saturated plateaus, Jumps
ORACLE	CHEVRU	ThetaProbe ML2X	0.06-0.06	Jan 2013 – May 2013	None
OZNET	M3	CS615	0.00-0.08	Mar 2010 – Apr 2010	Positive Jump
OZNET	Y4	Stevens Hydra Probe	0.00-0.05	Mar 2010 – Jul 2010	Spikes
RSMN	15408	5TM	0.00-0.05	Sep 2014 – Sep 2014	Negative Jump
RSMN	15425	5TM	0.00-0.05	Apr 2015 – May 2015	Jumps, Low Level Plateau
RSMN	15434	5TM	0.00-0.05	Apr 2014 – Sep 2014	Spikes
SCAN	2035	Hydraprobe Analog	0.10-0.10	Apr 2010 – May 2010	Noisy Data
SCAN	2053	Hydraprobe Analog	0.20-0.20	Apr 2013 – Jun 2013	Spikes, Jump, Saturated Plateau
SCAN	2091	Hydraprobe Analog	0.10-0.10	Jan 2012 – Apr 2014	Spikes
SCAN	2096	Hydraprobe Analog	0.05-0.05	Oct 2013 – Dec 2013	Noisy Data
SCAN	2103	Hydraprobe Analog	0.05-0.05	Jul 2011 – May 2012	Spikes, Jumps, Saturated Plateau, Noisy data
SCAN	2178	Hydraprobe Digital	0.20-0.20	Feb 2014 – May 2014	Spikes, Jumps, Saturated Plateau, Noisy data

SNOTEL	336	Hydraprobe Analog	0.05-0.05	Oct 2011 – May 2014	Spike
SNOTEL	428	Hydraprobe Analog	0.20-0.20	Jun 2013 – Sep 2013	Spike
SNOTEL	541	Hydraprobe Analog	0.50-0.50	Feb 2012 – Mar 2013	Spikes
SOILSCAPE	node412	EC5	0.20-0.20	Dec 2012 – Jan 2013	Jumps, Saturated Plateau
SOILSCAPE	node512	EC5	0.05-0.05	Feb 2014 – Apr 2014	Spike
SOILSCAPE	node804	EC5	0.15-0.15	Apr 2014 – Jun 2014	Spikes, Jumps
SOILSCAPE	node818	EC5	0.30-0.30	Apr 2014 – Apr 2014	Low Level Plateau
TERENO	ME_BK_001	Hydraprobe II Sdi-12	0.05-0.05	Jan 2014 – Jan 2014	Spike
TERENO	RU_BK_004	Hydraprobe II Sdi-12	0.50-0.50	Jun 2013 – Jul 2013	Negative Jump, Low Level Plateau
TERENO	SE_BK_001	Hydraprobe II Sdi-12	0.05-0.05	Oct 2013 – Oct 2013	Negative Jump
USCRN	Manhattan_6	Stevens Hydrapr. II Sdi- 12	0.10-0.10	Jan 2014 – Feb 2014	Jumps, Low Level Plateau, Noisy Data
USCRN	Monroe_26	Stevens Hydrapr. II Sdi- 12	0.05-0.05	Feb 2015 – Apr 2015	Jumps, Spikes, Saturated Plateau, Noisy Data
USCRN	Montrose_11	Stevens Hydrapr. II Sdi- 12	0.10-0.10	Nov 2012 – Dec 2012	Spikes, Jumps, Saturated Plateau, Noisy Data
USDA_ARS	RC	Hydraprobe Analog	0.00-0.05	Apr 2008 – Jul 2008	Low Level Plateau, Jumps
VAS	Melbex_II	ThetaProbe ML2X	0.00-0.05	Oct 2011 – Dec 2012	Low Level Plateau, Negative Jump
WEGENERNET	50	pF-Meter	0.30-0.30	Apr 2011 – Jul 2011	Saturated Plateaus, Jumps
WEGENERNET	99	pF-Meter	0.30-0.30	Sep 2011 – Nov 2011	Low Level Plateau, Negative Jump
WSMN	3	CS615	0.02-0.02	Oct 2013 – Nov 2013	Jumps

The performance of the automated quality control procedures applied to the new validation data sets (Table 4) is described in a classical error matrix presented in Table 5. The columns

“erroneous measurements” and “correct measurements” stand for the data which was visually classified as erroneous or correct, respectively. The columns “erroneous” and “correct” represent the percentage of observations which were classified as erroneous or correct by the automated quality control procedures. The overall percentage of flagged observations with respect to the whole available data can be found in the column “flagged observations”. It is clearly visible that the overall percentage of flagged data is very low, not even reaching 10% of the observed datasets. As some of the erroneous events represent only occurrences of a single incident this outcome does not surprise. Also the percentages of correct observations detected as erroneous measurements by the automated quality control procedures are extremely low and lie under 2%. The detection accuracy of erroneous observations varies between 42 and 92%.

Table 5: Results of the flagging performance (all values given in percentage).

Flagging results	Erroneous measurements		Correct measurements		Flagged observations
	,Erroneous‘	,Correct‘	,Erroneous‘	,Correct‘	
Spikes	80.0	20.0	0.0	100.0	0.0
Positive Jumps	41.5	58.5	0.0	100.0	0.0
Negative Jumps	57.4	42.5	0.0	100.0	0.0
Low Level plateaus	59.6	40.4	1.4	98.6	2.8
Saturated Plateaus	92.2	7.7	1.6	98.4	6.3

4.2 Discussion

4.2.1 Spikes

The percentage of detected spikes is with eighty percent the second highest result of the flagging performance test. If clear peaks are present in the datasets, the algorithm for spike detection works very well (Figure 23 - Figure 26). Problems occur when the level of random noise in the time series is high or when spikes are not isolated. Due to the properties and the used parameter set of the Savitzky-Golay filter the noise of the original time series is transferred to the derivatives (Figure 27). The characteristic shape of the spike is superimposed by the noise in the derivatives and consequently the detection method fails (Figure 28). This is also the case if consecutive spikes are present in the soil moisture time series (Figure 30).

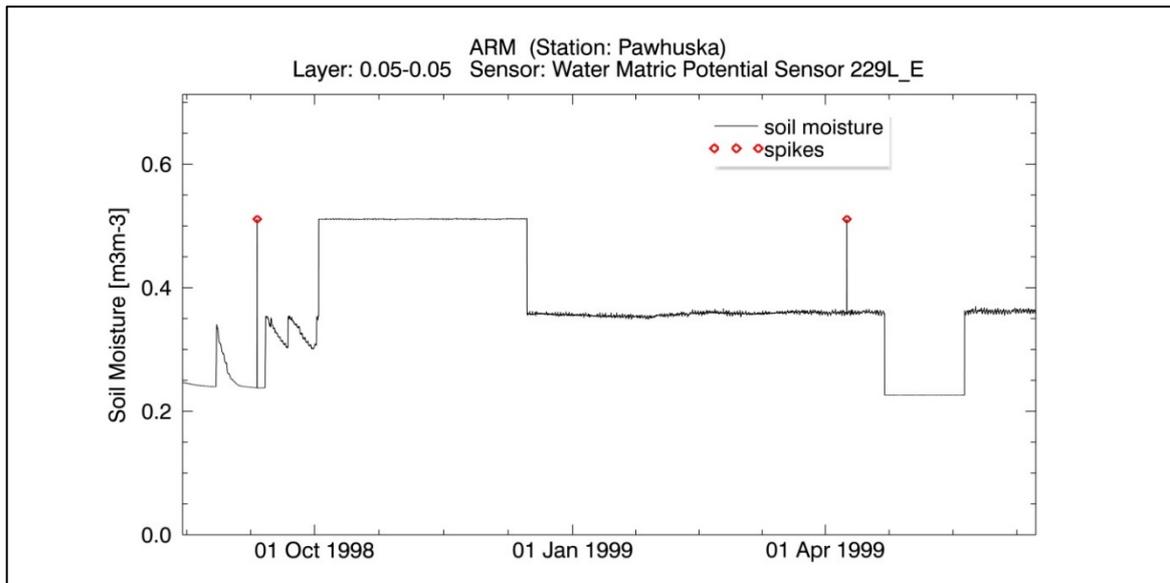


Figure 23: Example for flagging results of a spikes (Station Pawhuska of network ARM, Water Matric Potential Sensor 229L_E, 0.05-0.05m depth)

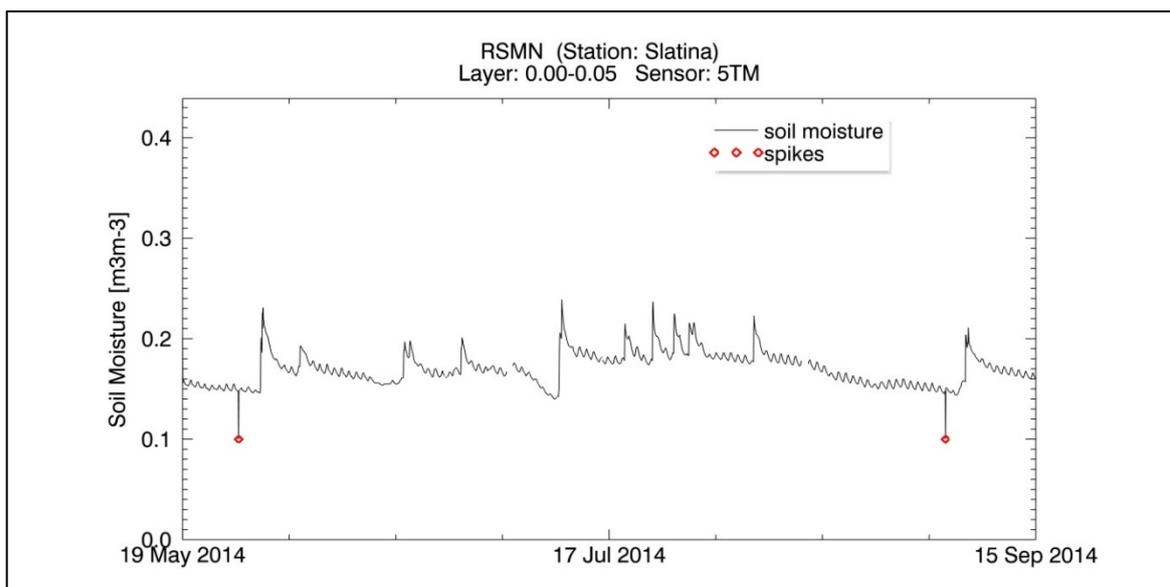


Figure 24: Example for flagging results of spikes (Station Slatina of network RSMN, 5TM, 0.00-0.05m depth)

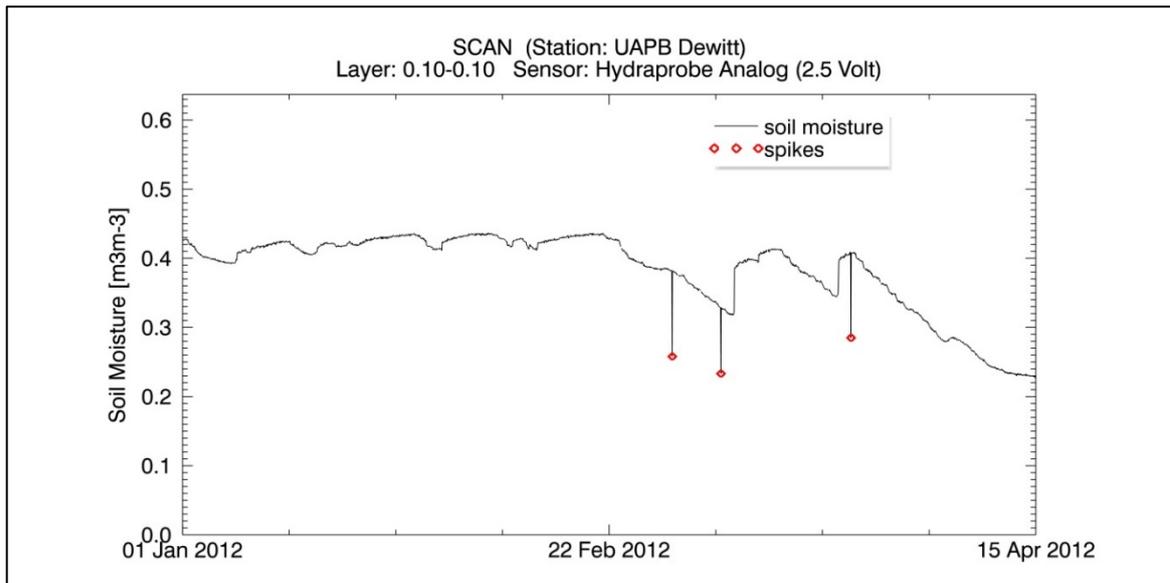


Figure 25: Example for flagging results of spikes (Station UAPB Dewitt of network SCAN, Hydraprobe Analog (2.5 Volt), 0.10-0.10m depth)

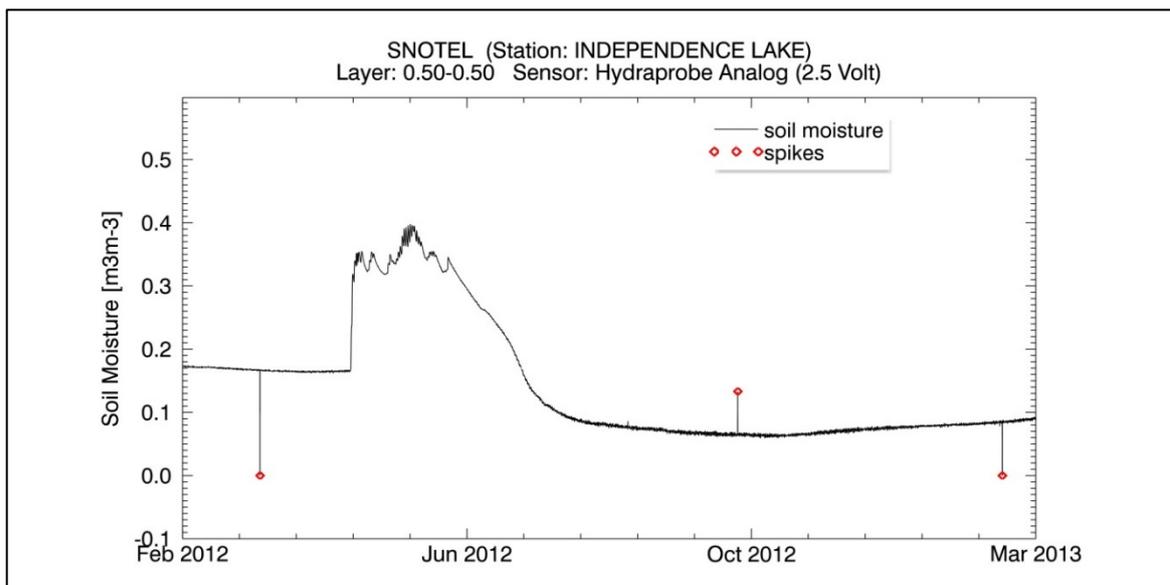


Figure 26: Example for flagging results of spikes (Station INDEPENDENCE LAKE of network SNOTEL, Hydraprobe Analog (2.5 Volt), 0.50-0.50m depth)

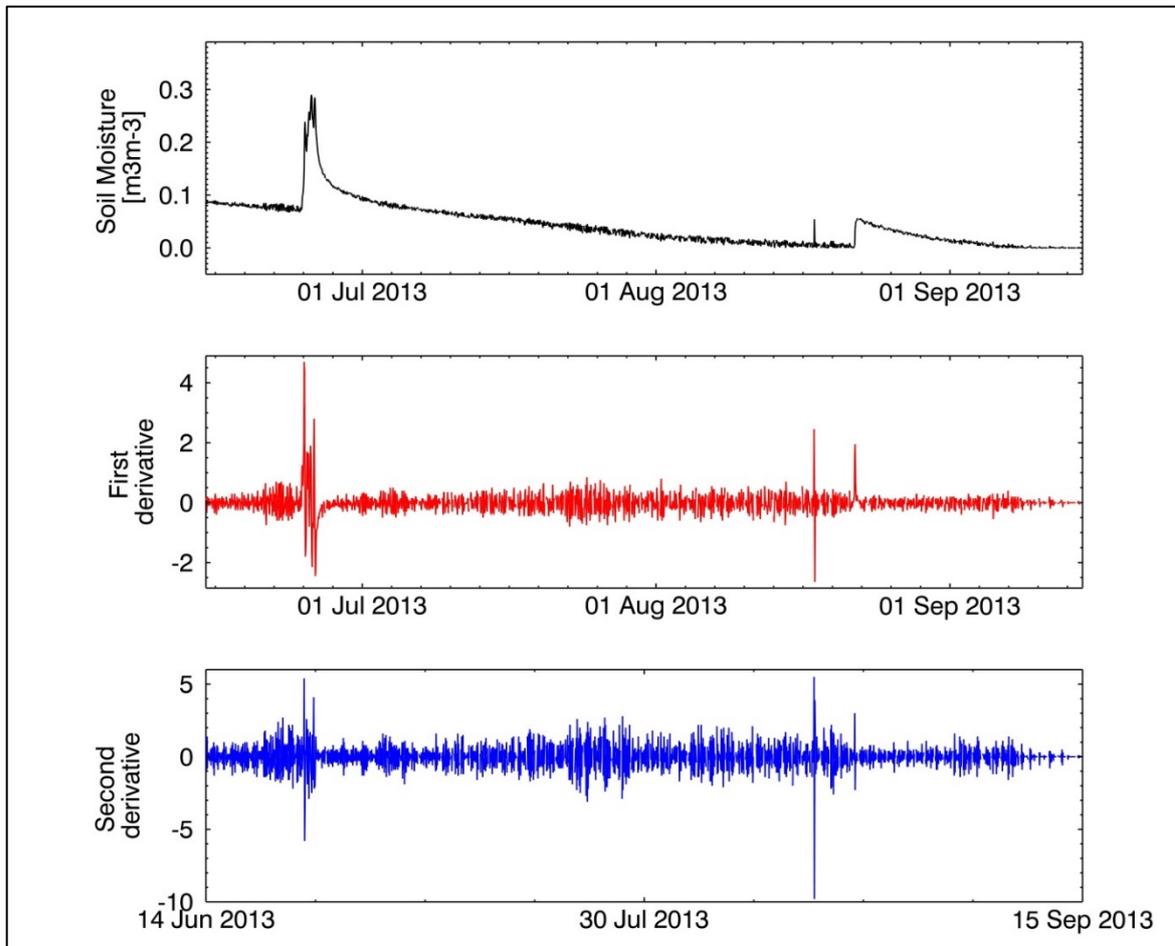


Figure 27: Shape of first (middle) and second derivative (bottom) for a soil moisture time series with random noise (Station CSS LAB of network SNOTEL, Hydraprobe Analog (2.5 Volt), 0.20-0.20m depth)

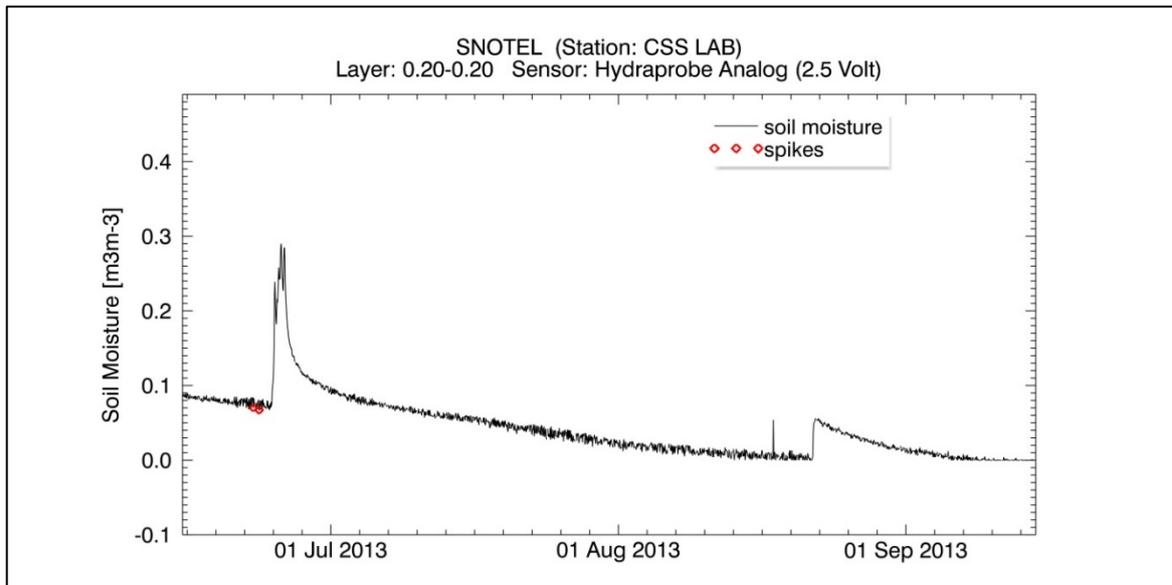


Figure 28: Instead of the spike in late August 2013 two correct but noisy measurements were flagged using the spike detection algorithm (Station CSS LAB of network SNOTEL, Hydraprobe Analog (2.5 Volt), 0.20-0.20m depth)

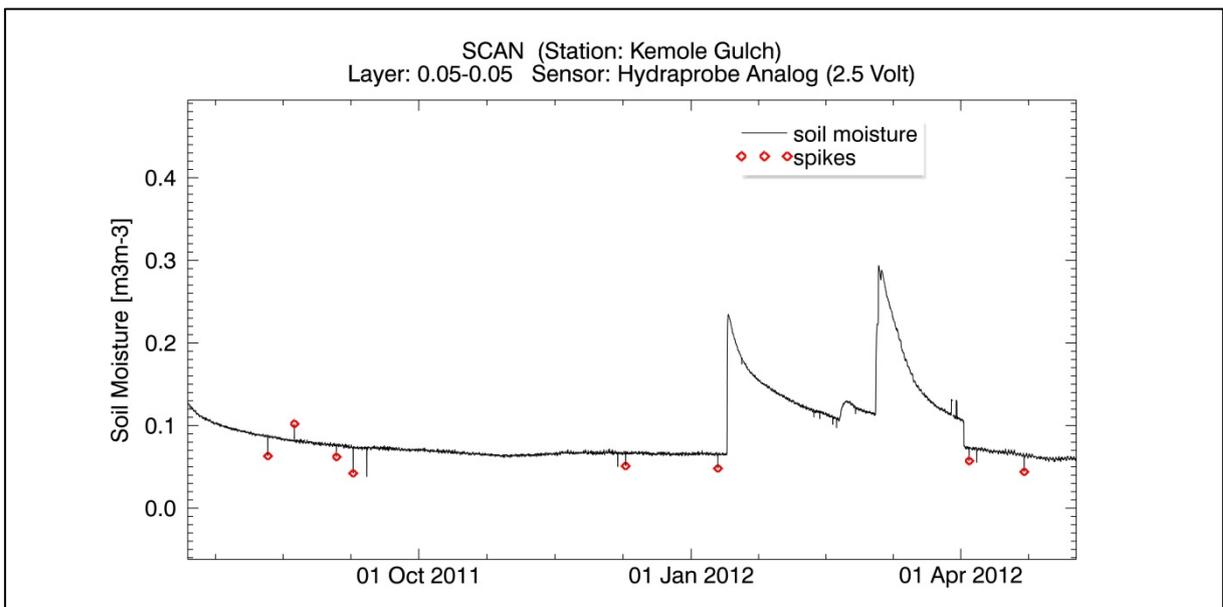


Figure 29: Due to noisy data not all of the spikes were correctly detected by the automatic quality control procedure (Station Kemole Gulch of network SCAN, Hydraprobe Analog (2.5 Volt), 0.05-0.05m depth) erroneous events (Station node828 of network SOILSCAPE, EC5, 0.30-0.30m depth)

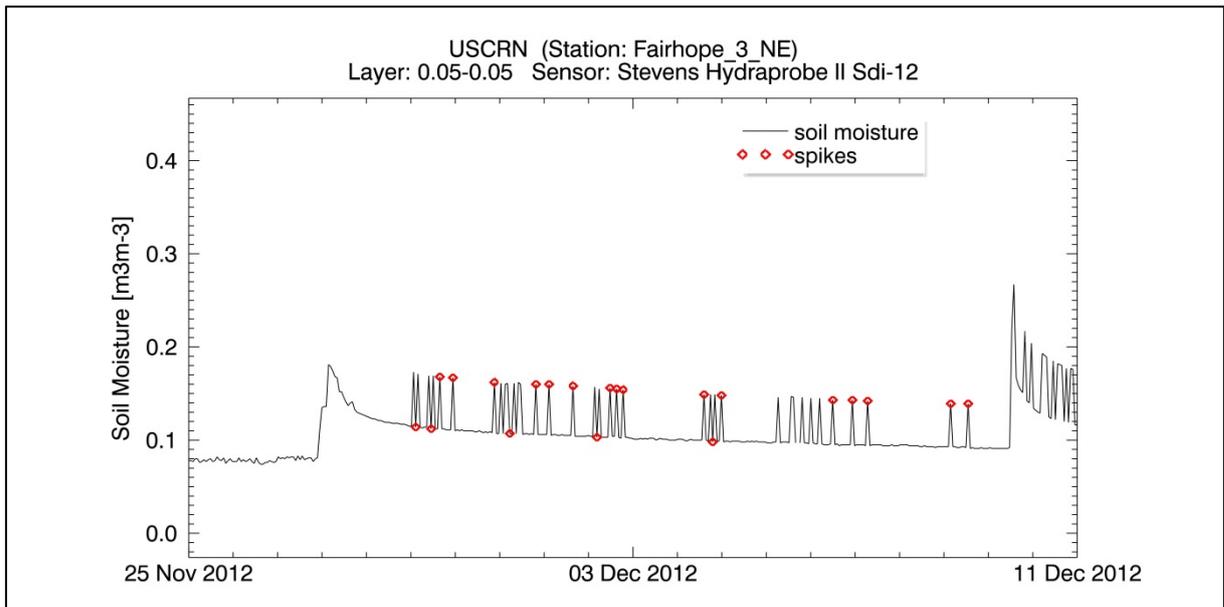


Figure 30: Example for a flagging result of spikes for a soil moisture time series with consecutive erroneous events (Station Fairhope_3_NE of network USCRN, Stevens Hydraprobe II Sdi-12, 0.05-0.05m depth)

4.2.2 Positive Jumps

The detection rate for positive jumps is relatively low, only 42%. This is not surprising, as it is very challenging to distinguish jumps from sudden high rises in soil moisture caused by precipitation events (Figure 35). Sometimes even visual inspection of rises in the soil moisture spectrum does not lead to a clear decision (Figure 36). The natural cause of the present rise in soil moisture could be verified by examination of the additional variable precipitation.

Examples for flagging results of the positive jump detection method are shown in Figure 31 to Figure 34.

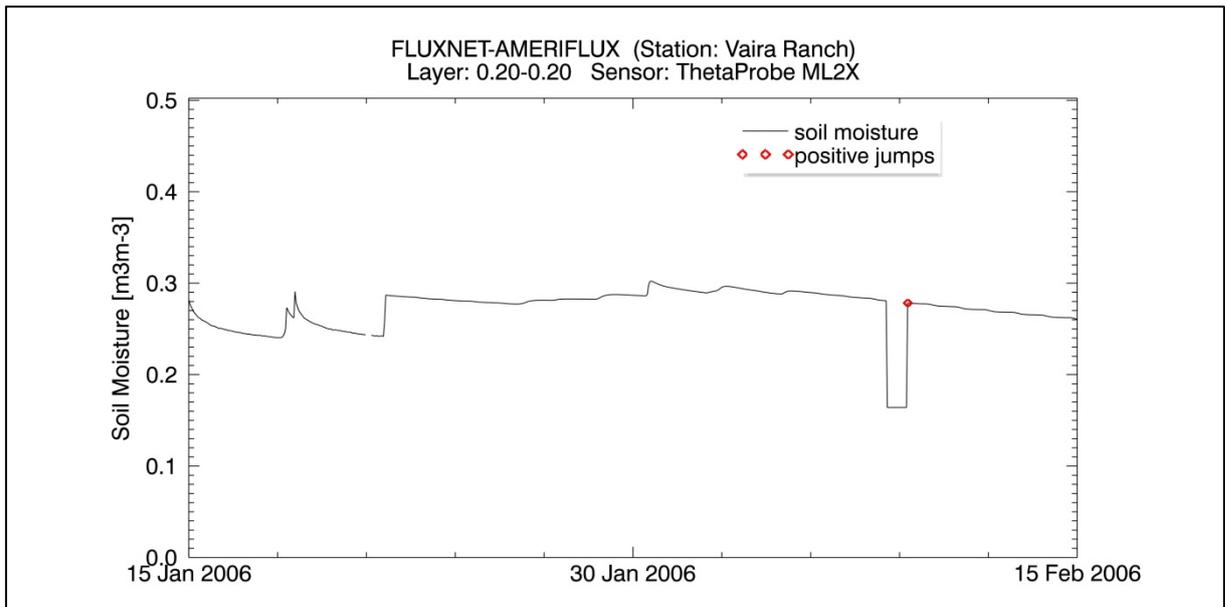


Figure 31: Example for flagging results of a positive jump (Station Vaira Ranch of network FLUXNET-AMERIFLUX, ThetaProbe ML2X, 0.20-0.20m depth)

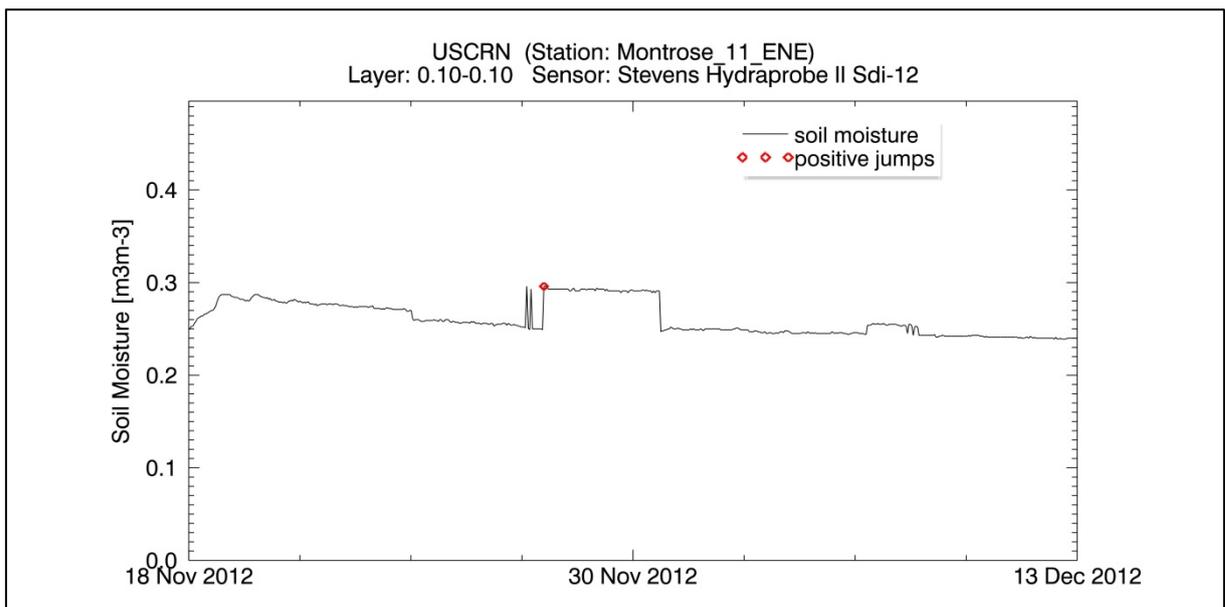


Figure 32: Example for flagging results of a positive jump (Station Montrose_11_ENE of network USCRN, Hydraprobe II Sdi-12, 0.10-0.10m depth)

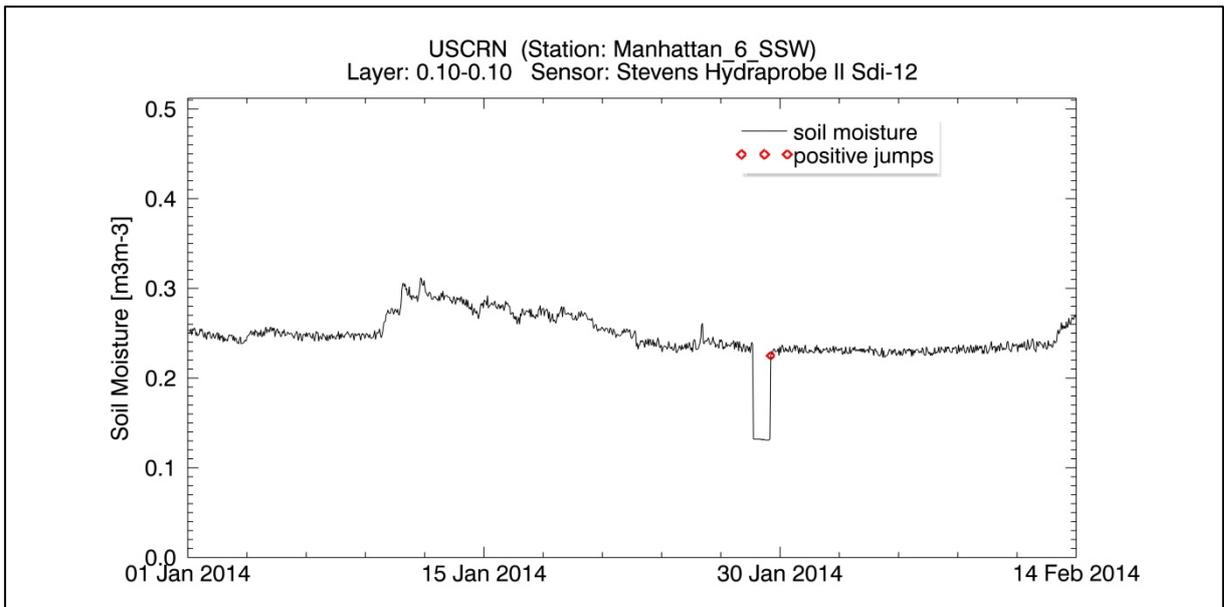


Figure 33: Example for flagging results of a positive jump (Station Manhattan_6_SSW of network USCRN, Stevens Hydraprobe II Sdi-12, 0.10-0.10m depth)

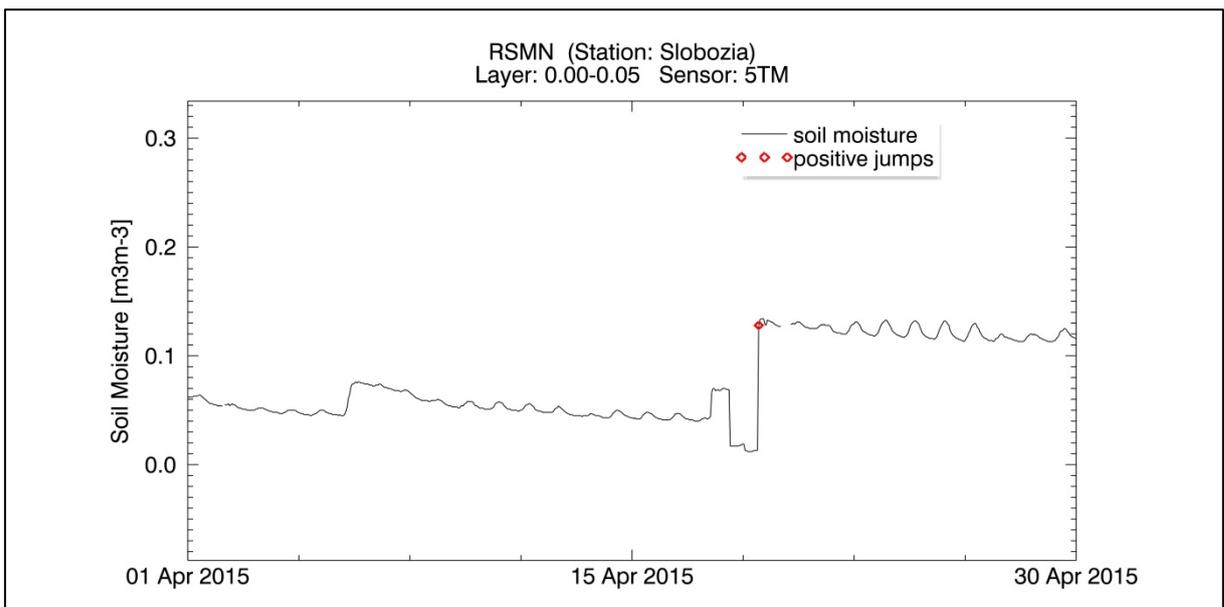


Figure 34: Example for flagging results of a positive jump (Station Slobozia of network RSMN, 5TM, 0.00-0.05m depth)

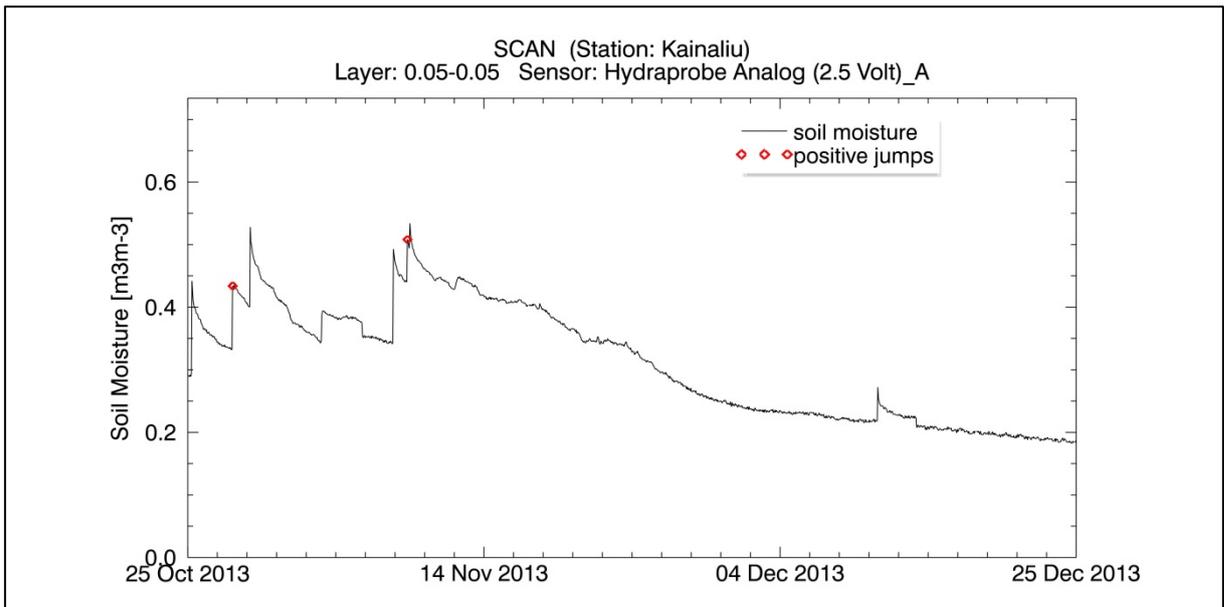


Figure 35: Example for erroneously flagged positive jumps (Station Kainaliu of network SCAN, Hydraprobe Analog (2.5 Volt), 0.05-0.05m depth)

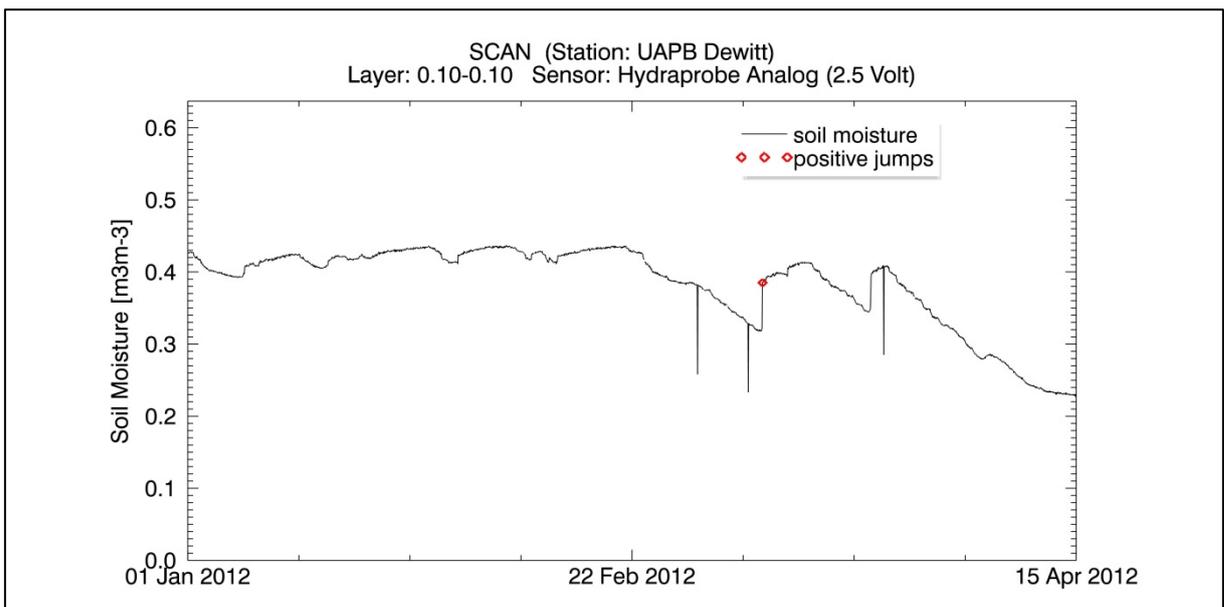


Figure 36: Example for an erroneously flagged positive jump (Station UAPB Dewitt of network SCAN, Hydraprobe Analog (2.5 Volt), 0.10-0.10m depth)

4.2.3 Negative Jumps

Almost 60% of negative jumps were detected correctly. Examples are shown in Figure 37 to Figure 40.

The reasons for failure of the flagging method for jumps are the same as for the spike detection. For noisy data or during the occurrence of consecutive erroneous events the detection method for negative jumps works insufficiently (Figure 41). An interesting result is shown in Figure 42. Two negative spikes have been detected as negative jumps.

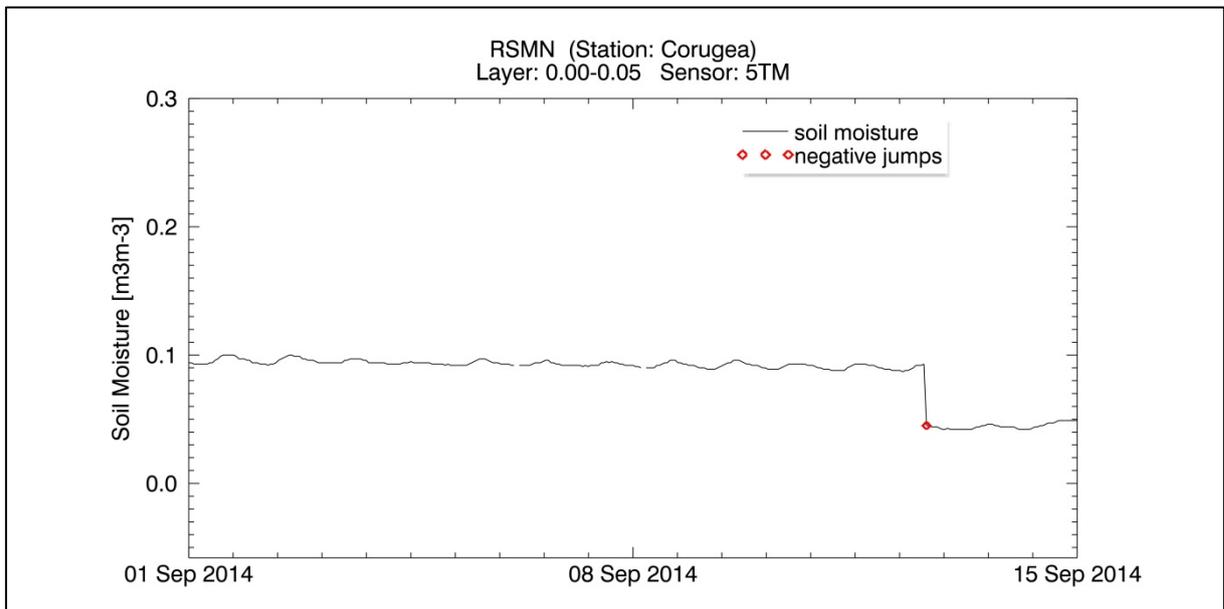


Figure 37: Example for flagging results of a negative jump (Station Corugea of network RSMN, 5TM, 0.00-0.05m depth)

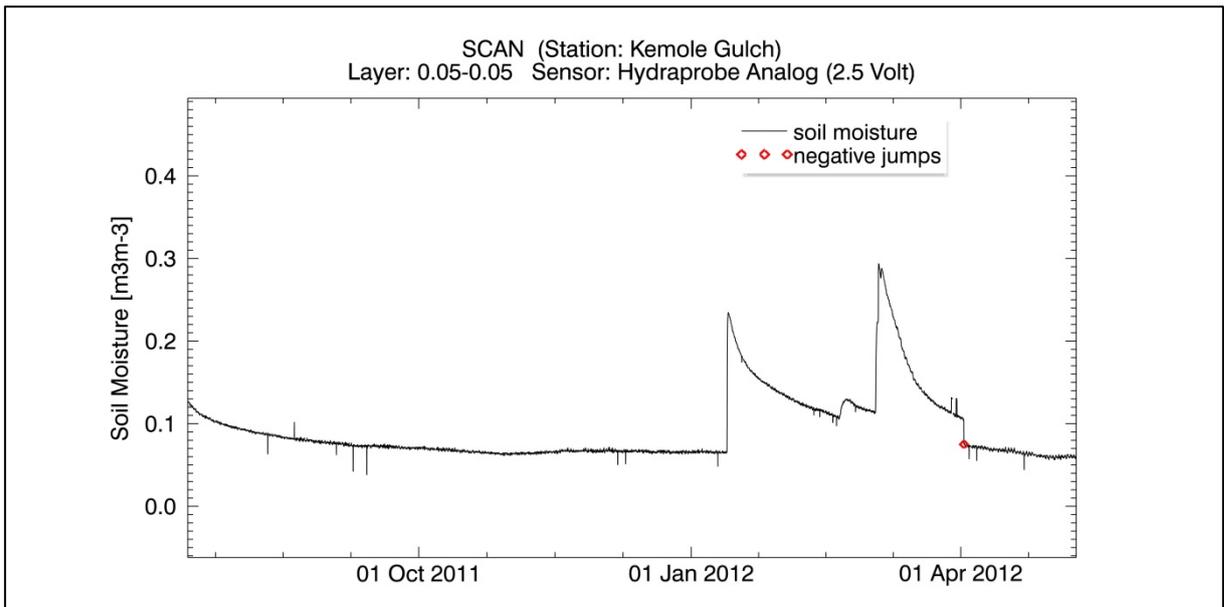


Figure 38: Example for flagging results of a negative jump (Station Kemole Gulch of network SCAN, Hydraprobe Analog (2.5 Volt), 0.05-0.05m depth)

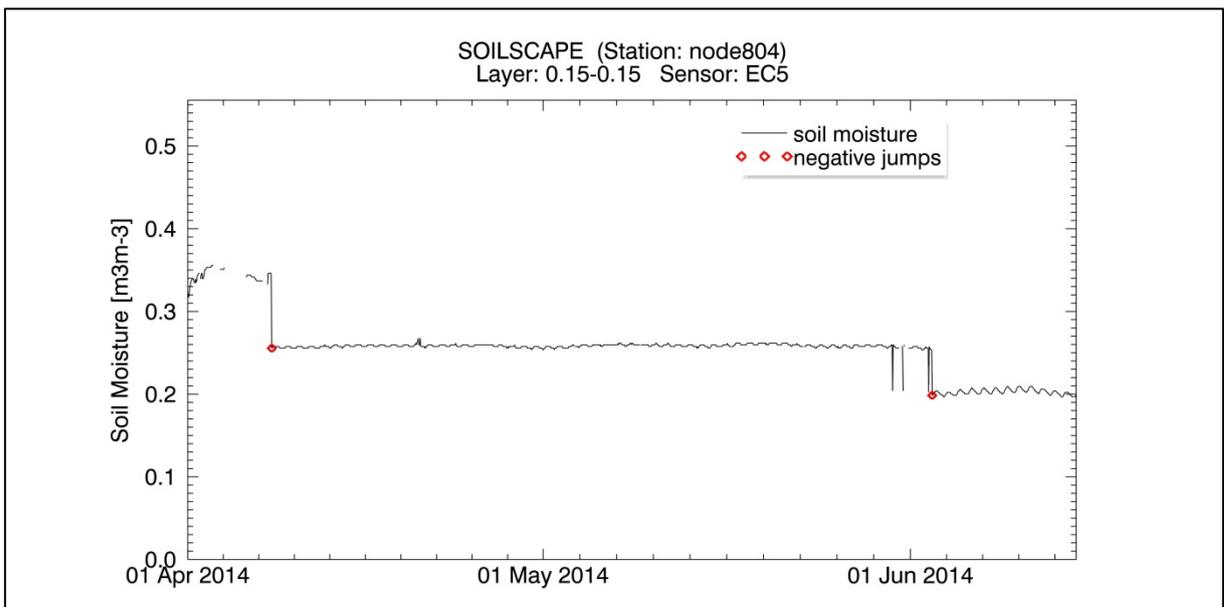


Figure 39: Example for flagging results of a negative jump (Station node804 of network SOILSCAPE, EC5, 0.15-0.15m depth)

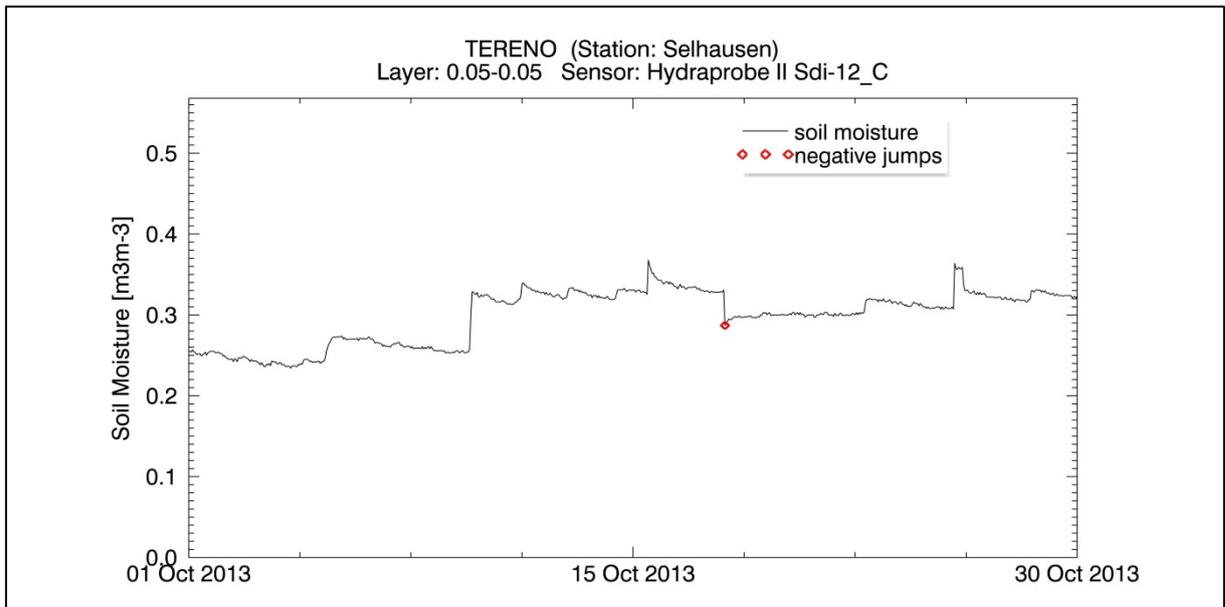


Figure 40: Example for flagging results of a negative jump (Station Slobozia of network RSMN, 5TM, 0.00-0.05m depth)

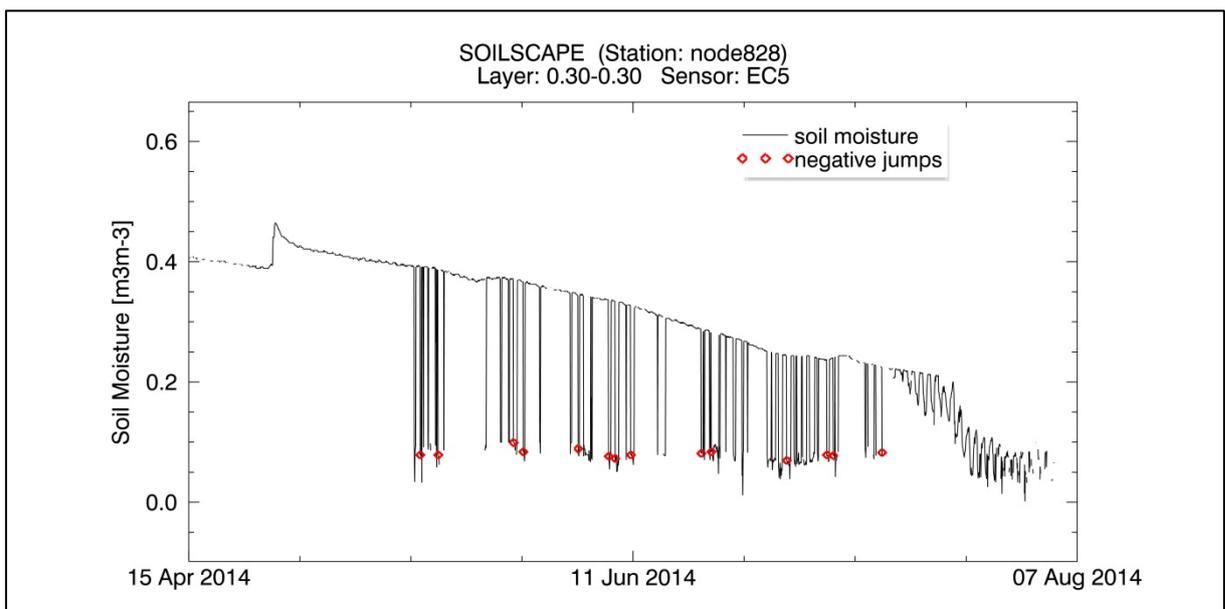


Figure 41: Example for a flagging result of negative jumps for a soil moisture time series with consecutive erroneous events (Station node828 of network SOILSCAPE, EC5, 0.30-0.30m depth)

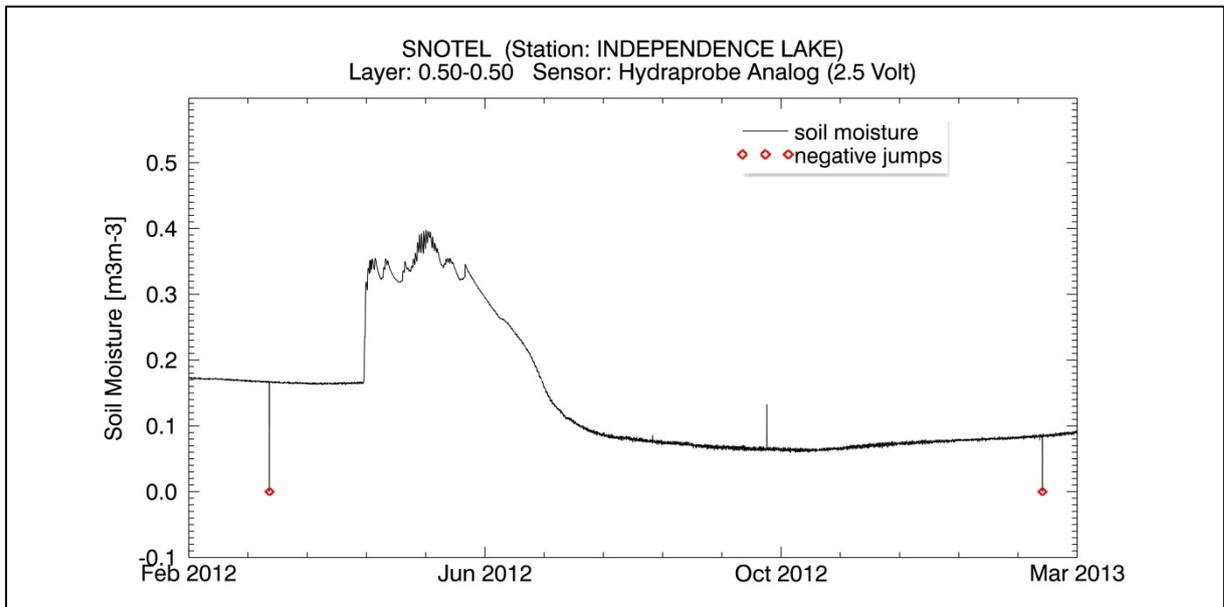


Figure 42: Example for erroneously flagged negative jumps, catching some negative spikes (Station INDEPENDENCE LAKE of network SNOTEL, Hydraprobe Analog (2.5 Volt), 0.50-0.50m depth)

4.2.4 Low Level Plateaus

The detection rate of low level plateaus is almost 60%, examples are shown in Figure 44 and Figure 45. Reasons for failure of this detection method are as follows:

- The relative standard deviation of the soil moisture level is higher than defined in equation [10] (Figure 46).
- Missing values occur within the low level plateau (Figure 46).
- Low level plateaus caused by frozen soil have been identified, which do not satisfy the criterion of the existence of a negative jump.

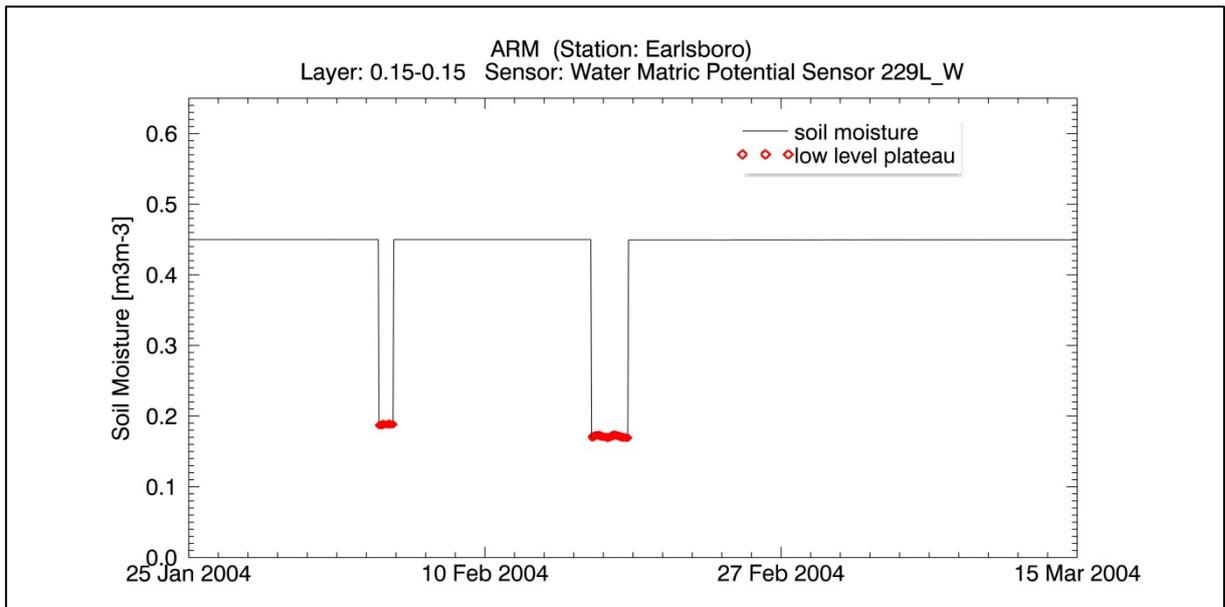


Figure 43: Example for flagging results of low level plateaus (Station Earlsboro of network ARM, Water Matric Potential Sensor 229L_W, 0.15-0.15m depth)

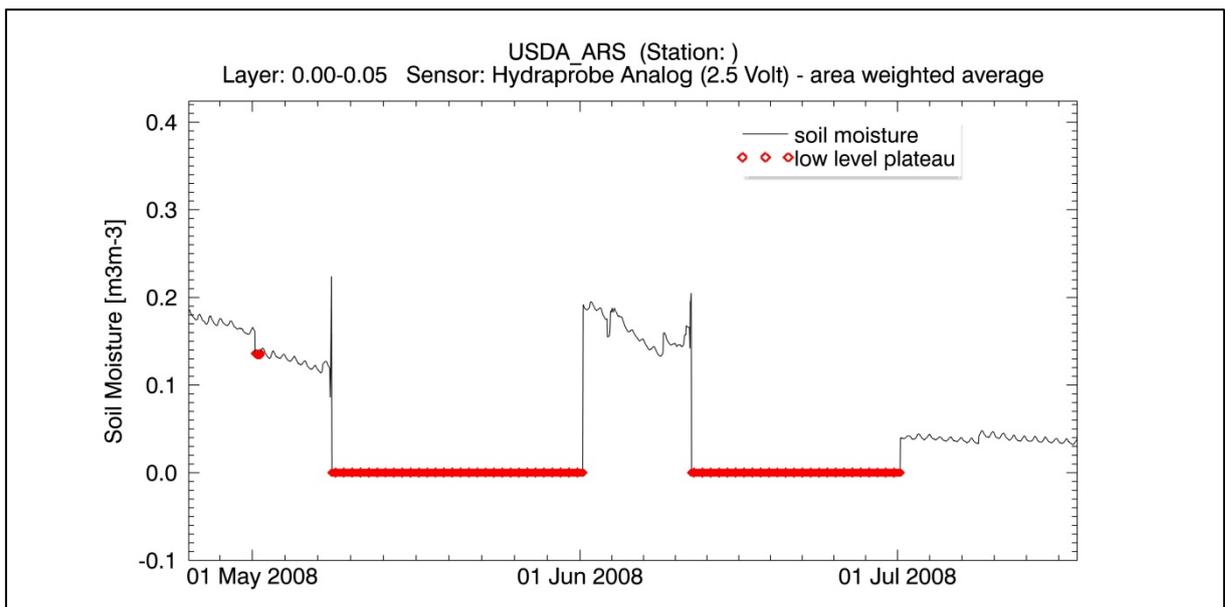


Figure 44: Example for flagging results of low level plateaus (Station RC of network USDA_ARS, Hydraprobe Analog (2.5 Volt), 0.00-0.05m depth)

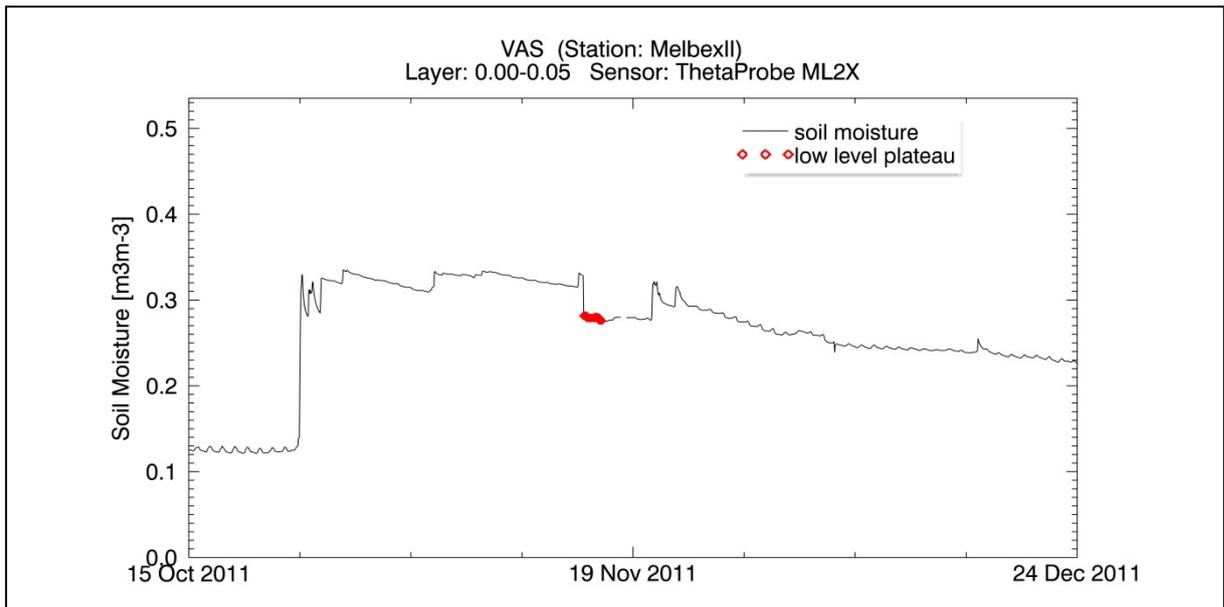


Figure 45: Example for flagging results of a low level plateau (Station MelbexII of network VAS, ThetaProbe ML2X, 0.00-0.05m depth)

4.2.5 Saturated Plateaus

More than 90% of the saturated plateaus in the validation datasets could be detected. Examples are shown in Figure 46 to Figure 49. Typical reasons for failure of this method are:

- The variation of soil moisture is higher than defined in equation [7].
- The average level of soil moisture is lower than defined in equation [9].
- The saturated plateau lasts less than the defined minimum interval of twelve hours.
- Missing values occur within the saturated plateau (Figure 50).
- The saturated plateau does not show a clear end.

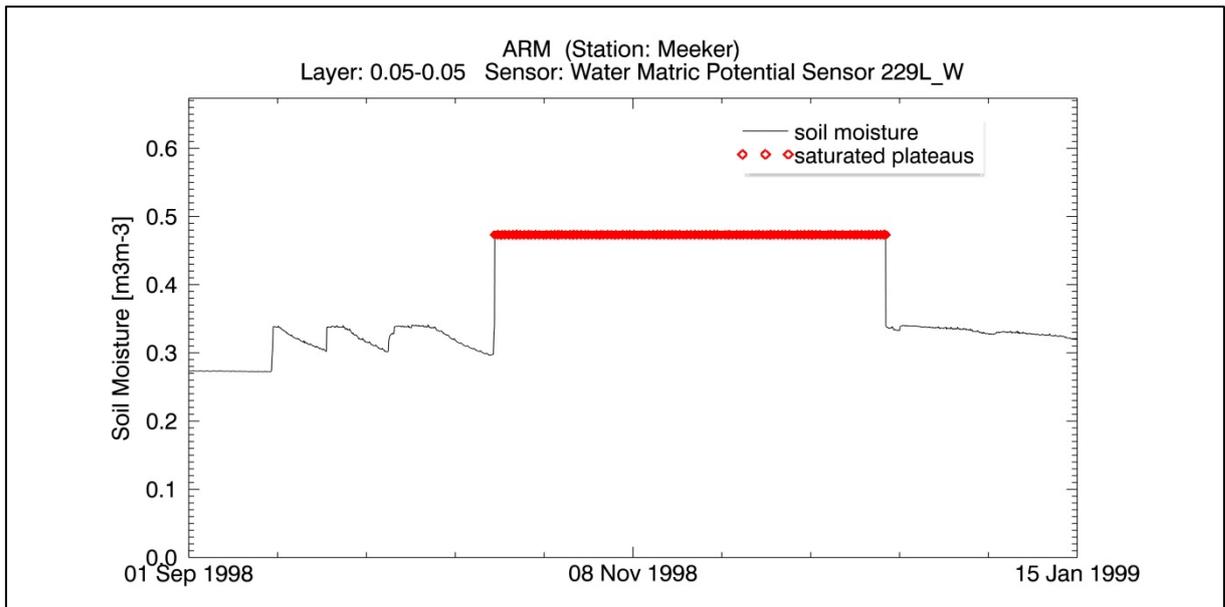


Figure 46: Example for flagging results of a saturated plateau (Station Meeker of network ARM, Water Matric Potential Sensor 229L_W, 0.05-0.05m depth)

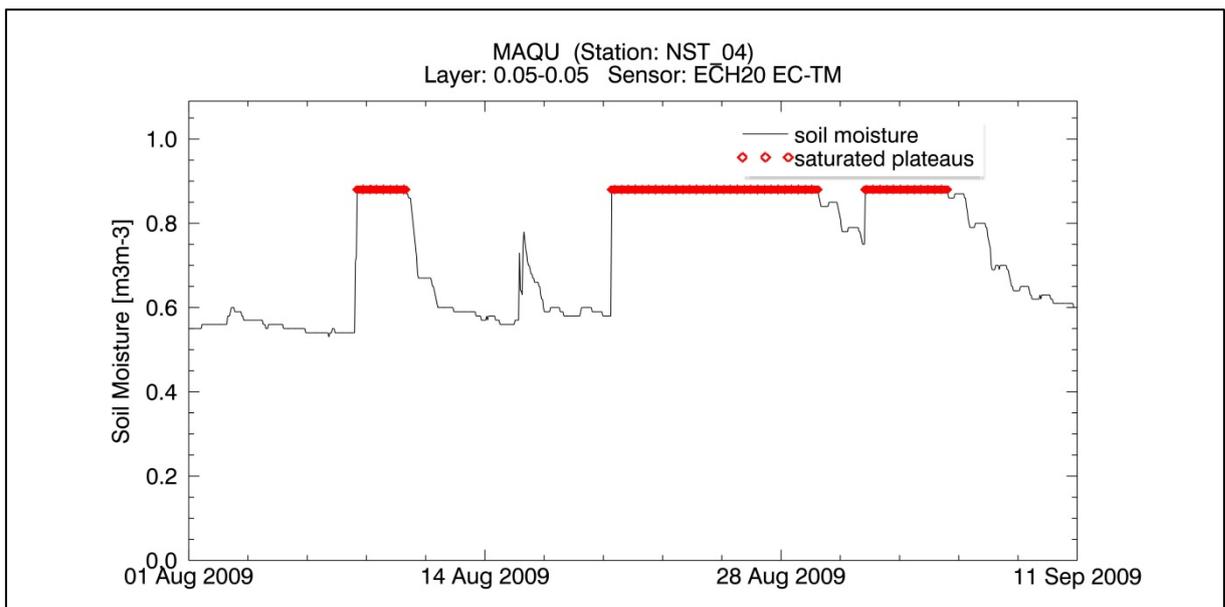


Figure 47: Example for flagging results of saturated plateaus (Station NST_04 of network MAQU, ECH20 EC-TM, 0.05-0.05m depth)

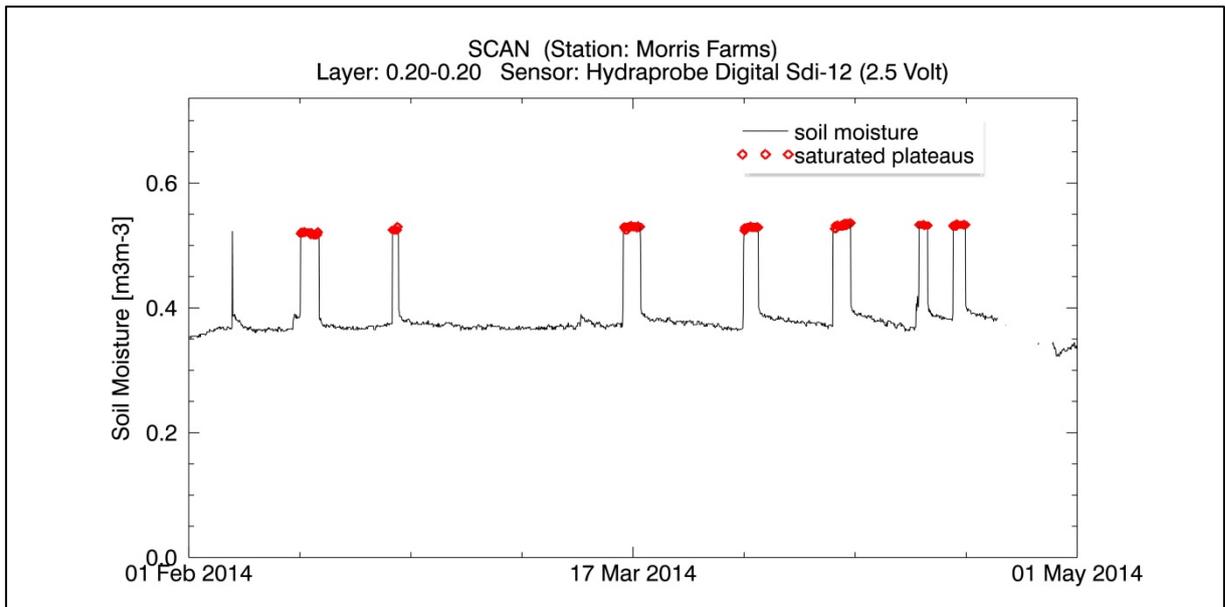


Figure 48: Example for flagging results of saturated plateaus (Station Morris Farms of network SCAN, Hydraprobe Digital Sdi-12 (2.5 Volt), 0.20-0.20m depth)

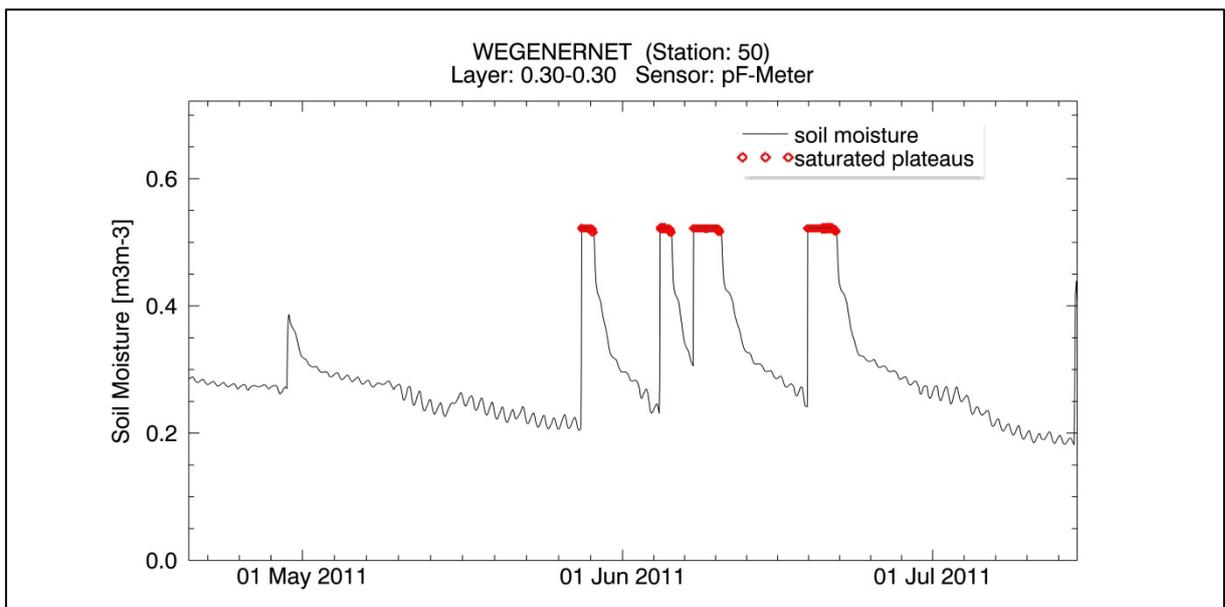


Figure 49: Example for flagging results of saturated plateaus (Station 50 of network WEGENERNET, pF-Meter, 0.30-0.30m depth)

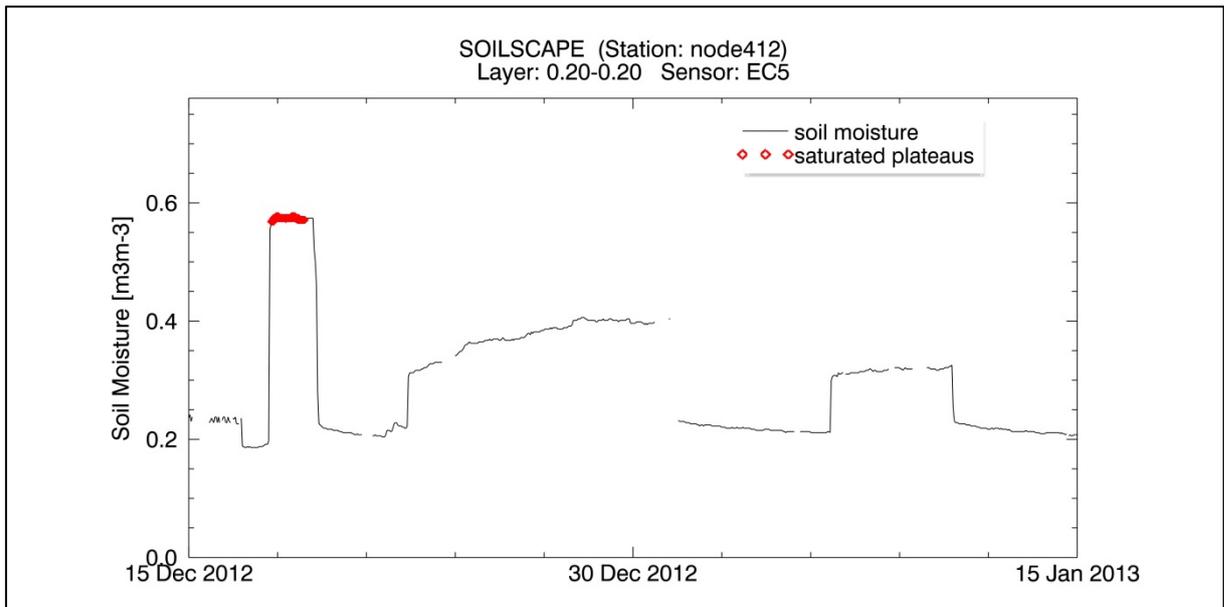


Figure 50: Example for flagging results of saturated plateaus (Station node412 of network SOILSCAPE, EC5, 0.20-0.20m depth)

4.3 Results for the entire ISMN

An overview of the spectrum-based quality flags detected in the datasets of the entire ISMN can be found in Figure 51. Presented are the percentages of flagged observations in respect to the available soil moisture observations for all measurement depths per network. For a better readability of the graphic, networks with an overall flagging percentage of 0.0% have been excluded.

The most frequently detected spectrum-based quality flags are the saturated plateaus, followed by the low level plateaus. Obviously, the percentages of flagged events vary between the networks, expressing their certain characteristics. Saturated Plateaus are most prominent for the networks CAMPANIA, SWEX_POLAND and WEGENERNET. The reasons are different. Soil moisture measurements within CAMPANIA are observed at one depth layer only, which is 30 centimetres. Naturally, saturated plateaus are more likely to occur in deeper layers than in surface soil layers. While several stations of the network SWEX_POLAND are situated in marsh land, the saturated plateaus within WEGENERNET result from the usage of pF-Meters, mainly in deeper soil layers. Low level plateaus were detected the most for ARM and USDA-ARS. It was already stated in Dorigo et al., (2013) that several occurrences of repeated sensor drop outs were observed for both networks.

Although the percentages of detected positive and negative jumps are much smaller than for plateaus, the detection of these events should lead to a careful investigation of the datasets

since they may indicate a permanent change of the level of soil moisture. This permanent offset can influence the validation of satellite-retrieved and model-based soil moisture datasets significantly.

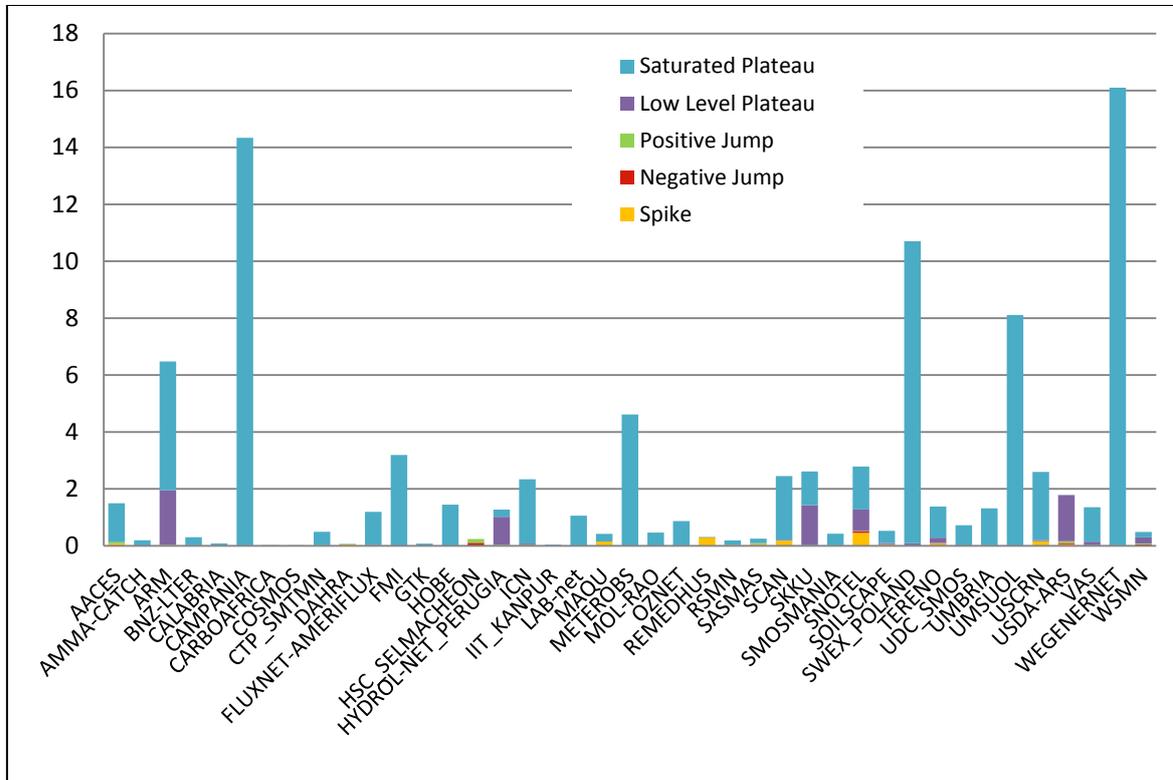


Figure 51: Percentages of spectrum-based quality flags per network.

5 Conclusions

The results presented in the previous chapter clearly show that the developed quality control procedures do not catch all erroneous events. Obviously, automated quality control procedures cannot be expected to work perfectly, but an overall good percentage of erroneous measurements could be identified by the introduced methods and the percentages of correct observations which were flagged erroneously are extremely low. Even if not all erroneous observations can be detected the spectrum-based quality control algorithms act as an indicator for quality. Especially in combination with the consistency tests of the geophysical range and the geophysical consistency checks described in (Dorigo, et al., 2013), the spectrum-based quality control procedures provide a relatively reliable statement about the quality of the examined dataset.

In general, the quality control algorithms are working very well for data which behaves very similar to the datasets used for calibration. Those datasets were chosen because of the clear appearance of erroneous measurements. Reasons for the failure of the flagging procedures can be summarized as follows:

- **Existence of missing values:** A general problem during the detection of suspicious events is the existence of missing values. Single missing values can be overcome by the introduced algorithms, but multiple missing values or whole periods of missing data lead to a disturbance of the Savitzky-Golay filter and the characteristic shapes in the first and second derivatives. This is the case for all described types of quality control mechanisms. It has to be noticed that this is also valid for measurements obtained at a lower temporal resolution than hourly data, e.g. for historical networks or network PBO_H2O with daily observations obtained from GPS measurements.
- **Imperfect definition:** Some thresholds had to be defined for the different quality control procedures to detect as much erroneous events as possible and avoid over-flagging of correct data. Of course, real data does not always follow exactly those definitions and therefore some erroneous measurements cannot be detected automatically. Such definitions which sometimes prevent the algorithms from detecting flags are:
 - Spikes were introduced as short-time events, lasting only for a single time step. In the datasets spikes sometimes last for two time steps. This could be caused by the temporal harmonization step within the ISMN processing chain. Only hourly measurements are implemented into the ISMN, even if observations at a higher temporal resolution exist.

- Saturated plateaus are only identified if the level of soil moisture is higher than 95% of the maximum soil moisture value of the entire data period.
- Plateaus in general are defined to be only detected if they last at least 12 hours.
- Low level plateaus are only detected if a preceding negative jump was identified.
- **Noisy data:** Noisy data causes also relatively noisy first and second derivatives, resulting from the properties of the Savitzky-Golay filter and the used parameters. Therefore, the characteristics of the two derivatives which are used to identify the different suspicious measurements are superimposed by the random noise.
- **Consecutive erroneous events:** Similarly to noisy data the characteristic shape of the derivatives is disturbed and the defined methods fail to detect erroneous measurements.

The here described quality control procedures are implemented within the ISMN processing chain since spring 2014. For each case of the described quality detection methods (i.e., flags, jumps, and plateaus) separate flags are defined, which are attached to the soil moisture measurements and provided to all registered users of the ISMN in addition to the actual measurements, which themselves remain unchanged.

The quality flags are not only of interest for data users, but also to the data providers, which was revealed by a survey conducted amongst all data providers of the ISMN in October 2014. Thus the quality control procedures will help to improve the reliability of validation studies based on in situ soil moisture observations on the one hand, and will help to identify problems (e.g. energy supply, sensor drop out) which may occur at the measurement sites on the other hand.

Bibliography

- Albergel, C., de Rosnay, P., Gruhier, C., Munoz-Sabater, J., Hasenauer, S., Isaksen, L., et al. (2012). Evaluation of remotely sensed and modelled soil moisture products using global ground-based in situ observations. *Remote Sensing of Environment*, *118*, 215-226.
- Congalton, R., & Green, a. K. (2009). Assessing the accuracy of remotely sensed data: Principles and practices. 2nd ed. *CRC Press/Taylor and Francis*.
- Dorigo, W., Gruber, A., De Jeu, R., Wagner, W., Stacke, T., Loew, A., et al. (2015). Evaluation of the ESA CCI soil moisture product using ground-based observations. *Remote Sensing of Environment*, *162*, 380-395.
- Dorigo, W., Oevelen, P. v., Wagner, W., Drusch, M., Mecklenburg, S., Robock, A., et al. (2011a). A new international network for in situ soil moisture data. *Eos Transactions AGU*, *92*(17), 141-142.
- Dorigo, W., Wagner, W., Hohensinn, R., Hahn, S., Paulik, C., Xaver, A., et al. (2011b). The International Soil Moisture Network: A data hosting facility for global in situ soil moisture measurements. *Hydrology and Earth System Sciences*, *15*(5), 1675-1698.
- Dorigo, W., Xaver, A., Vreugdenhil, M., Gruber, A., Hegyiová, A., Sanchis-Dufau, A., et al. (2013). Global Automated Quality Control of In situ Soil Moisture data from the International Soil Moisture Network. *Vadose Zone Journal*.
- Eilers, P. (2003). A Perfect Smoother. *Analytical Chemistry*(75), 3631-3636.
- (1992). Savitzky-Golay Smoothing Filters. In B. Flannery, S. Teukolsky, & W. Vetterling, *Numerical Recipes in C: The art of scientific computing*. Cambridge University Press.
- Hillel, D. (1998). Environmental soil physics. *Academic Press*.
- Hubbard, K., Goddard, S., Sorensen, W., Wells, N., & Osugi, a. T. (2005). Performance of quality assurance procedures for an applied climate information system. *J. Atmos. Ocean. Technol.*, *22*, 105-112.
- Legates, D. R., Mahmood, R., Levia, D., DeLiberty, T. L., Quiring, S., Houser, C., et al. (2011). Soil moisture: A central and unifying theme in physical geography. *Progress in Physical Geography*, *35*, 65-86.

- Mittelbach, H., Casini, F., Lehner, I., Teuling, A., & Seneviratne, S. (2011). Soil moisture monitoring for climate research: Evaluation of low-cost sensor in the framework of the Swiss soil moisture experiment (SwissSMEX) campaign. *J. Geophys. Res. D Atmos.*, 116(D05111).
- Persson, P.-O., & Strang, G. (2003). Smoothing by Savitzky-Golay and Legendre Filters. In D. S. Gilliam, *Mathematical systems theory in biology, communications, computation, and finance* (S. 301). Springer.
- Robinson, D., Campbell, C., Hopmans, J., Hornbuckle, B., Jones, S., Knight, R., et al. (2008). Soil moisture measurement for ecological and hydrological watershed-scale observatories: A review. *Vadose Zone Journal*, 7, 358-389.
- Savitzky, A. and Golay, M.J.E. (1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry*, 36, 1627-1639.