AUTOMATED QUALITY CONTROL OF IN SITU SOIL MOISTURE AND SOIL TEMPERATURE DATA FROM THE TIBET-OBS NETWORKS

JIAQI YANG
February, 2017

SUPERVISORS:
Dr.Y.Zeng
Phd: S.Lv

AUTOMATED QUALITY CONTROL OF IN SITU SOIL MOISTURE AND SOIL TEMPERATURE DATA FROM THE TIBET-OBS NETWORKS

JIAQI YANG
Enschede, The Netherlands, February 2017

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Water Resources and Environmental Management

SUPERVISORS:
Dr.Y. Zeng
Phd: S.Lv
Dr. ir. C. van der Tol

THESIS ASSESSMENT BOARD:
Prof. Dr. Z. Su (Chair)
Dr.Y.Zeng
ir. A.M. van Lieshout
Dr.V Bense

ABSTRACT

Soil moisture as a crucial property highly related to land-atmosphere systems. Soil moisture enormously impacts the climatic system, including meteorological, hydrological and geographical systems, especially for high plateau like Tibet. Since 1990s, various campaigns and in situ networks were established to monitor soil moisture and soil temperature over Tibetan Plateau. However, due to multiple sensor settings, different installation depths and distribution locations, ground measured soil moisture obtained over large-scale area are difficult to ensure the data quality. In this case, quality management of in situ soil moisture data would provide comprehensively enhancement for appropriately validating and calibrating remote sensors gathered or models simulated soil moisture. The main objective is to check if the current existing QC methods is applicable to detect QC issues over TP or not. Quality control methods proposed by Dorigo et al., 2013 revealed comprehensive capability in solving in situ soil moisture over large spatial scales. Accurate and effectively quality controlled data for the Tibet Plateau as the highest and largest plateau is if great relevance. Four major variables and inputs are tested, including threshold, temperature limit, precipitation detection and spurious constant value identification. The threshold is effective but could not be applicable to all regions over TP. Temperature limit at -5 degree showed relatively good result over humid region, but for more arid regions, the temperature limit need to be set lower. The result of using different sets of precipitation products is not promising, due to the fact that the estimate of precipitation over TP is highly challenging. Constant value detection with the interval of 24 hours could be applied for entire Tibet-Obs soil moisture network. Overall, this method could be used for Tibet-Obs network but requires modifications, as well as to consider the difference in the multiple climatic zones over TP, in order to do the QC for the soil moisture observed at various depths across the TP.

Keywords: Quality control, soil moisture, Tibet plateau.

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my deep sense of gratitude to my ITC supervisor Dr.Yijian.Zeng and Phd Lv.Shaoning for their constant support and invaluable guidance. Many thanks to their timely help and great patience no matter how busy they are.

My foremost thanks the Faculty of Geo Information Science and Earth Observation (ITC), University of Twente, the Netherlands for giving me a chance to chase this course. I would like to thank ir. A.M. van Lieshout, Course Director of Water Resource and Management for his pushing and encouragement.

My sincere thanks to Prof. Dr. Z. Su for his preciousness advice and selfless help.

Also I am grateful my friends for their constant support and academic suggest without reservation.

Last but not least, I would like to thank my parents for their continuous support, absolutely understanding and long-term encouragement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1. Background

Soil moisture as a crucial property is highly related to land-atmosphere systems, and exerts strong influence on the rate of soil respiration, carbon transportation, microbiological activities as well as the precipitation occurrence. Tibetan plateau, a plateau of elevation of 4000m above sea level and size of 3 million square kilometres, has a tremendous impact over the entire Asia on climate conditions. It is indicated that the Tibetan plateau has a wetter climates to the south and east of the plateau and drier climates to the north and west of the plateau (Kutzbach et al.,1993), whereas soil moisture closely correspond with climate change by soil moisture- soil temperature and soil moisture-precipitation processes(May et al., 2015).

Ground measured soil moisture are invaluable often collected from long-term measurement with tremendous spatial coverage. Compared with other surface measurements, gathering soil moisture data is more sophisticated due to several soil depths demanded. The utilization of in situ soil moisture are validation and calibration for the land surface model simulation or remote sensor collected observation (i.e., Sun et al., 2016; Rajib et al., 2016). The other usefulness of time series of in situ soil moisture is capable of demonstrating climate change.

However, the quality of in situ soil moisture is highly variable due to the deployment of sensors, sensor depth, sensor types and spatial setup. As the highest plateau worldwide, data collection is more complicated than in other areas. The first influenced situation is the spatial setup because of the multiple climatological, geographical, hydrological conditions, i.e. three climates which are cold semiarid, cold humid and cold arid exist over the entire Tibetan plateau. Furthermore, the freeze-thaw process would result in an overestimation or an underestimation of soil moisture collected during summer and winter, respectively (Su et al., 2011; Su et al., 2013). Even though remote sensed soil moisture data has the advantage of monitoring soil moisture content at regional scale, the data just could gathered from shallow depth, i.e., about 0 to 5 cm(Schmugge, 1983). Guaranteeing the quality of in situ soil moisture is a decisive point for other researches. In order to ensure the quality of the ground measured data, the best way is checking manually by visually inspecting every record. However, due to quantity measurement with long-term and large-spatial coverage, manually checking in situ data from networks are almost impossible. Therefore, the automated qualify process provides the possibility of dealing with massive data.

1.2. Literature review

Quality control checked the quality of all variables in the production which are widely used in engineering and science. Quality control have two levels including control of raw data and control of processed data. Quality control of raw data will remove errors of technical device, while for processed data is aim to check consistency, bias and noisy (Geiger et al., 2002; MEEK et al., 1994; WMO, 2004). For example, for long term soil moisture, there will be missing value when sensor is drop out at any time, and a consistency phenomenon may present incorrect correlation between nearby stations. Thus, the task of quality control of observational data is to find out missing data, erroneous data and to correct the potential data, and to ensure a high level of accuracy for all users.

To achieve the goal of quality control, numerous methods are discovered to address the incorrect values. MEEK , 1994 developed three validation approaches for detecting dynamic high and low boundary, checking the rate of change and witnessing a continual no observed change in defined period for each variables. Detecting dynamic high and low boundary refer to as threshold test is a common method employed by many quality control researches (Geiger et al., 2002; Hubbard et al., 2005; Shafer et al., 2000; Jinshing You et al., 2010). Jinshing You et al., 2010 used a physical relationship to define the local boundary of soil water content which are not appropriate for global scale or ever for continental scale. Hubbard et al., 2005 developed a limit based high and low boundary calculated from historical data. This threshold method is restricted with data from long period to ensure its reliability. Besides, step change from Hubbard et al., 2005 assessed the different between two continuous steps to find the rate of change which are especially useful to identify datalogger problems. Step change can address potential incorrect value but cannot directly recognize errors.

The other methods to check geophysical consistency are extreme event detection and temporal test (Eischeid et al., 1995; Jinsheng You et al., 2006a). Extreme event test was proposed by Jinsheng You et al., 2006a by focusing on association between climatological data and extreme events to find out its weather patterns. The similiarity of climatological monthly data from multiple years should remain high level to ensure the consistency for the long-term (Eischeid et al., 1995). The weakness of extreme event test is that the data of extreme event required are very detailed. For temporal test, only stations existed over long period could meet the coniditon. Moreover, spatial linear regression approach supports to check whether the value fall within plausible range correlated with surrounding station (Eischeid et al., 1995; Geiger et al., 2002; Hubbard et al., 2005; Shafer et al., 2000; Jinsheng You et al., 2006b; Jinshing You et al., 2010).

Dorigo et al. (2013), employed comprehensive quality procedures to investigate soil moisture from International Soil Moisture Network (ISMN). The soil moisture data are obtained through global networks, mainly distributed in North American and Eurasia. The methods used in Dorigo's study are categorized into two sections, dynamic range check and consistency check. The dynamic range check used the threshold method while consistency check used spectrum-based methods, including spike detection, breaks detection and constant value detection (Inan et al., 2007; Journée & Bertrand, 2011; Malyshev & Sudakova, 1994). Even though the result was acceptable, they could not satisfactorily flag the values where the soil temperature close to zero. For those soil moisture stations over Tibetan Plateau, the QC method developed by Dorigo et al were used only for summer seasons, as during winter most of stations were experiencing frozen grounds. In this thesis, Tibetan Plateau, lying in the southwest of China and presenting few precipitation due to a cold and dry weather, will be the study area for this research.

1.3. Reserch Problem

1.3.1. General problem

Automated quality control process provides a relatively accessible tool to monitor the in-situ soil moisture data quality. Even though in the article of Dorigo et al., (2013), the researchers developed a considerably reliable tool for detecting errors of soil moisture over global regions, there still remain more space for improvement, particularly for its application over the Tibetan Plateau

In the previous article, they had just checked the Maqu network from the Tibet-Obs network when evaluating the performance of the global soil moisture data. They did not include data from Naqu an

Ngari over the Tibet due to the irregular time interval. However, in this study we will use the soil moisture time series over the Tibet-Obs network.

In order to check the consistency of soil moisture, the relationship between precipitation and soil moisture data would be employed to ensure the quality of soil moisture. However, noisy precipitation data would negatively influence on the consistency checking. Dorigo et al., 2013, also mentioned that the capability of automated quality control tool could be improved by using a finer precipitation product.

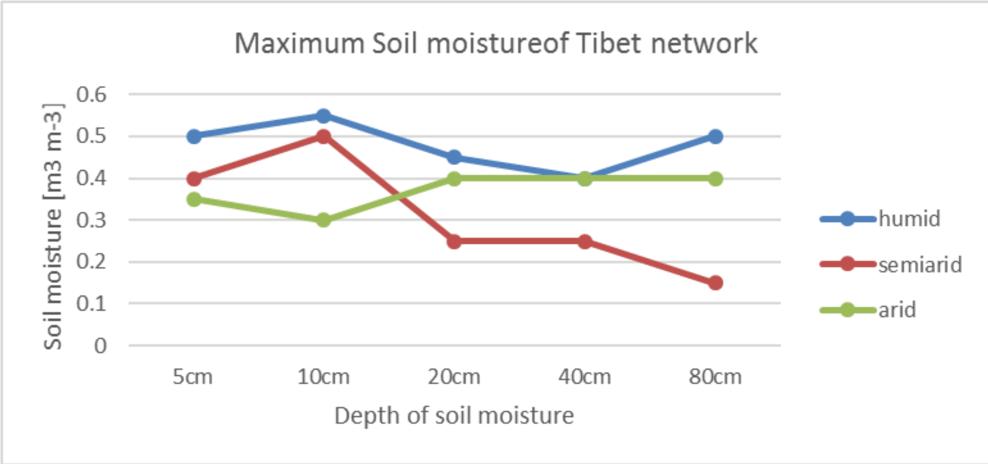
Regional climatic differences would also have consequences. Soil temperature close to zero would lead to a negative QA flag, as well as the removal of corresponding soil moisture measurements. This situation occurs abundantly in cold climatic zones. For different climatic regions, the QC procedure should be carefully evaluated.

Land covers, soil organic matters and soil types are highly related to soil porosity which is an important parameter in quality control tool. Soil porosity is a crucial factor by restricting the upper bound and lower bound of soil water storage. Inappropriate soil porosity could impact enormously on checking the dynamic range of whole time series of soil moisture.

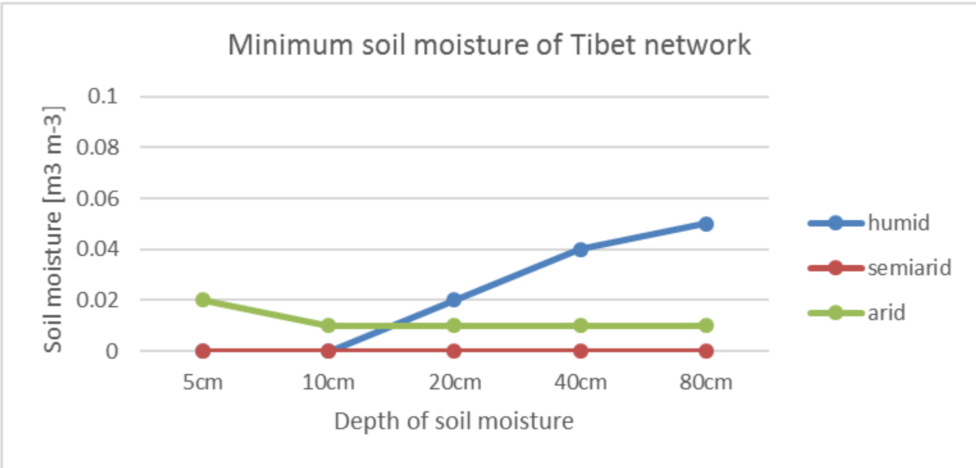
1.3.2. In situ soil moisture data quality over Tibetan Plateau

As the highest plateau, worldwide, in situ soil moisture data quality encounters with severe problems. The climatic condition of Tibet plateau is divided into three type, humid, semiarid and arid. These distinctive regions significantly affect the heat transfer system leading that soil moisture over these areas are different as well. In order to inspect the specific errors occurred over Tibetan plateau, classifications applied for three climatic zones and more detailed for soil types, elevation, soil organic matters should be implemented to figure out whether the existing soil moisture data quality issues were correlated with different conditions.

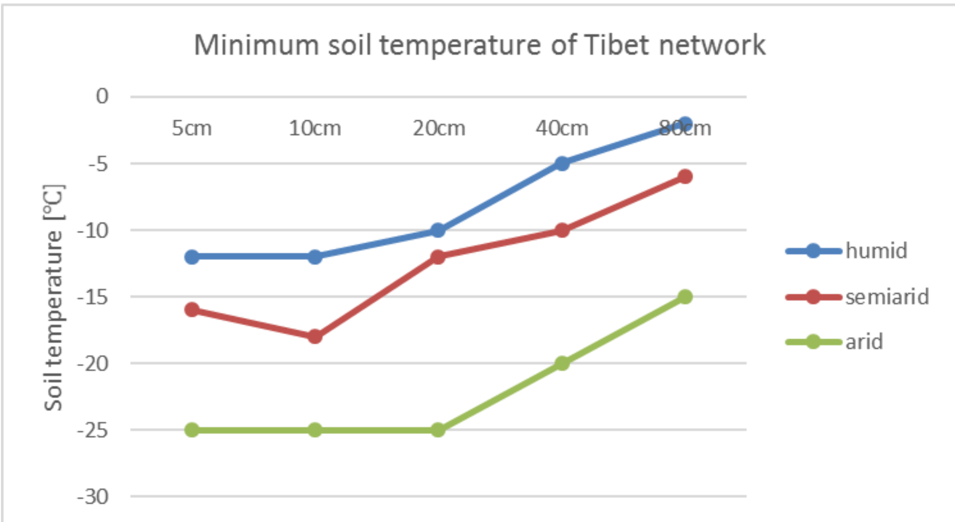
According to Zeng et al., (2016), the subnetworks belong to Tibet-Obs network are located in three diverse climatic zones which are humid, semiarid and arid area. Maqu is in humid climate while the climate of Naqu is semiarid and the arid area is Ngari network. As shown in Figure 1, the highest soil moisture and the lowest soil moisture of various climatic zones are distinctive. A summary in Table 1 shows that the highest soil moisture of humid arid is higher than that of semiarid and arid. The trend of soil moisture through the vertical profile of the soil over various climate is also different, and it indicates that the highest soil moisture of humid at deep layer is lower than that at low layers, conversely, maximum soil moisture of arid at deep layer is higher than that at low layer. This diverse phenomenon for soil moisture is a change to apply unified standard for multiple climatic zones. The soil temperature is also different, resulting in spurious flags for quality control based on a same standard.



(a)



(b)



(c)

Figure 1, (a) The maximum of soil moisture over multiple climatic zones, (b) The minimum soil moisture over multiple climatic zones, (c) The minimum soil temperature when the soil moisture still beyond 0 m^3 m^{-3}.

Table 1, A summary of minimum soil moisture value and maximum soil moisture and the lowest soil temperature value in humid, semiarid and arid area.

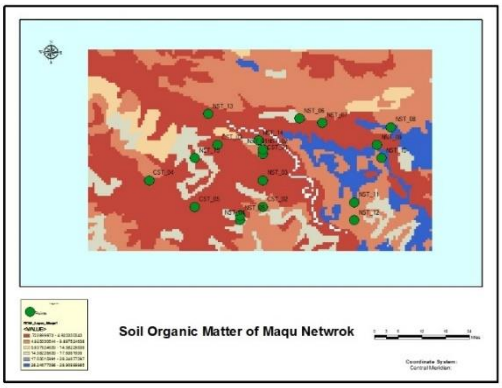
CLIMATE	HUMID			SEMIARID			ARID		
	minSM	maxSM	minT	minSM	maxSM	minT	minSM	maxSM	minT
5CM	0	0.5	-12	0	0.4	-16	0.02	0.35	-25
10CM	0	0.55	-12	0	0.5	-18	0.01	0.3	-25
20CM	0.02	0.45	-10	0	0.25	-12	0.01	0.4	-25
40CM	0.04	0.4	-5	0	0.25	-10	0.01	0.4	-20
80CM	0.05	0.5	-2	0	0.15	-6	0.01	0.4	-15
SUMMARY	0	0.55	-12	0	0.5	-18	0.01	0.4	-25

Soil types, soil organic matters and elevation are closely bounded with climatological conditions of the area. The soil organic matters are highly related to soil water content of shallow depth. Thus, diversely soil organic matters would affect surface soil moisture in multiple ways. Thus, to ensure the comprehensiveness and accuracy of the research, a classification map of soil organic matter was established in Figure 2a as well as in Table 2.

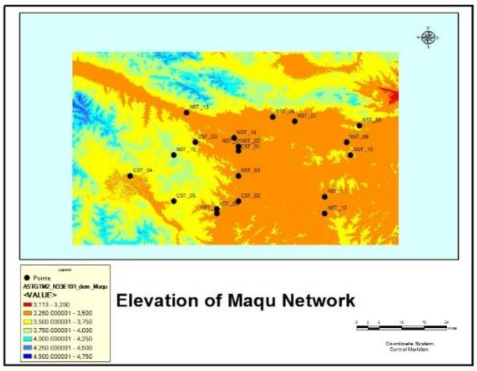
Kurylyk et al., (2013) concluded that liquid soil moisture in a frozen soil is significantly correlated with soil types. As the highest plateau, Tibet-Obs has a comparably low temperature result in frozen soil frequently discovered in winter. To avoid unpredictable fault flagging, a classified flagging standard for different soil types is inevitable. Figure 2c, 2d,2e give the three classification maps of Maqu network for various soil types. While, table 3 is a summary based on the classification maps over the TP.

Elevation is another factor introduced to research as its close association with temperature. The elevation of three subnetworks is distinctive, showing that the average elevation of Naqu is the highest and Maqu has the lowest average elevation. From table 4, it clearly shows that soil temperature of these subnetworks is totally divergent. In Figure 2b, A classification map of elevation for Maqu and its summary of classified types were created in table.

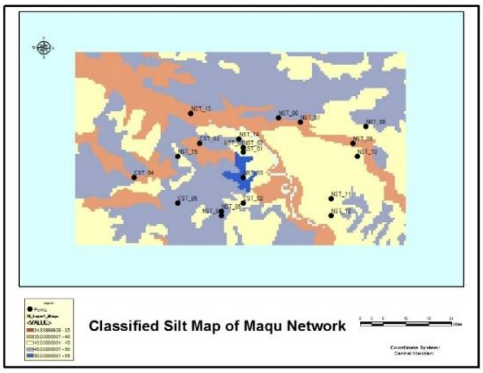
Therefore, in order to appropriately apply the quality control procedure for in situ soil moisture over Tibetan Platea, these potential factors will be evaluated.



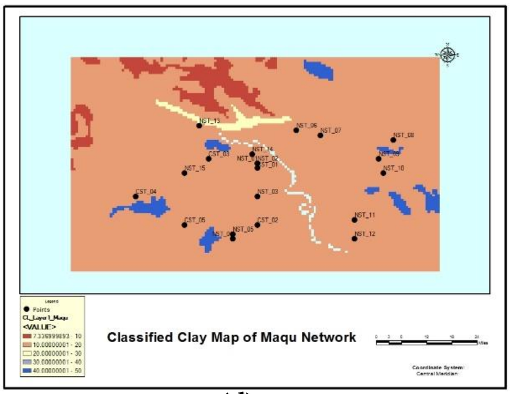
(a)



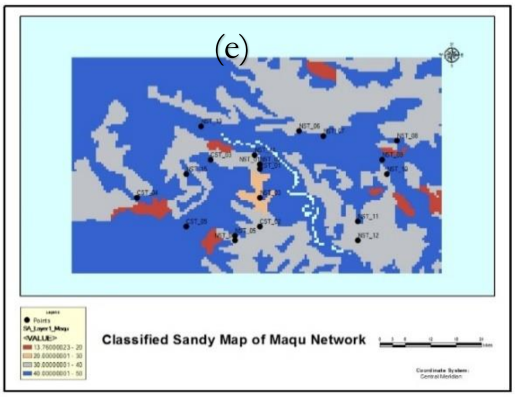
(b)



(c)



(d)



(e)

Figure 2, (a) A classification map for soil organic matters in Maqu network, (b) classification map for elevation in Maqu network, (c), (d) and (e) show three classification map for soil types in Maqu networks.

Table 2, A classified types of different soil organic matters for Tibet network

SOM NETWORK	NAQU(20 STATIONS)	MAQU(5 STATION)	SHIQUANHE(16 STAYIONS)	ALI(4 STATIONS)
TYPE 1 (0-5 G/100G)	100%	55%	100%	100%
TYPE 2 (5-10 G/100G)	0%	25%	0%	0
TYPE 3 (10-15 G/100G)	0%	10%	0%	0%
TYPE 4 (15-20 G/100G)	0%	0%	0%	0%
TYPE 5 (20- G/100G)	0%	10%	0%	0%

Table 3, Percentages of sand, silt and clay at the measurement sites in the Tibet network

SOIL TYPE	NAQU(5 STATIONS)	MAQU(20 STATION)	SHIQUANHE(16 STAYIONS)	ALI(4 STATIONS)
SAND	75%	25%	75%	100%
SILT	25%	75%	18.75%	0%
CLAY	0%	0%	6.25%	0%

Table 4, A classified table for different elevations in Tibet networks over various subnetworks

ELEVATION NETWORK (M AMSL)	NAQU(5 STATIONS)	MAQU(20 STATION)	SHIQUANHE(16 STAYIONS)	ALI(4 STATION
3000-3500 TYPE1	0	80%	0	0
3500-4000 TYPE2	0	20%	0	0
4000-4500 TYPE3	20%	0	100%	100%
4500-5000 TYPE4	80%	0	0	0

2. OBJECTIVES

2.1. General Objective

To evaluate the performance of using an automated quality control process for in situ soil moisture data over the Tibetan Plateau.

2.2. Subobjectives

1. To investigate sources and appearance of errors of the in-situ soil moisture data over Tibet.
2. To investigate the usefulness of an automated quality control from Dorigo et al., 2013 when applying for soil moisture data of Tibet-OBS network.
3. To study the influence of using different precipitation datasets;.
4. To evaluate the performance of the existing automated quality control process when meeting different climatic conditions.;
5. To identify how much the influence will be when applied with multiple land cover, different soil types.

2.3. Research questions

- Does the existing automated quality control program satisfy the demand of detecting soil moisture quality issue over Tibetan Plateau?
- Which are the source and appearance of errors for the in-situ soil moisture data over Tibet?
- Can the precipitation obtained from China Meteorological Forcing Dataset could improve the performance of quality control process?
- What could be modified to improve the performance of the best QC methods for high elevation plateau?

3. DATA

3.1. Study Area

The Tibetan plateau is located in the Southwest of China with the highest evaluation worldwide, which average evaluation range from 4000 to 5000 meters. The size of the region approximately 3 million square kilometres, the East-West length is about 2900 kilometres (73°18'52"E-104°46'59"E), the North-South length is 1500 kilometres (26°00'12"-39°46'50").

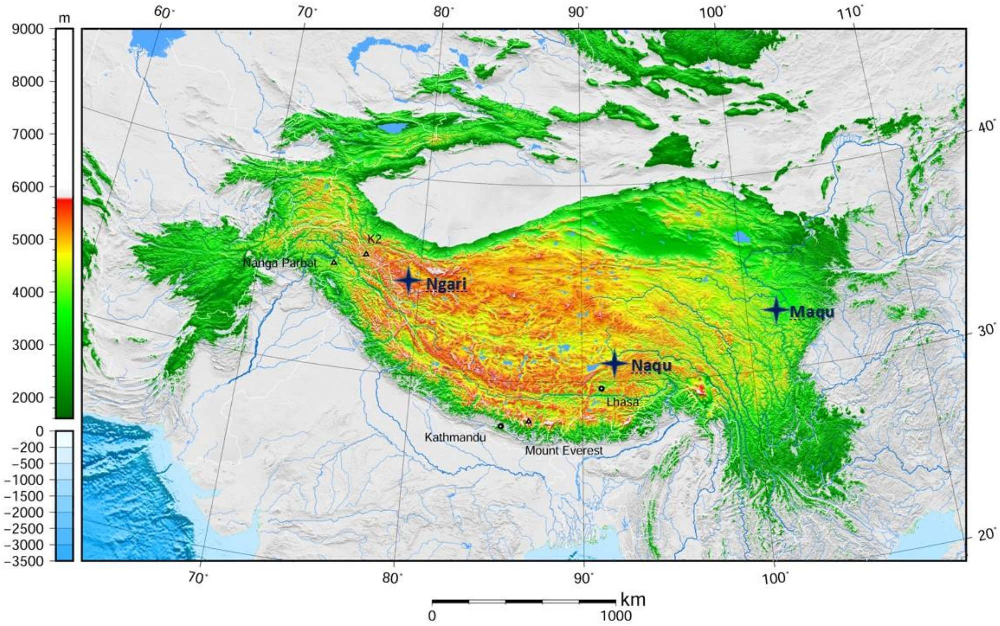


Figure 3, *A geographical illustration about Tibetan plateau. 3 in situ network in plateau are showed, Ngari, Maqu and Naqu. Reference:* Tibet and surrounding areas above 1600 m. topographic map *https://en.wikipedia.org/wiki/Tibetan_Plateau#/media/File:Tibet_and_surrounding_areas_topographic_map.png*

The Tibetan plateau is located in the Southwest of China with the highest evaluation worldwide, which average evaluation range from 4000 to 5000 meters. The size of the region approximately 3 million square kilometres, the East-West length is about 2900 kilometres (73°18'52"E-104°46'59"E), the North-South length is 1500 kilometres (26°00'12"-39°46'50").

3.2. Data

3.2.1. General

For the purpose of investigating the performance of automated quality control process when using different conditions, soil moisture data, soil temperature data, precipitation dataset from multiple sources and porosity value will be introduced. Soil moisture data and soil temperature data were obtained from the Tibet-OBS network. GLDAS precipitation data and data of China Meteorological Forcing Dataset will be used for comparison of impacts under multiple precipitation datasets. As for studying the influence of distinctive underlying surfaces, a soil particle-size distribution dataset would be employed.

3.2.2. Soil moisture and soil temperature data

Soil moisture data and soil temperature data were obtained from the Tibet-OBS network. Tibet-Obs network is the Tibetan Plateau observatory of plateau scale soil moisture and soil temperature (Tibet-Obs), which consists of three regional scale in-situ reference networks, including the Naqu network in a cold semiarid climate, the Maqu network in a cold humid climate and the Ngari network in a cold arid climate. The periods of each stations are showed in Figure 4, the red bar shows the original observation, while the purple bars are records after restarting observation.

3.2.2.1. Naqu network

Naqu network is located in the Naqu basin with an elevation of 4500 m over a high saturated hydraulic conductivity in soil propriety. It contains five stations, one in the central of network and the other four station is 10km away from central over four direction separately. Table 5 is the information of Naqu network.

Table 5, Soil moisture information of Naqu network.

Station name (sensor ID)	*Lat/Lon Elev. (m)*	*Depth of sensor*	*below surface (cm)*	*TPC*	*LC*	*BD (kgm−3)*	*STX*
Naqu Station (SM5)	31° 22'/91° 53'	4509	2.5, 7.5, 15, 30, 60	Plain	Grassland	NA	high organic matter, loamy sand
West Station (TSM1)	31° 20'/91° 49'	4506	2.5, 7.5, 15, 30, 60	Plain	Grassland	NA	loamy sand
South Station (TE2, SM2)	31° 19'/91° 52'	4510	2.5, 7.5, 15, 30, 60	Slope of wet land	Wetland	NA	high organic matter, loamy sand
North Station (TE3, SM3)	31° 22'/91° 52'	4507	2.5, 7.5, 15, 30, 60	Flat hill top	Grassland	NA	high organic matter, loamy sand
East Station (TE4, SM4)	31° 22'/91° 55'	4527	2.5, 7.5, 15, 30, 60	Flat hill top	Grassland	NA	high organic matter, loamy sand

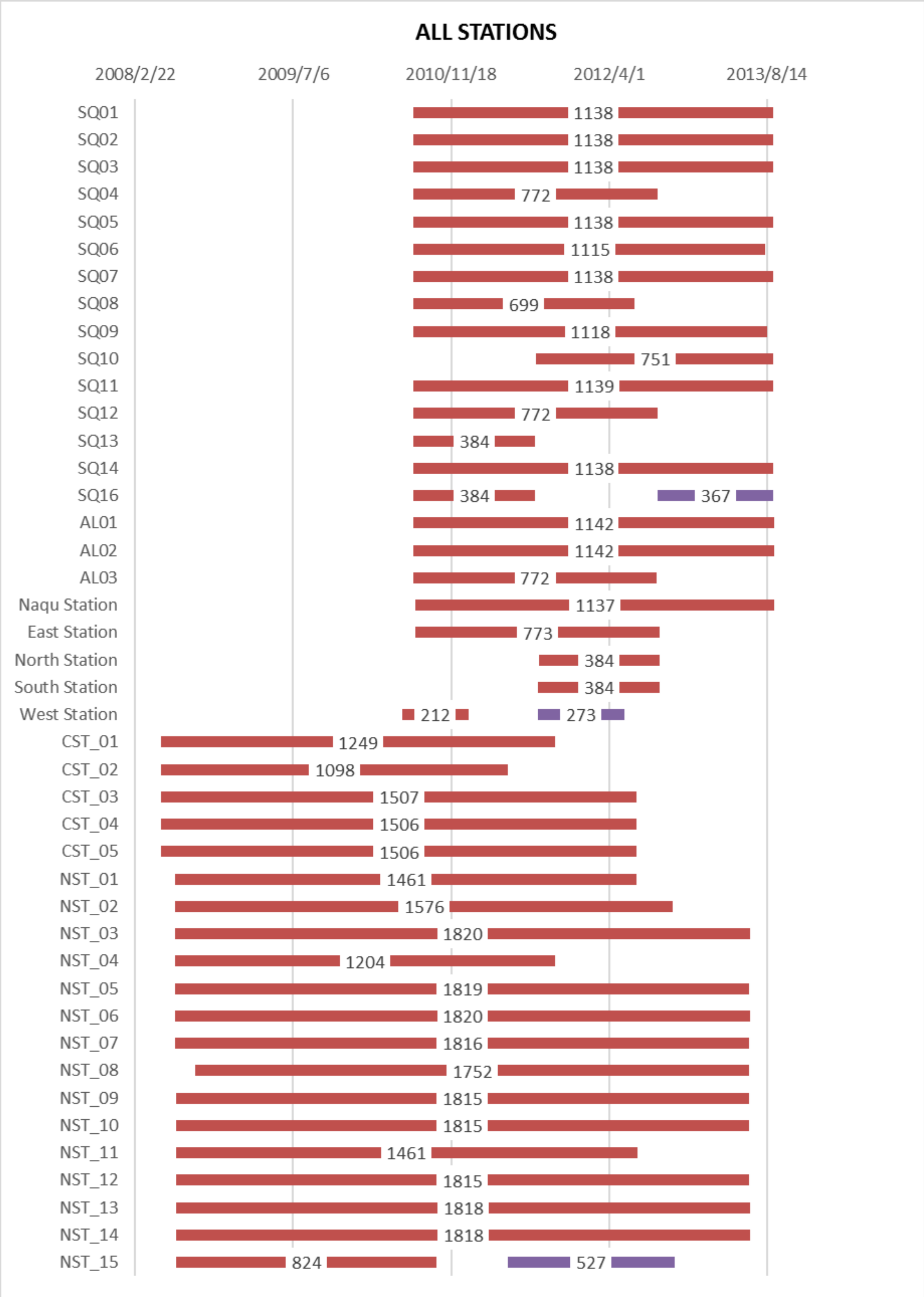


Figure 4, Measurement period of each stations from three subnetworks

3.2.2.2. Maqu network

Maqu network has an elevation of 3500 m and its location is on the north-eastern edge of Tibetan plateau. The land cover of the network is mostly grass with a soil texture of silt loam. It contains 20 stations, 11 stations were installed in the valleys of the Yellow River and Black River, 3 stations in the valleys between hills, 4 stations on steep hill slopes and 2 stations in wetlands.(Su et al., 2011). Table 6 indicates the information of soil moisture and soil temperature of Maqu network.

Table 6, Soil moisture information of Maqu network.

Station Name/ID	*Lat/Lon*	*Elev. (m)*	*Depth (cm)*	*TPG*	*LC*	*BD (kgm−3)*	*STX*
CST_01	33°53´/102°08´	3431	5, 10, 20 40, 80	River valley	Grass	NA	NA
CST_02	33°40´/102°08´	3449	5, 10, 20 40, 80	River valley	Grass	NA	NA
CST_03	33°54´/102°58´	3507	5, 10, 20 40, 80	Hill valley	Grass	NA	NA
CST_04	33°46´/102°43´	3504	5, 10, 20 40, 80	Hill valley	Grass	NA	NA
CST_05	33°40´/102°53´	3542	5, 10, 20 40, 80	Hill valley	Grass	NA	NA
NST_01	33°53´/102°08´	3431	5, 10, 20 40, 80	River valley	Grass	0.96	Silt loam
NST_02	33°53´/102°08´	3434	5,10	River valley	Grass	0.81	Silt loam
NST_03	33°46´/102°08´	3513	5,10	Hill slope	Grass	0.63	Silt loam
NST_04	33°37´/102°03´	3448	5,10	River valley	Wetland grass	0.26	Silt loam
NST_05	33°38´/102°03´	3476	5, 10, 20 40	Hill slope	Grass	0.75	Silt loam
NST_06	34°00´/102°16´	3428	5, 10, 20 40	River valley	Grass	0.81	Silt loam
NST_07	33°59´/102°21´	3430	5,10	River valley	Grass	0.58	Silt loam
NST_08	33°58´/102°36´	3473	5,10	valley	Grass	1.06	Silt loam
NST_09	33°54´/102°33´	3434	5,10	River valley	Grass	0.91	Sandy loam
NST_10	33°51´/102°34´	3512	5, 10, 20 40	Hill slope	Grass	1.05	Loam-silt loam
NST_11	33°41´/102°28´	3442	5,10	River valley	Wetland grass	0.24	Silt loam
NST_12	33°37´/102°28´	3441	5, 10, 20 40, 80	River valley	Grass	1.02	Silt loam
NST_13	34°01´/102°56´	3519	5, 10, 20 40	valley	Grass	0.67	Silt loam
NST_14	33°565´/102°07´	3432	5,10	River valley	Grass	0.68	Silt loam
NST_15	33°51´/102°53´	3752	5,10	Hill slope	Grass	0.78	Silt loam

3.2.2.3. Ngari Network

Ngari is located in the western part of Tibetan plateau with an elevation of 4300 m. This network is in a cold arid environment, but some of the stations located near the Shiquanhe River. It consist of 20 stations shown in Table 7.

Table 7, Soil moisture information of Ngari network.

Station Name /ID	*Lat/Lon*	*Elev. (m)*	*Depth (cm)*	*TPG*	*LC*	*STX*
SQ01	32°29´/80° 04´	4306	5,5,5,10, 20	flat	Desert	Fine sand with gravel(0–10 cm)
SQ02	32°30´/80° 01´	4304	5,5,5,10, 20	Gentle slope	Desert	Fine sand with gravel(0–15 cm)
SQ03	32°30´/79° 58´	4278	5,5,10,20,40	Gentle slope	Desert	Fine sand with gravel(0–15 cm)
SQ04	32°30´/79° 57´	4269	5,5,10,20,40	Edge of a wetland	Sparse grass	Loam to loamy sand
SQ05	32°30´/79° 55´	4261	5,5,10,20,40	Edge of a marsh	Sparse grass	Loam with roots
SQ06	32°30´/79° 52´	4257	5,10,20, 40,80	flat	Sparse grass	Sand
SQ07	32°31´/79° 50´	4280	5,5,10,20,40	flat	Desert	Sand
SQ08	32°33´/79° 50´	4306	5,10,20, 40,60	flat	Desert	Fine to coarse sand
SQ09	32°27´/80° 03´	4275	5,5,10,20,40	flat	Desert/river bed	Fine sand with gravel and bigger rocks (0–5 cm)
SQ10	32°25´/80° 00´	4275	5,10,20, 40,80	flat	Grassland	Fine sand with some thick roots (0–20 cm)
SQ11	32°27´/79° 58´	4274	5,10,20, 40,60	flat	Grassland with bushes	Fine sand (0–5 cm), loamy sand with roots (5–30 cm)

SQ12	32°27´/79° 56´	4264	5,10,20, 40,60	flat	Edge of riverbed	Fine to coarse sand (0–5 cm), loamy with roots (5–40 cm)
SQ13	32°26´/79° 54´	4292	5,10,20, 40,60	flat	Valley bottom	Coarse sand (0–5 cm), fine to coarse sand with roots (5–30 cm)
SQ14	32°27´/80° 10´	4368	5,5,10,2 0,40	flat	Desert	Fine sand with gravel (0–10 cm), fine to coarse sand with roots (10–30 cm)
SQ15	32°26´/80° 11´	4387	5,10,20, 30,50	flat	Bushes	Fine sand (0–15 cm), loam (15–30 cm)
SQ16	32°26´/80° 04´	4288	5,10,20, 40,60	flat	Desert	Loam with gravel and with some clay layers (0–30 cm)
Ali Station	33°23´/79° 42´	4288	5, 10, 20 40, 80	flat	Grass	Loamy sand with roots (0–20 cm)
AL01	33°26´/79° 44´	4262	5,10,20, 40,60	flat	Sparse grass	Fine to coarse sand with roots (0–10 cm)
AL02	33°27´/79° 37´	4266	5,10,20, 30,50	flat	Sparse grass	Coarse sand with gravel (0–35 cm)
AL03	33°27´/79° 37´	4261	5,10,20, 40,60	flat	Grassland close to wetland	Coarse sand with gravel (0–35 cm)

3.2.3. Precipitation data

3.2.3.1. GLDAS Precipitation

GLDAS precipitation is derived from GLDAS Noah Land Surface Model L4 which is simulated by blending of NOAA/GDAS atmospheric analysis fields, spatially and temporally disaggregated NOAA Climate Prediction Center Merged Analysis of Precipitation (CMAP) fields, and observation based downward shortwave and longwave radiation fields derived using the method of the Air Force Weather Agency's Agricultural Meteorological modeling system (AGRMET). Due to the time interval of soil moisture collected in Tibet-OBS is 15 minutes, a 3 hourly GLDAS Noah Land Surface Model with a 0.25° ×0.25° spatial-scale will be employed(Table 8).

Table 8, Precipitation information of GLDAS Noah Land Surface Model.

NAME	GLDAS_NOAH025SUBP_3H
SPATIAL COVERAGE	(-60.0 to 90.0; -180.0 to 180.0)
TEMPORAL COVERAGE	2000-02-24 to Present
SPATIAL RESOLUTION	0.25 degree x 0.25 degree
TEMPORAL RESOLUTION	3 hours

3.2.3.2. Precipitation of China Meteorological Forcing Dataset

China Meteorological Forcing Dataset (CMFD) is a combination of multiple data sources, including CMA (China Meteorological Administration) station data, TRMM satellite precipitation analysis data, GEWEX-SRB downward shortwave radiation, GLDAS downward shortwave radiation data, Princeton forcing data and GLDAS data (Jie & Kun, 2011). The precipitation data of China Meteorological Forcing Dataset used in this study has a 3 hourly temporal interval with a 0.1°×0.1°spatial-resolution.

3.2.4. Porosity

The porosity is retrieved from The Soil Database of China for Land Surface Modeling. The spatial resolution is 30 arc-seconds. The vertical variation of the soil was categorised by 8 layers and the depth of vertical variation is 2.3 m.

3.3. Source and appearance of errors

For better investigating the performance of the automated quality control program over Tibetan plateau, multiple classification maps will be built, including land cover, soil types, elevation and soil organic matters. According to multiple classification maps, a statistic matrix would be established to figure out the influences of these factors on soil moisture data quality. Then, a process would be built to identify the relationship between erroneous value and multiple factors. This would help to identify and classify the erroneous value types.

4. METHODOLOGY

4.1. Quality Control method

Manual inspection of soil moisture data is always the best way to identify spurious values. It could identify various errors by changing rules for different erroneous situations, and then provides a very accurate quality control result. However, when dealing with vast amount of data, like data from several years with a 15 minutes interval, visual checking is unable to successfully handle with because of limited number of observers with limited time could not timely processing. Therefore, automated quality control provides an advanced method to avoid the problem of manual inspection by identifying with some specified rules.

Historical methods of quality control include spatial regression technique (Jinshing You et al., 2010), triple collocation Gruber et al., 2013 to identify erroneous value in time series of soil moisture. The approach used here is developed by Dorigo et al., 2013 through which errors are flagged against with several rules. There are two different subsets are involved to fulfil requirement of quality control, geophysical consistency methods and spectrum-based algorithms respectively.

4.1.1. Geophysical dynamic range

Before geophysical consistency check, a prerequisite dynamic range checking is applied to detect whether there is value beyond plausible range or not. The threshold method described in Hubbard et al. (2005) proposes to use a boundary system derived from historical recorded observations. However, You et al. (2011) employed a porosity-based threshold method to control for local and regional monitored data instead of using empirical threshold. Due to the spatial coverage of Tibet-Obs Network we applied for, the using plausible range for soil moisture here is from 0 to 0.5 m^3m^{-3} for all stations of 3 different climatic zones.

The soil moisture is normally defined as a volumetric fraction of water contained in the unsaturated soil zones. Thus, the lower limit of soil moisture is 0 presented by no water in the layer. Meanwhile, the upper limit of soil moisture strongly depends on the porosity. As soil cannot be fully filled with water, porosity could not reach 1 in reality instead by typically less than 0.5. Besides, this program is employed by SMAP to qualify satellite-based soil moisture product, so the upper limit need to satisfy the requirement globally. Based on points above, the upper limit is proposed to be 0.5.

4.1.2. Geophysical consistency check

The purpose of checking geophysical consistency is to found out the anomalies by comparing soil moisture with closely related coefficients. In order to appropriately use coefficients, coefficients should be either temporally approximate to in situ soil moisture or derived from locations within a feasible radius. Because of in situ variables are uneasily to access, external dataset provides an available solution for consistency, but may not be coincide with in situ data. Thus, interpolation is inevitable for maintaining coherence. Within this technique, precipitation and air temperature are required variables for consistency checking and both of them are interpolated to hourly interval.

4.1.2.1. Checking by precipitation

Precipitation is of great importance in causing a rise in soil moisture. After precipitation, water from atmosphere enters into the pedosphere to fill soil water content and consequently raise the soil moisture.

This strongly bounded relationship between rainfall and soil moisture is adopted as a regulation for identifying anomalies.
An obvious rise can be detected by these following equations expressed as:

$$\theta_t > \theta_{t-1} \quad (1)$$
$$\theta_t - \theta_{t-24} > 2\sigma_{\theta[t-24,t]} \quad (2)$$

Where θ_t is soil moisture at time step t, while θ_{t-1} represents soil moisture backward one time step from t and $\sigma_{\theta[t-24,t]}$ is the standard deviation of the difference in soil moisture of the preceding 24 hour.
The equation (1) used here is to detect significant rise in soil moisture. After first checking, equation (2) will test if the daily variation of soil moisture exceeds defined range. However, this variation may be caused by precipitation. If a precipitation event is detectable, this precipitation should beyond a minimum value correlated with depth of installed sensor, porosity of the area and the accuracy of sensors. These parameters are employed to define minimum precipitation expressed as:

$$P_{min} = PDA \quad (3)$$

Where P is soil porosity, A is the accuracy of sensors using the unit of m^3m^{-3} and D is depth of sensor. If a detected soil moisture value meets the criteria of equation (1) and (2), after applied by equation (3), which is less than minimum precipitation value will be flagged.

4.1.2.2. Checking by temperature

Soil temperature is another factor to check for geophysical consistency. Soil moisture would be significantly lower freezing of the soil, thus, soil temperature have a capability to detect typical frozen soil. The in-situ soil moisture and in-situ soil temperature used here are both measured using a thermistor from the same probe installed in Tibet-Obs network. Any soil moisture below 0 will be flagged of which the units is degree.

4.1.3. Spectrum Based Approach

4.1.3.1. Spike detection

Spike defined as a sudden rise (fall) in just one step, followed by a fall (rise) with same magnitude to return is initial value. Sometimes a spike may have an opposite peak with same intensity, or may surrounded some negative small peaks. A sensor failure or sudden not supply would generate a spike in time series of soil moisture.

The magnitude of a substantial rise or fall detected in a spike should b at least 100%, which expressed as:

$$\frac{\theta_t}{\theta_{t-1}} > 1.15 \quad (4)$$

or

$$\frac{\theta_t}{\theta_{t-1}} < 0.85 \quad (5)$$

A second criterion introduced to avoid overflagging by build a minimum change in value expressed as:

$$|\theta_t - \theta_{t-1}| \geq 0.05 \tag{6}$$

Where $\theta_t$soil moisture at step t is, θ_{t-1} is soil moisture one step before step *t*.

Equation (4) flags value with enormous change from previous value to distinguish from other error types. However, when a soil moisture is too small, even though the value of a change between two continuous steps is critical small, the ratio of the change could beyond 100%. Thus, equation (5) is built up to avoid this situation by defining a minimum change of 0.05

A positive (negative) spike normally surrounded by two negative or positive peaks of which the intensity are almost same. Based on this typical behavior of spikes, an equation is introduced to distinguish spike from precipitation events. This equation compares the magnitude of second derivative of surrounded peaks, soil moisture in time step *t-1* and step *t+1*, and the variation of this comparison range from 0.8 to 1.2 expressed as:

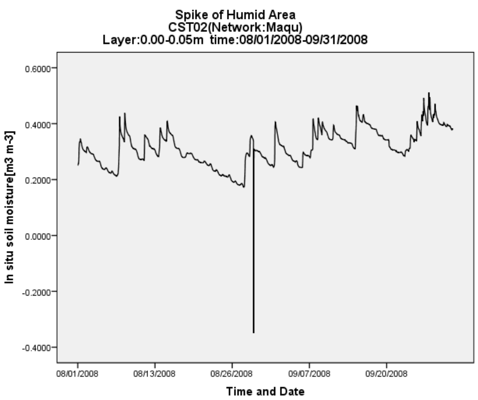
$$0.8 < \left| \frac{\theta''_{t-1}}{\theta''_{t+1}} \right| < 1.2 \tag{7}$$

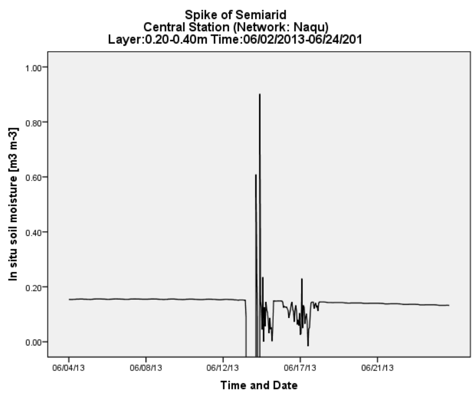
However, for noisy data, these equations could not appropriately detect spike. Thus, another equation is employed to check the variation of the surrounding values, so that spikes can be detected from a noisy data. The equation expressed as:

$$\left| \frac{\sigma^2{}_{(\theta_{t-12},\theta_{t+12})}}{\mu_{(\theta_{t-12},\theta_{t+12})}} \right| < 1 \tag{8}$$

Where $\sigma^2{}_{(\theta_{t-12},\theta_{t+12})}$ is the variance of values over 24 hours but except soil moisture at *t*. and μ is the average of the values coincide with that of variance.

Whenever all these conditions are meet, an in situ soil moisture could be correctly identified as a spike. The Figure 5 below shows three typical spikes of three climatic zones.





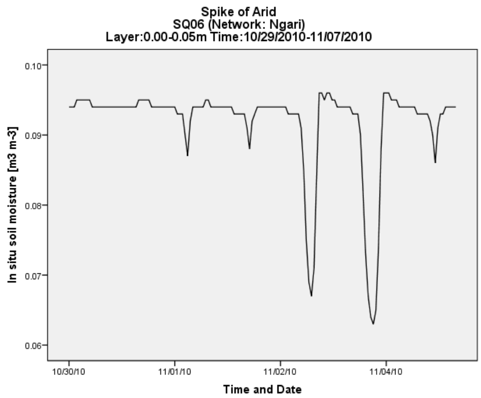


Figure 5, Three typical spikes from Humid, Semiarid and Arid areas.

4.1.3.2. Break detection

A normal break existed in time series expresses as a sudden rise or fall. The difference between a common break and a typical spike is that break is directly from positive value to negative while spike would have an obvious opposite value to offset after rise or fall. This phonemon is often caused by reduced power supply or by sudden change of environment in which the sensor is installed. The following criteria have been established to identify breaks through monitoring its universal characteristics.

According to behavior of breaks, a threshold is used for breaks defines that the relative change should more than 30%. Besides, the absolute variation of soil moisture in a break need to be at least 5% to avoid overflagging. The equation for threshold check expressed below as:

$$\left| \frac{\theta_t - \theta_{t-1}}{\theta_t} \right| \geq 0.03 \quad (9)$$

And

$$\left| \theta_t - \theta_{t-1} \right| > 0.05 \quad (10)$$

Even though these equations are able to detect typical rise or fall in time series, but for break, it also requires changes of preceding time steps so that it could distinguish from spikes. Thus, a first derivative criteria introduced assuming that the first derivative of θ_t should be 10 times larger than all first derivatives of 24 hours centered at time step *t*. leading to:

$$\theta'_t > 10 \frac{1}{n} \sum_{k=-12}^{12} \theta_{t+k} \quad (11)$$

Where θ'_t is the first derivative of soil moisture at time *t*. $\sum_{k=-12}^{12} \theta_{t+k}$ is all derivatives of preceding 12 hours.

Another research found that the peak of comparison of second derivatives between a *t* and a *t+1* is approximate to one in ratio. However, when switching to a comparison of second derivatives between a *t+1* and a *t+2*, the ratio of increase of more than 10 is implausible. According to these two phenomena, two conditions are built up expressed as:

$$\left\| \frac{\theta''_t}{\theta''_{t+1}} \right\| = 1 \quad (12)$$

And

$$\left| \frac{\theta''_{t+1}}{\theta''_{t+2}} \right| \geq 10 \qquad (13)$$

Where θ''_{t} is the second derivative at t, θ''_{t+1} is the second derivative at t+1. θ''_{t+2} is the second derivative at t+2.

After these three break detecting condition is fulfilled, a value could be appropriately flagged as break. Figure 6 below shows three typical breaks of three climatic zones.

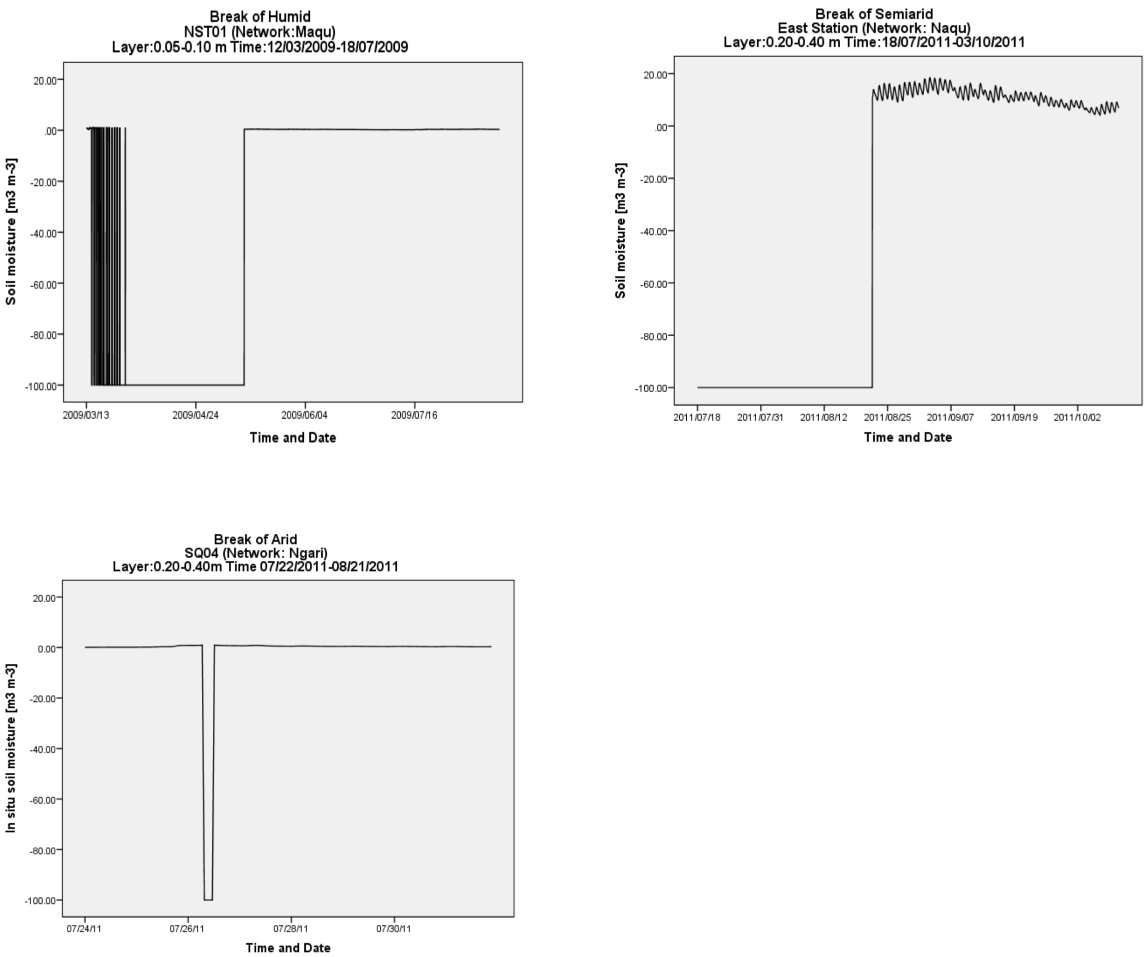


Figure 6, Three typical breaks from Humid, Semiarid and arid areas.

4.1.3.3. Constant value detection

Constant value witnessed as a plateau or a basin over a time series may cause by soil water content exceeds the upper boundary of sensors' sensitivity or on account of long time sensor drop outs and frost periods. To separate constant value from other types of erroneous value, a value should achieve 3 conditions. First criteria defines the lag time and the variance of a constant value as:

$$\mathrm{IN} = [t - n, t + n]$$

while

$$n \geq 6$$

and

$$\sigma^2{}_{[\theta_{t-n},\theta_{t+n}]} \leq 0.0005 \quad (14)$$

Where IN is the whole data period interval centred at t, $\sigma^2{}_{[\theta_{t-n},\theta_{t+n}]}$ is the variance of the interval.

In order to differ constant value from wetting events or spikes, a constant value need to be lasted 12 hours for minimum. Besides, the variance among the whole period of time are not allowed to beyond $0.05m^3m^{-3}$ which is 1% of sensor uncertainty. After checking for the first criteria, a value and its surrounded values over 12 hours are flagged as potential constant value.

Since a plateau normally comes after an intense precipitation, there exist two enormous variations on the edges. These variations placed on the edges are able to be used as indicators to separate a plateau. Thus, the start of a plateau is derived by the first derivative in a local maximum while the end of a plateau is acquired by that of a local minimum, expressed as:

$$\mathrm{t} = t_{plateau\_start} \Leftrightarrow \exists \max([\theta'{}_{t-n-12}, \theta'{}_{t-n+12}]) \geq 0.0025 \quad (15)$$

$$\mathrm{t} = t_{plateau\_end} \Leftrightarrow \exists \min([\theta'{}_{t-n-12}, \theta'{}_{t-n+12}]) \leq 0 \quad (16)$$

However, a plateau is characterized by the highest value over the whole period of a time series. Therefore, another condition is added to detect plateau with a magnitude of 95% of the maximum value over the whole period:

$$\mu\left(\theta_{\left[t_{plateau_{start}} ? t_{plateau_{end}}\right]}\right) > \max\left(\theta_{[t_0,t_{end}]}\right) * 0.95 \quad (17)$$

After 'Break Detection', some values are flagged as breaks, whereas these value are constant value in reality. With the purpose of discriminating constant values from potential breaks, a division between variance and average is measured as second regulation showed below:

$$\left| \frac{\sigma^2{}_{(\theta_t,\theta_{t+n})}}{\mu_{(\theta_t,\theta_{t+n})}} \right| < 0.01 \quad (18)$$

Where $\sigma^2{}_{(\theta_t,\theta_{t+n})}$ is the variance of values over a period of time from t to $t+n$, n has a minimum length of 12. While μ is the average of the values coincide with that of variance.
With these two condition, a constant value could be identified. The Figure 7 below shows three typical constant values of three climatic zones.

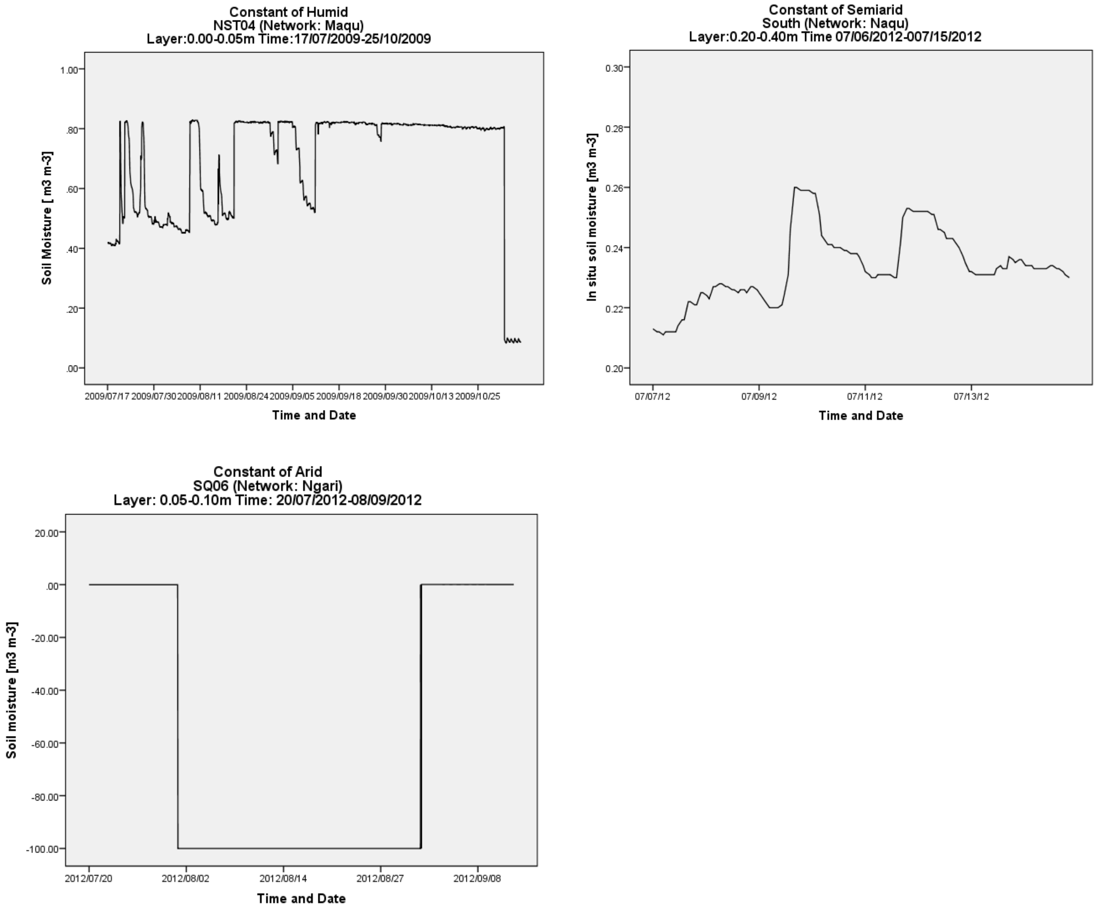
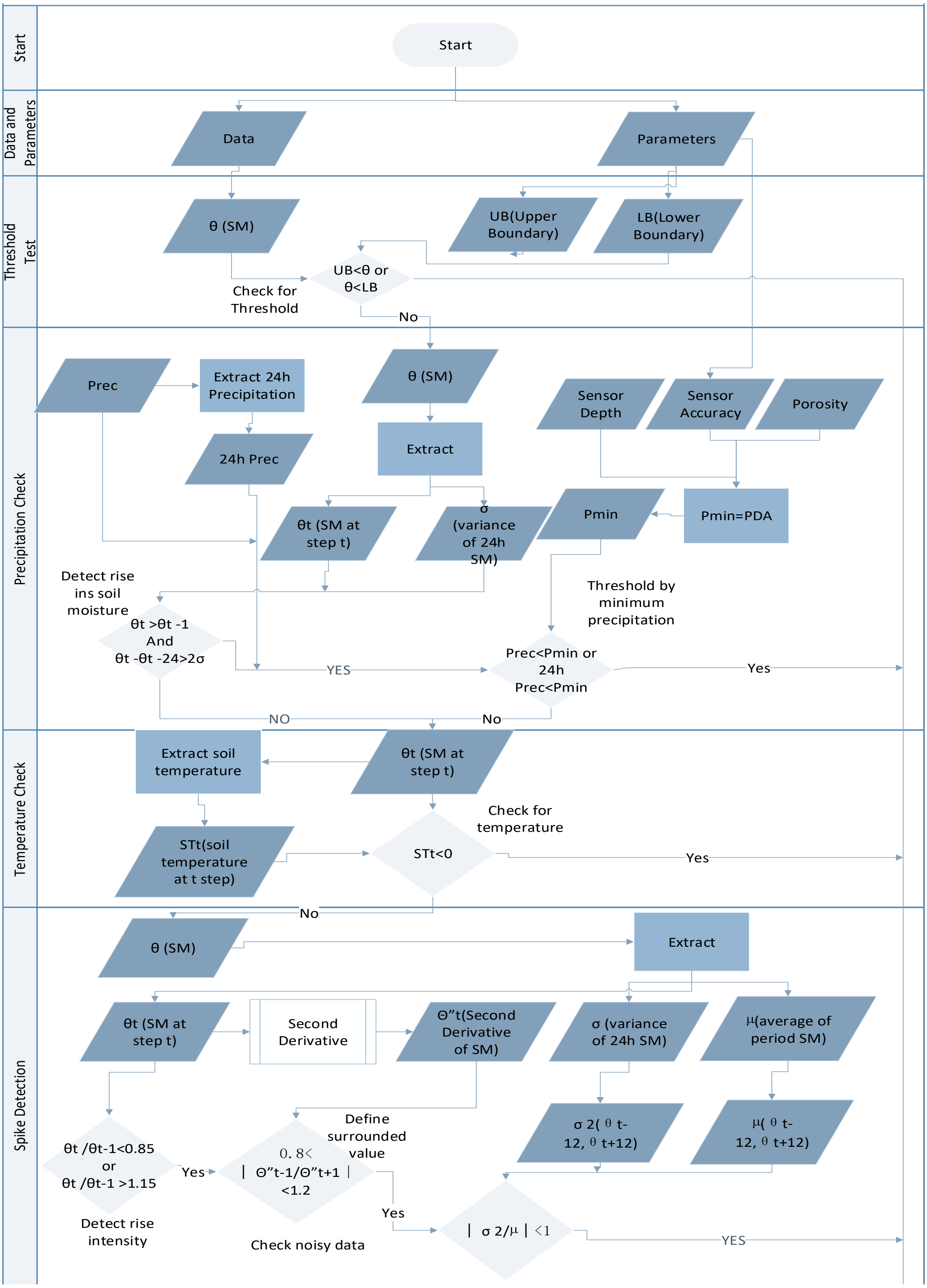
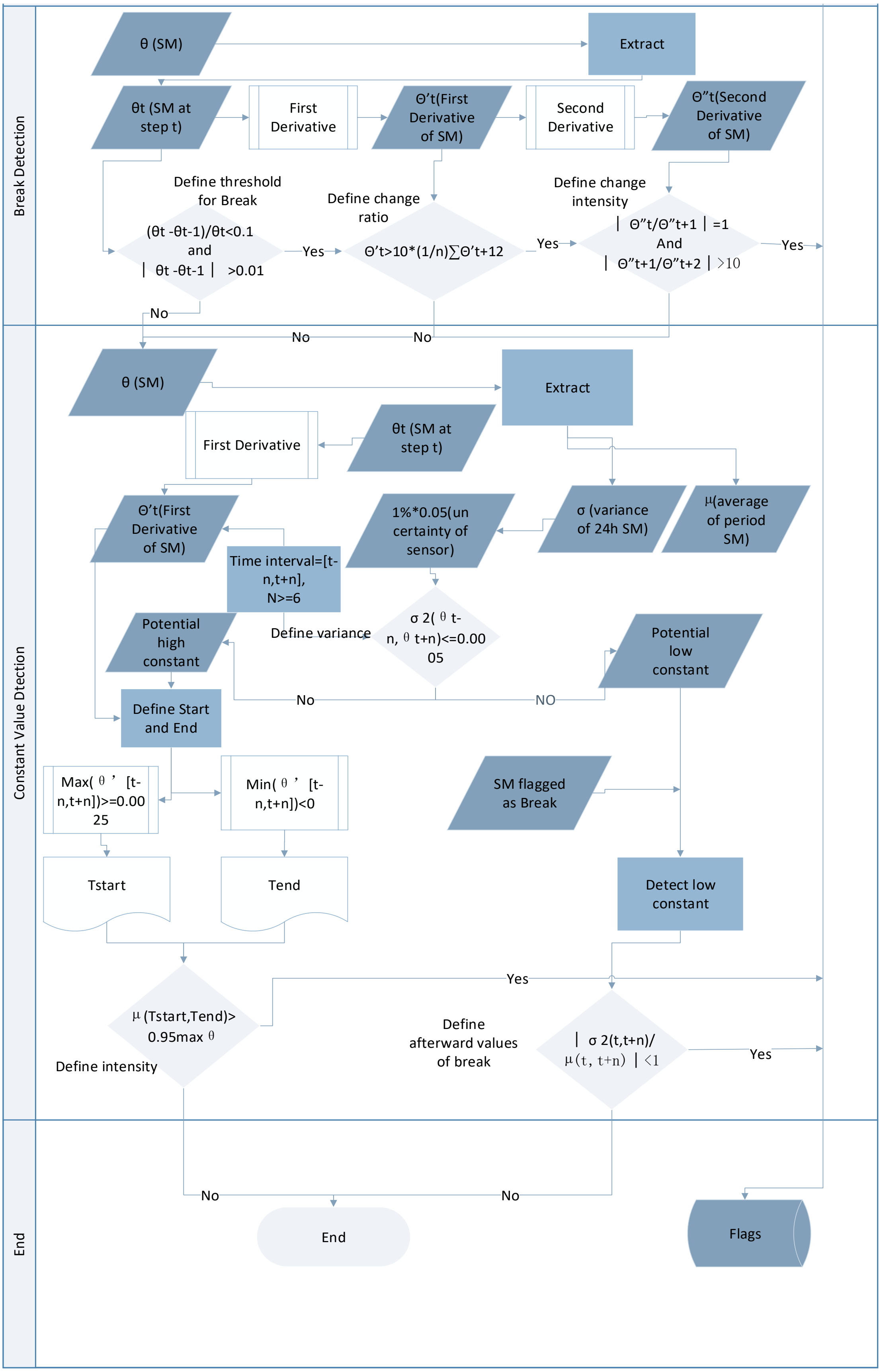


Figure 7, Three typical breaks from Humid, Semiarid and arid areas.

4.1.4. Flowchart of QC program


Start
Data and Parameters
Threshold Test
Precipitation Check
Temperature Check
Spike Detection
Start
Data
Parameters
θ (SM)
UB(Upper Boundary)
LB(Lower Boundary)
UB<θ or θ<LB
Check for Threshold
No
Prec
Extract 24h Precipitation
24h Prec
θ (SM)
Extract
Sensor Depth
Sensor Accuracy
Porosity
Pmin
Pmin=PDA
θt (SM at step t)
σ (variance of 24h SM)
Detect rise ins soil moisture
θt >θt -1 And θt -θt -24>2σ
Threshold by minimum precipitation
YES
Prec<Pmin or 24h Prec<Pmin
Yes
NO
No
Extract soil temperature
θt (SM at step t)
STt(soil temperature at t step)
STt<0
Check for temperature
Yes
No
θ (SM)
Extract
θt (SM at step t)
Second Derivative
Θ"t(Second Derivative of SM)
σ (variance of 24h SM)
μ(average of period SM)
σ 2(θ t-12, θ t+12)
μ(θ t-12, θ t+12)
θt /θt-1<0.85 or θt /θt-1 >1.15
Yes
Define surrounded value
0. 8< | Θ"t-1/Θ"t+1 | <1.2
Yes
| σ 2/μ | <1
YES
Detect rise intensity
Check noisy data

The upper flowchart shows four tests and the lower flowchart shows two detection.



4.2. Procedure

4.2.1. Flowchart of procedure

Quality control process helps develop the correctness of data by identifying the anomalous values. Here, the quality control procedure especially employed to investigate the abnormal in situ soil moisture data from various stations over long period. The main objective of the study is to evaluate the performance of using an automated quality control process for in situ soil moisture data over Tibetan Plateau. The following chart shows methodology:

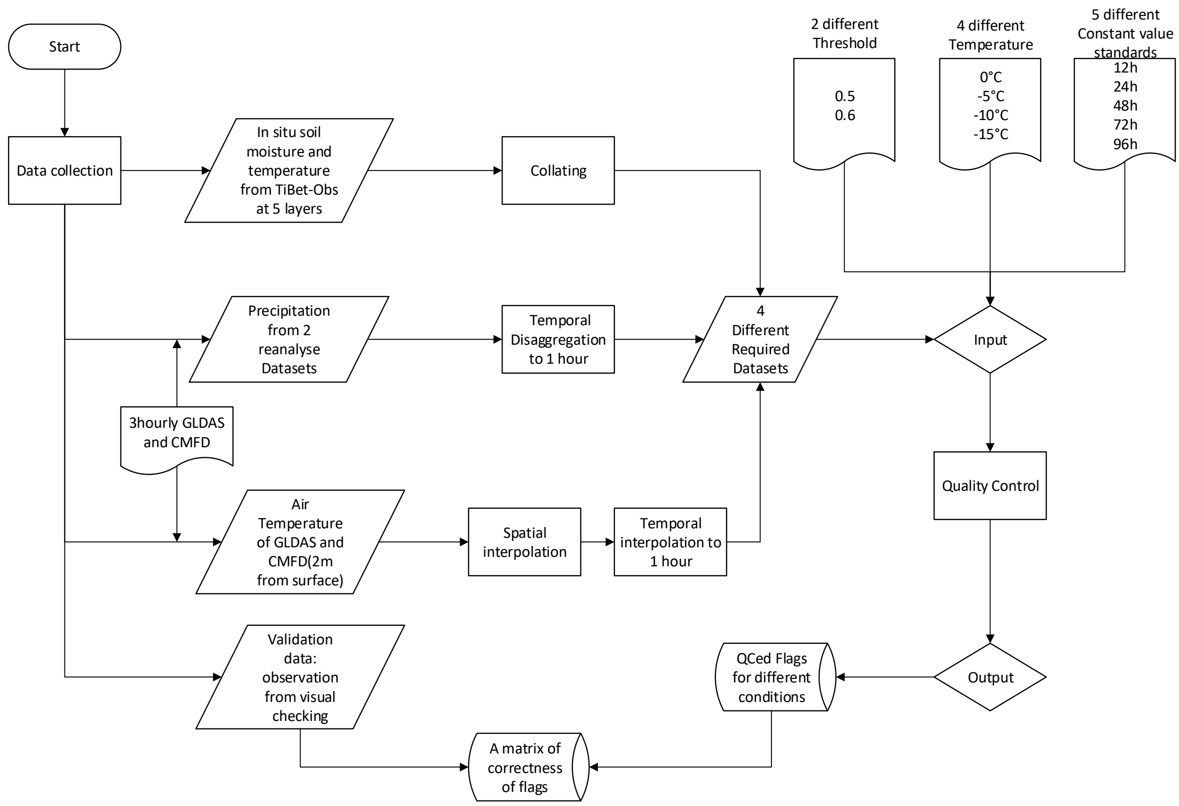


Figure 8, a flowchart shows the procedure of sensitive analyse of quality control

The flow chart of quality control is shown in Figure 8. The collected data are four types including, in situ soil moisture and soil temperature, precipitation and air temperature. After collecting, in situ soil moisture and soil temperature would take a collation procedure to homogenize into a standard format. Due to 2 reasons, precipitation data are disaggregated into 1 hourly to fit with in situ soil moisture. Firstly, precipitation data collected from GLDAS and CMFD both are 3 hourly. Moreover, this quality control systems only based on hourly data to support its accuracy. Air temperature was also resampled into 1 hourly. However, air temperature was spatial interpolated to several stations where there is no value in CMFD dataset. 3 different types of conditions (Threshold, Temperature and Constant values) were used to qualify 4 various datasets. The validation dataset used to compare with qualified flags was obtained by visually inspection of time series of all stations over whole time. A matrix resulted from the comparison concludes the correctness of flags by whether the qualified flags against the validation or not.

4.2.2. Collating Data

4.2.2.1. In situ soil moisture and soil temperature

The in-situ soil moisture of Tibet-Obs network consist of three subnetworks including Maqu, Naqu and Ngari of which the installation time of sensors dispersed variously. This brings an evitable fact that both recording period and recording formats among these networks are different.

Thus, multiple time recording format, different time interval and missing time are derived products of the fact. The time interval issues existed in soil moisture contains 4 types, including utilization of 10 minutes and 15 minutes two different time intervals, monthly recording for just 24 hours, minutely observation recordings and one hour recording at 15 minute with one hour pause for next recording. Exchanging the time interval directly from 15 minutes and 10 minutes to 1 hour could solve most of the problems. While for one hour pause, an interpolation is applied by using linear interpolation method. The monthly records are directly removed. Moreover, missing values have been interpolated for short time span based on time consistency, while for long time span, the value with missed time would be directly disposed. And time records format is another crucial issue obstruct appropriately extract soil moisture. The involved time format is based on Chinese date and time types including both long term format and short time format. For keeping the consistency of soil moisture, the time format of all stations from various networks are unified into same format (i.e. mm/dd/yyyy hh:mm) and later restored in different columns.

Besides, the erroneous records of automated meteorological station result in a mistakenly repeated registration which would be simply removed. Empty data defines a data on various layers at any time step either missing soil moisture or missing soil temperature. For solving this problem, linear interpolation is used as an effectively way for short term, while for long term, removing data is more suitable. All above procedures for in situ soil moisture allow data appropriately input quality control program in a requested format.

4.2.2.2. Precipitation

There are two precipitation datasets used in this study, one from China Meteorological Forcing Dataset (CMFD) and another one from Global Land Data Assimilation System (GLDAS). Based on the latitude and longitude of each station, precipitation data are extracted on local scale by using MATLAB software. Both precipitation data have fully covered the whole period of soil moisture but at a temporal resolution of 3 hour. The required time interval of quality control procedure is one hour which ensures step check. Thus a disaggregation for precipitation is implemented based on a simplified discrete disaggregation model (Ormsbee, 1989) by equal distribution of 3 hour accumulated rainfall to 1 hour showed in Figure 9.

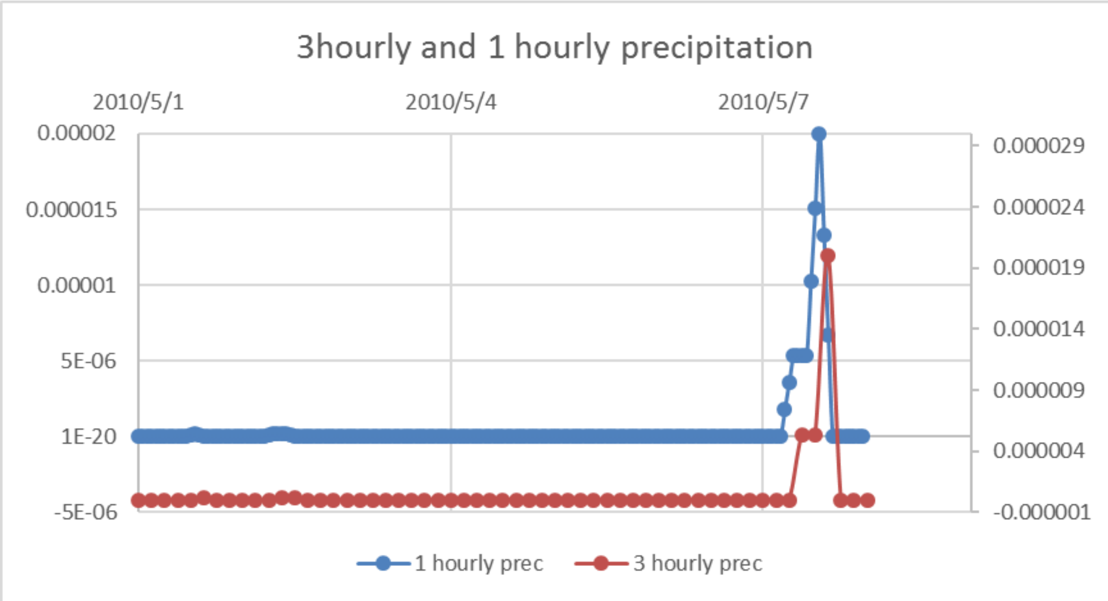


Figure 9, Comparison of interpolated 1 hourly and original 3 hourly precipitation

4.2.2.3. Air temperature

In order to keep the consistency of precipitation and temperature, the air temperature also employed two types, CMFD and GLDAS datasets. The time scale of air temperature is coincided with in situ soil moisture. Due to the limit of spatial scale, specific points on the boundary of coverage may have no value. Thus, to solve this problem, a spatial interpolation based on nearest neighbour is used to acquire air temperature for no value regions. After that, utilization of linear temporal interpolation allows air temperature consistent with precipitation.

4.2.3. Resetting conditions

Resetting conditions is used to reset quality control parameters to change quality control system for fitting with soil moisture of different environmental conditions. Three typical parameters, threshold, temperature and constant value are closely correlated with flagged results.

4.2.3.1. Resetting threshold

Threshold defines the upper boundary and lower boundary of soil moisture which is normally from 0.0 to 0.6 m3m-3. However, this maybe not infeasible for the entire networks, for example the maximum soil moisture of NST_04 station of Maqu Network at 0.05m layer is 0.8 which is much higher than threshold. Notice that this is not just one case for NST_04 station, there still exist large amount of value beyond threshold. For different climatic zones, the highest value of soil moisture of each climatic regions is also different. Thus, this variable in this study are separated into 3 different values,0.5,0.6 and 0.7 to investigate how much is the influence of threshold for various climatic conditions.

4.2.3.2. Resetting temperature

During winter, temperatures below 0 lead to ground freezing, however, the soil moisture of these regions still above 0 (Su et al., 2011). In this study, using four types temperature conditions, 0, -5, -10 and -15 degree instead of 0 degree to flag erroneous data. These temperature conditions are obtained from the relationship of soil temperature and soil moisture which are showed in Figure 1. Considering below -20, all regions may already be frozen, could resulting in overflagging for large amounts. For the purpose to avoid this situation, only four types range from 0 to -15 is employed to investigate.

4.2.3.3. Resetting constant value

Constant value often happens where there is long time sensor failure or long time saturated soil moisture. The original variable of constant value defines the constant should last at least 12 hours so that can be identified. However, this may overflag values just have short period of high values. Thus, in this study, an investigation of effects of using different constant time by applying 12 hours, 24 hours, 48 hours, 72 hours and 96 hours for research.

4.2.4. Result analyse

According to the objective, 4 different required datasets have been done at 3 categories with 11 types. Each of them is hourly data covers the period from 2008 to 2013. As installation time of each station were not totally the same, counting by numbers of each type of error is inappropriate. Thus, the first step of validation is calculating the percentage of each station for various error types at multiple depths. Based on the subobjectives, the investigation is mainly focus on the influence of different geophysical conditions on quality control flagging. Thus, there are three ways in analysing the results.

4.2.4.1. Soil type, soil organic matters and elevation analyse

Soil types, soil organic matters and elevation as three influence factors are analysed. The purpose of these analyse is to carry out the soil condition associate to the flag patterns. Thus, these analyses will build over various layers to figure out how the influence of deep layers compare with shallow layers. And this analyses could establish a retrievable standard for soil moisture of Tibet network.

4.2.4.2. Climatic analyse

Due to climatic condition defined mainly based on its precipitation and temperature,
there is a strongly link between climatic condition and soil moisture. The errors will be validated are categorized into three types based on the climatic conditions. Further research for various soil types will be applied by subdividing the errors into soil organic matters, soil texture and height.

4.2.4.3. Validation

The validation dataset is created by visually checking the original data. To compare with quality controlled data, a matrix is built by categorizing results into two types.
These two types define whether the quality controlled result is correct or not. However, through visually checking, only data with specific flags can be seen. Via this validation, the credibility of quality control method could be guaranteed.

5. RESULTS AND ANALYSE

5.1. Code modification

The quality control systems employed in this study is based on the spectrum-based methods. The code used in this study is from SMAP CalVal team, who aims to enhance the quality of soil moisture of near surface for remote sensing observations worldwide. Nevertheless, for the QC of in-situ soil moisture, the code has to be modified to follow the rules provide by Dorigo et al., 2013. Figure 10 is an example of CST_05 station of Maqu at layer of 5 cm before modification, while Figure 11 is the same time series but the equations are modified. It is clearly showed that large fraction of the values is flagged in Figure 10 but in Figure 11 only jumps and very low values are flagged. Though most of soil moisture after being modified looks like flags, there will be a discussion about it later.

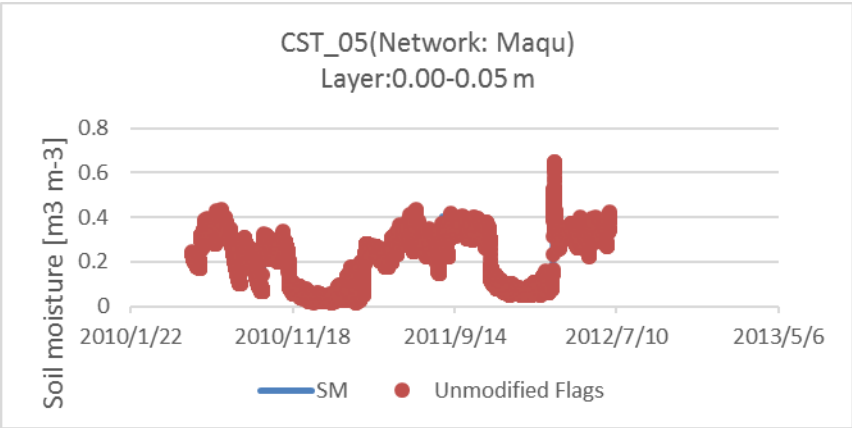


Figure 10, Example of flags before modification

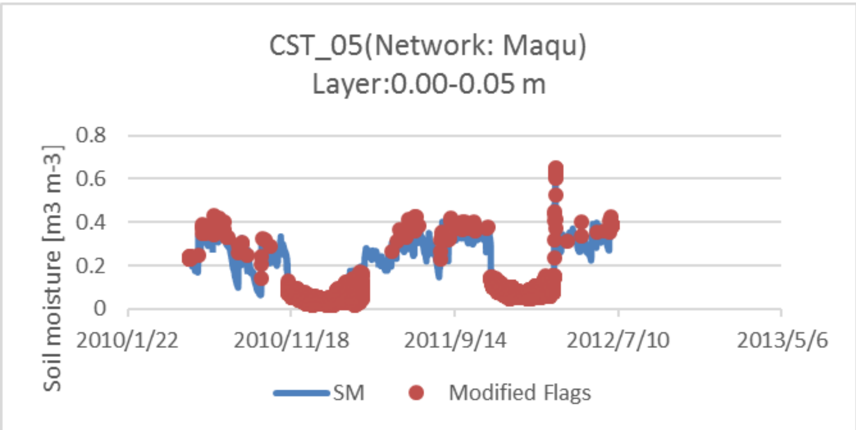


Figure 11, Example of flags after modification

5.1.1. Temperature modification

Three modifications were applied for the code, temperature threshold, spike detection and constant value detection. The change of temperature threshold is from 4 degree to 0 degree to fit the situation over TP, where one fined wide-spread frozen grounds. Moreover, according to Su et al., 2011, even though the temperature is below 0, there will still exist liquid water content, this poses a challenge to the applicability of current QC code for TP. The result in Figure 12 illustrates that after modifying the temperature threshold the fraction of flags declined significantly. Before modification, the values flagged fraction by temperature increased with deeper soil layers. However, after modification, this phenomenon improved much at layers beyond 40 cm, such as modification at 80cm decreased from 53% to 25%. The illustration in Figure 13 obviously exhibit that the difference between each change is raised by the depth. A possible reason registered for this case is that large amount of soil moisture accompanied with soil temperature relatively low at deep layers. Though the modification gives an enormous improvement, among three networks of Tibet network, the lowest temperatures of each network are different as well. The corresponding soil temperature at multiple depth of various climate also varies. The further research would be inspected in this part.

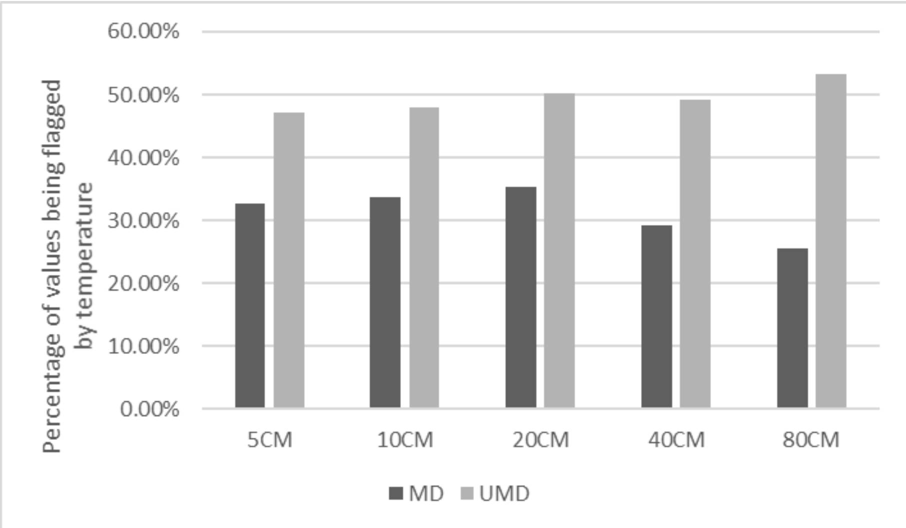


Figure 12, Comparison between unmodified (UMD) temperature threshold (dark gray) and modification (MD) result (light gray) at various depths.

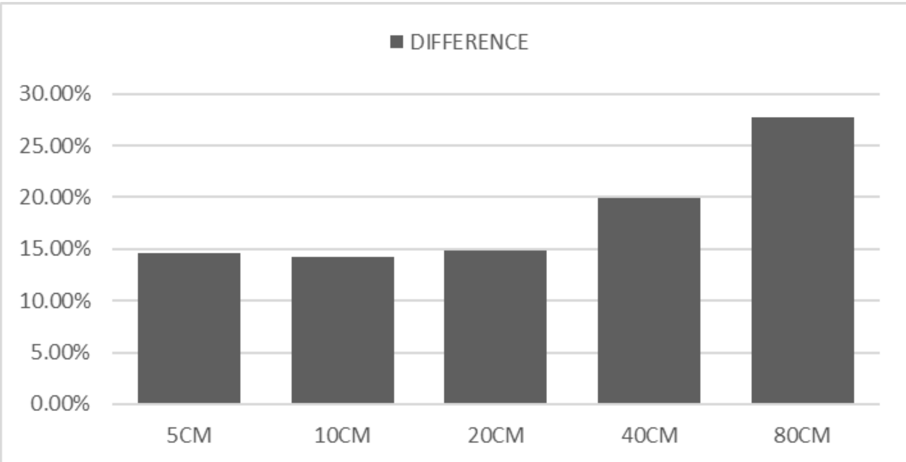


Figure 13, The difference of temperature threshold change at various depth of entire Tibet networks.

5.1.2. Spike modification

Spike detection of soil moisture would mix up with precipitation's consistency checking in the code. In the program, spike detection also use minimum value of precipitation to find out abnormal rise. However, to distinguish the precipitation checking and spike detection, the only difference of spike detection is that it employs the change rate of 100% to find out abnormal values (e.g. the current time step SM observation is two times larger than the previous time step SM observation). For a humid climatic region, it is tricky to use such spike detection criteria and may lead to fault flagging. In arid regions with low soil moisture, the change rate with 100% is a rational number. Thus, the spike detection is modified by following the rules of Dorigo et al., 2013. Figure 14 shows the result of spike detection for the original code and after modification. Comparing with temperature modification, the fraction of soil moisture flagged as spike is considerably lower with a change less than 10%, but the trend of which over soil profile is same as the trend of changes in temperature modification. It obviously shows that soil moisture of deeper layers would be flagged more as spike. The possible reason for this case is that the soil moisture in deep layers may frequently happens inconsecutive drop out which would be detected as spikes. Even though, the modification has not changed much on spike detection but Figure 15 indicate that difference of spikes in shallow layers and deep layers is still worth of attentions. The average difference of shallow layer is almost 0.2% which is two time higher than that of deeper layers.

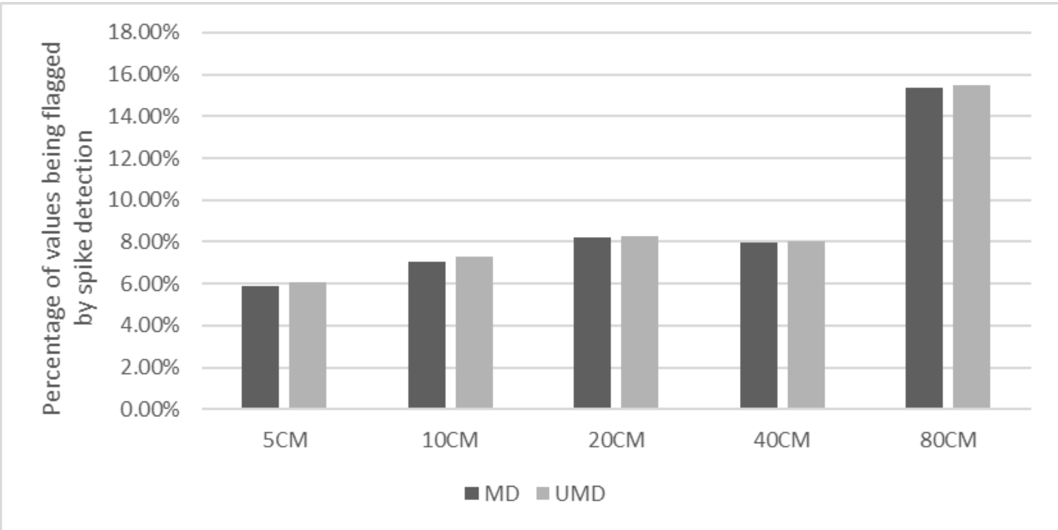


Figure 14, Comparison between unmodified (UMD) spike and modification (MD) result at various depths.

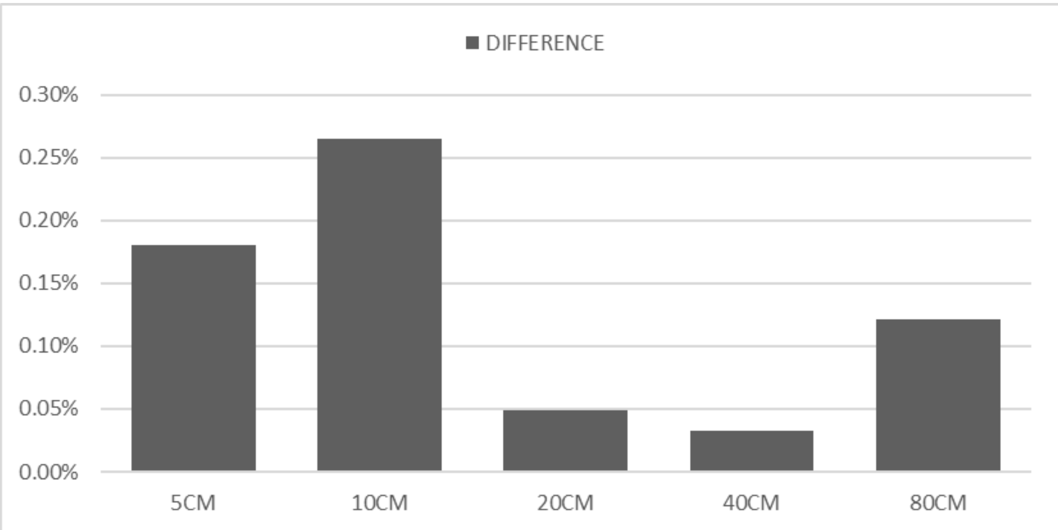


Figure 15, The difference of change in spike at various depth of entire Tibet networks.

5.1.3. Constant value detection modification

Another issue tremendously affects assurance of quality control is constant value detection. The original code has a potential drawback that the program only detects values in constant status. The modified code defined the start and end for positive constant values as well as the upper boundary of constant flags. For negative constant, it gives a more clearly definition by following with a negative break. The dramatically reduction between modified result and unmodified values is obviously revealed by Figure 16. In Figure 17, constant flags of all layers declined significantly with about 90%. The constant flags in original code is almost 100% of multiple layers while after modification the flags are less than 20%. As mentioned above, this is because that original program's overflagging.

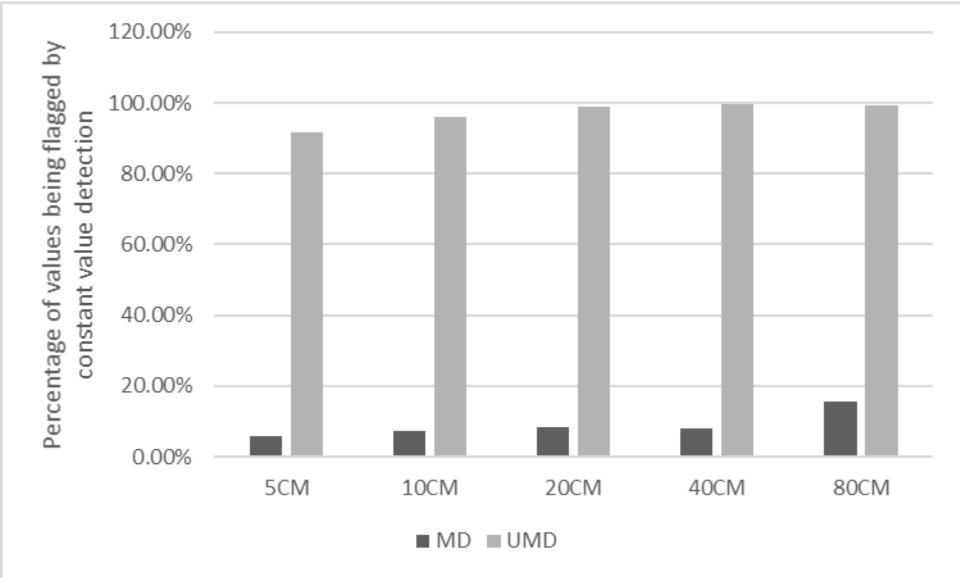


Figure 16, Comparison between unmodified (UMD) constant values and modification (MD) result at various depths.

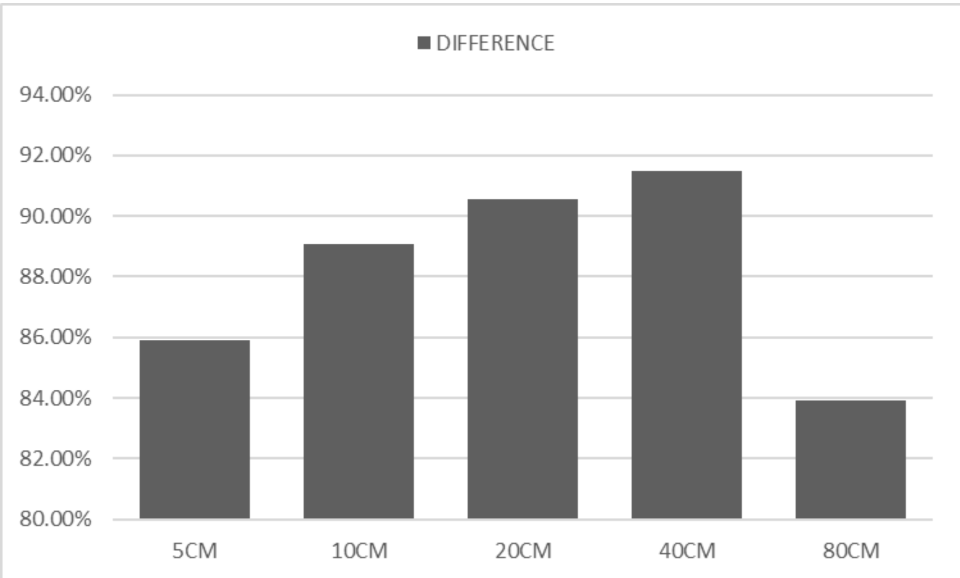


Figure 17, The difference of change in constant at various depth of entire Tibet networks.

Based on above specified modification, quality control result improved significantly. As it showed in Figure 18, flags of all layers decreased form above 100% to fraction of less than 40%, the difference between modified results and unmodified results revealed in Figure 19. However, to find out more specified information about this quality control program in Tibet Network, more detailed analyse is implemented.

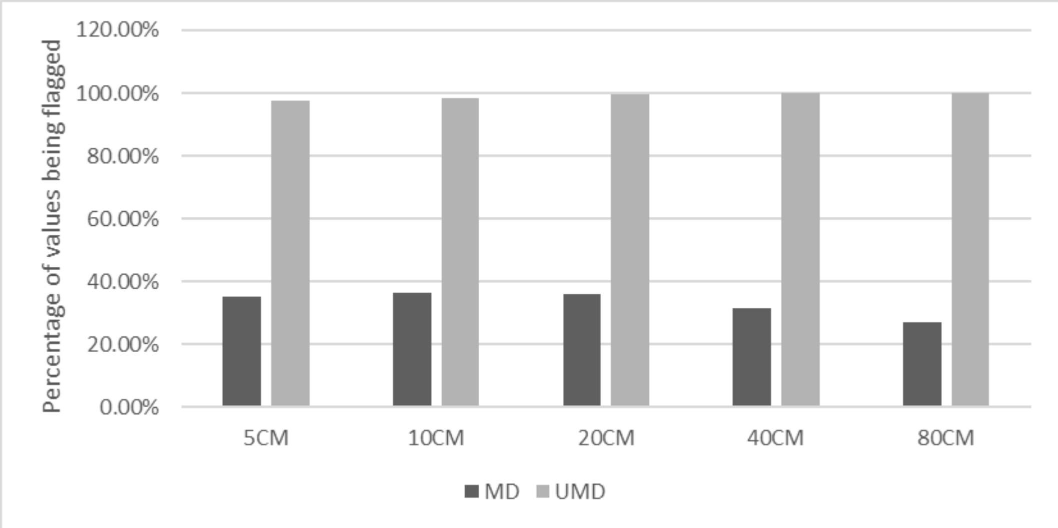


Figure 18, Comparison between unmodified (UMD) flags and modification (MD) result at various depths.

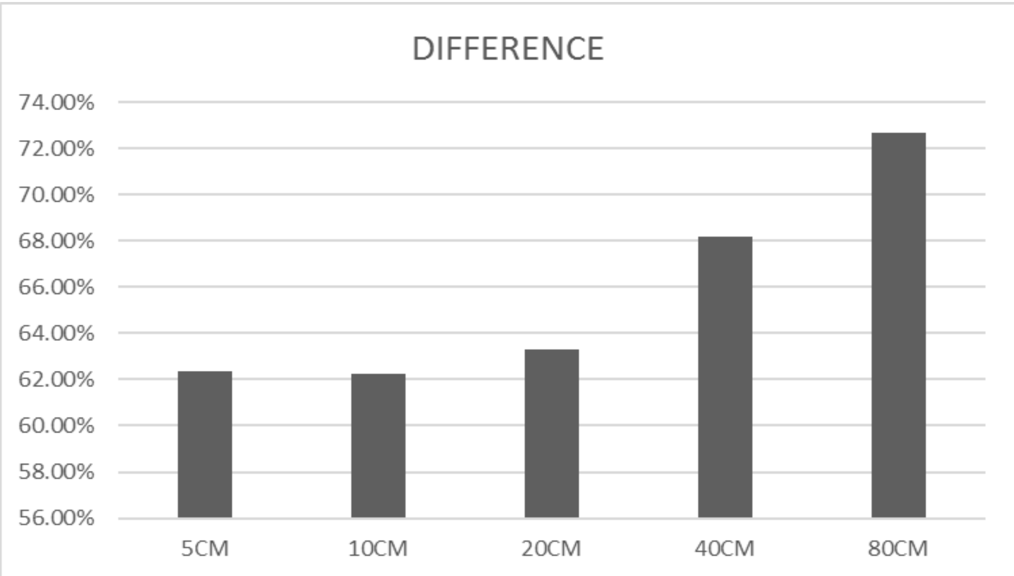
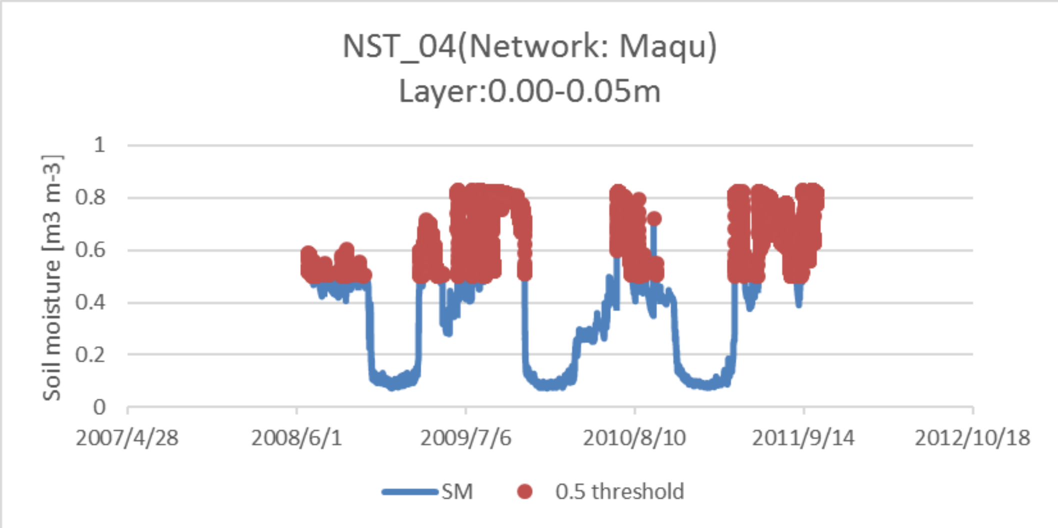


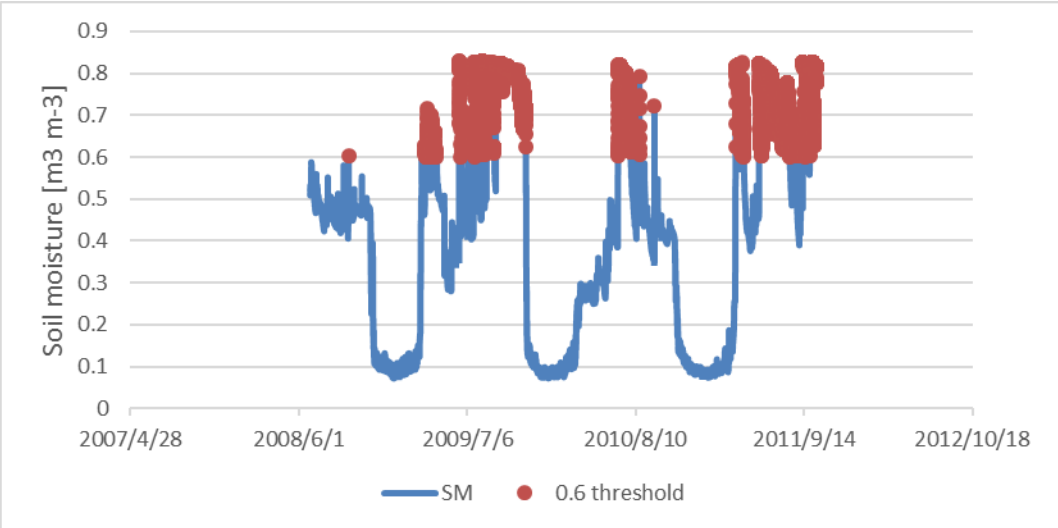
Figure 19, The difference of the comparison between unmodified flags and modified flags.

5.2. Threshold

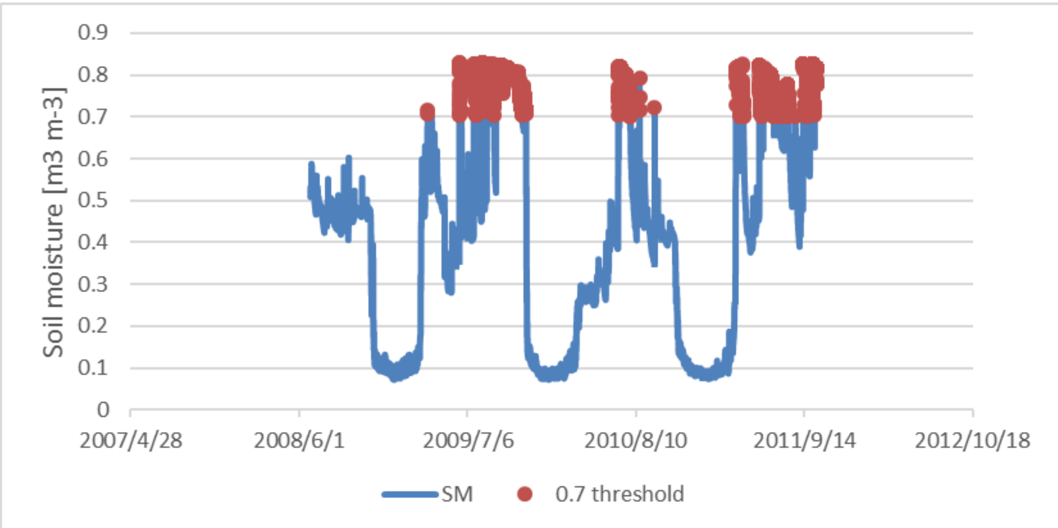
Threshold detection is employed in this program for checking geophysical dynamic range. To fully investigate the threshold, 3 threshold standards were established as 0.5, 0.6 and 0.7 m^3m^{-3}. Figure 19 shows three examples of soil moisture values beyond the range of different specifications. Figure 20a is the soil moisture exceed the range from 0.00 m^3m^{-3} to 0.50 m^3m^{-3}, while the range of 20b is from 0.00 m^3m^{-3} to 0.06 m^3m^{-3}. Figure 20c gives the flags in which flagged soil moisture are higher than 0.7 m^3m^{-3}.



(a)



(b)



(c)

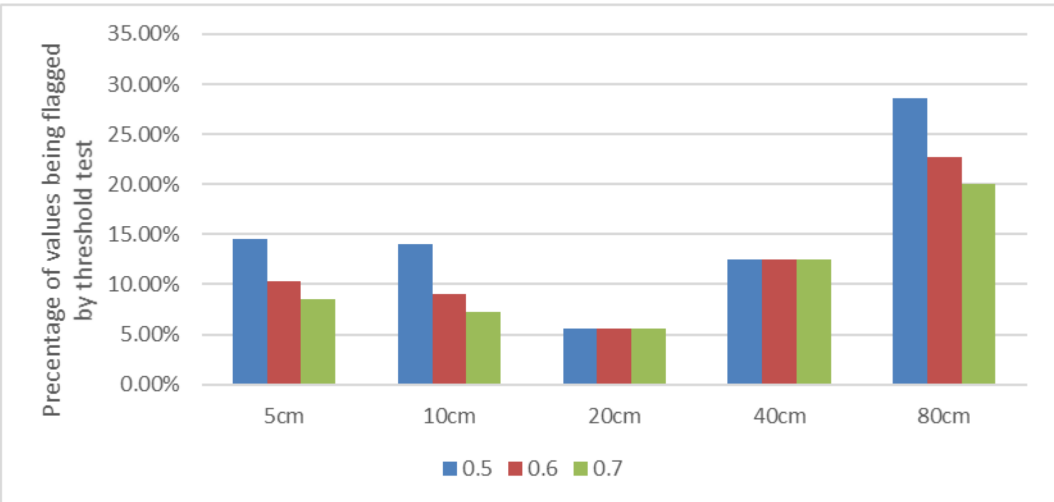
Figure 20, Three examples of threshold checking on different conditions, (a) an example of flags with threshold as 0.5, (b) an example of flagged values beyond 0.6, (c) an example of threshold testing by 0.7.

As we can see, the comparison of these flags is considerably obvious. All soil moisture beyond the threshold is flagged, showing that the threshold checking is significantly accurate and effective. However, quantity flags should not be recognized as spurious value. But the figures revealed that with the increase of threshold, the fraction of soil moisture as flags are decreased.

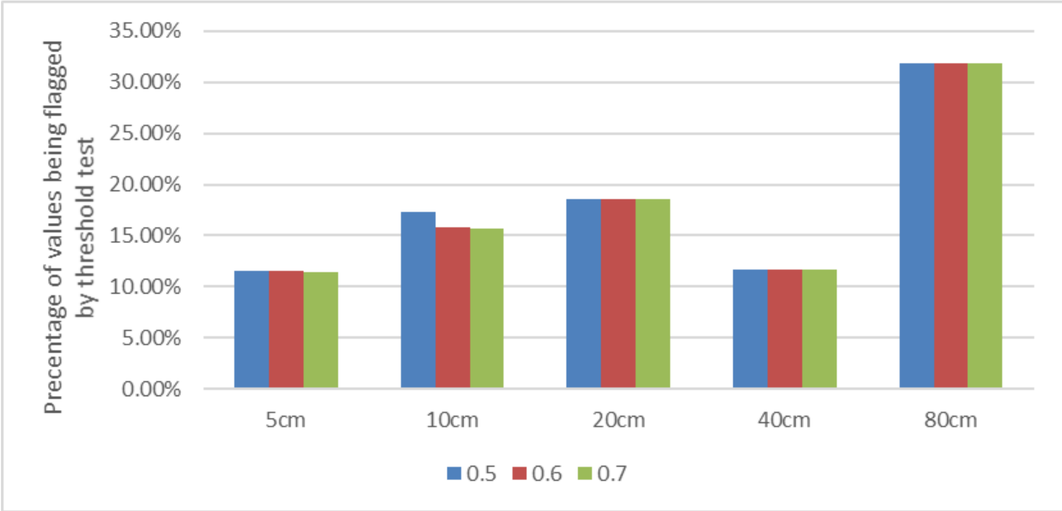
For different climatic zones, the sensitivity of resetting threshold is distinctively as well. In Figure 21a which is the comparison of humid areas over 5 layers, showing that no obvious change is witnessed at 20cm and 40cm, while at 5cm, 10cm and 80cm, there exist huge improvement. When threshold from 0.5 m^3m^{-3} changed to 0.6 m^3m^{-3}, the soil moisture of 80cm of humid flagged by threshold has largest reduction with 7% decrease from about 29% to 22%. Behind it, the flagged soil moisture of 5cm gives the minimum change with just 4.4% from 14.4% to 10%. Among them, the flagged soil moisture of 10cm layer in humid took up 5% from 15% to 9%. When threshold changed from 0.6 m^3m^{-3} to 0.7 m^3m^{-3}, the trends of decline for both flagged soil moisture of 5cm, 10cm and 80cm is almost same by a drop with 2%. A possible reason to explain that situation is that soil moisture of shallow layers of humid climate are relatively high and detected as high values beyond the upper boundary of threshold.

Figure 21b shows comparisons of semiarid area, only soil moisture of 10cm is affected by the changes from 17% to 15% after changing the threshold from 0.5 m^3m^{-3} to 0.6 m^3m^{-3}. The soil moisture of 80cm had a high percentage of flags because of soil moisture of deeper layer are suffering to sensor failure, which falling out of the range of threshold.

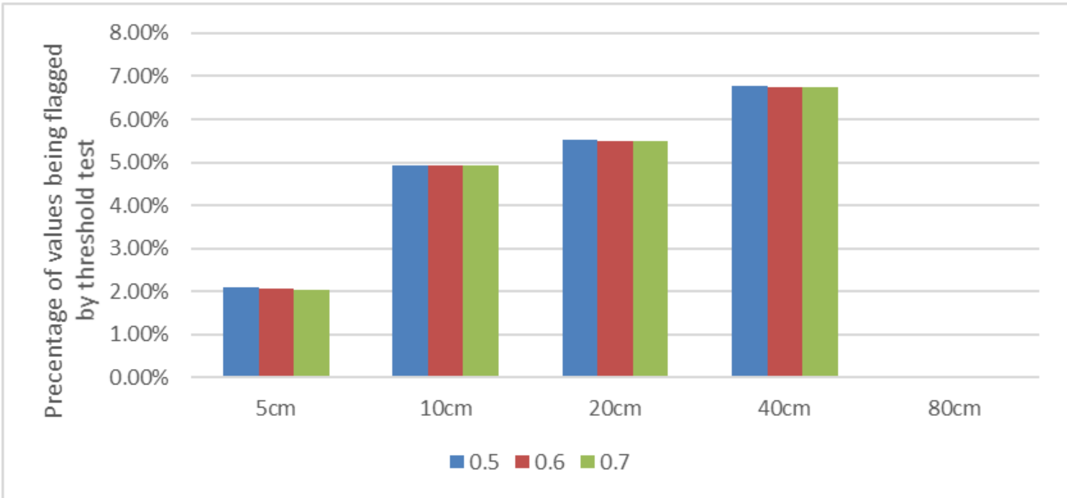
The number of flags indicating low quality data in the arid area is much lower than other climates over all layers which showed in Figure 21c, especially at layer of 80cm. Among the flagged soil moisture of arid area, the lowest flagged layer is 80cm with 0% being flagged, while soil moisture of 40cm reached the highest percentage of 7% as flags. The fact of 0% flags at 80cm is that soil moisture of deep layers in arid is too low to reach threshold as the high aridness lead to almost little soil moisture. However, there are slightly decrease over layers from 5cm to 40cm between each adjustment of threshold.



(a)



(b)



(c)

Figure 21, Comparison of threshold flags over three climatic zones of Tibet-Obs networks, (a) Comparison of flags of Humid area, (b) Comparison of flags over Semiarid Area, (c) Comparison for Arid regions.

Soil moisture of three distinctive climates over multiple layers is contaminated by senor failure resulting in low value that below 0. There was little soil moisture of arid area of all layers being flagged due to soil moisture of arid are much small which cannot go beyond upper boundary. Most obvious development was found at shallow soil layers after adjustment, i.e. soil moisture of 10cm of humid area.

Table 9, A summary of flagged percentage categorized to soil organic matters, soil types and elevations for threshold. Note that, the value in blod text present the highest percentage between each comparison of mean value of various layers.

THREHOLD	**SOIL ORGANIC MATTERS(G/100G)**					**SOIL TYPE(SAND,SILT,CLAY)**		**ELEVATION(M)**			
0.5		0-5	5-10	15-20	>20	SAND	SILT	<3500	3500-4000	4000-4500	>4500
	5cm	6.8%	8.9%	13.3%	0.5%	6.6%	7.8%	9.6%	10.0%	1.9%	19.2%
	10cm	6.2%	3.8%	32.1%	18.9%	6.0%	10.7%	10.4%	4.9%	4.6%	26.4%

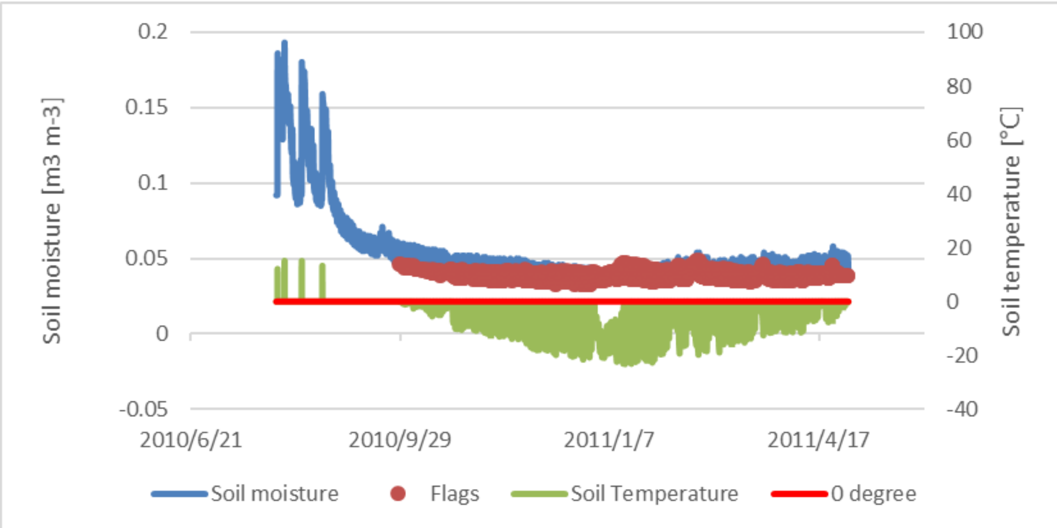
	20cm	7.6%	8.7%	0.0%	0.0%	8.7%	6.4%	3.1%	8.6%	5.2%	31.0%
	40cm	9.0%	16.2%	0.0%	0.0%	9.8%	10.4%	9.6%	15.8%	6.3%	19.5%
	80cm	16.7%	18.7%	0.0%	0.0%	15.4%	20.6%	17.6%	26.5%	0.0%	53.3%
	mean	9.2%	11.3%	**22.7%**	9.7%	9.3%	**11.2%**	10.1%	13.2%	3.6%	**29.9%**
0.6											
	5cm	8.3%	11.0%	18.4%	8.4%	7.2%	11.6%	13.7%	13.9%	2.0%	19.3%
	10cm	7.6%	6.2%	44.5%	28.6%	6.6%	15.5%	15.6%	8.6%	4.7%	28.9%
	20cm	7.6%	8.7%	0.0%	0.0%	8.7%	6.4%	3.1%	8.6%	5.2%	31.0%
	40cm	9.0%	16.2%	0.0%	0.0%	9.8%	10.5%	9.6%	15.8%	6.4%	19.5%
	80cm	16.8%	34.2%	0.0%	0.0%	15.4%	28.0%	28.3%	26.5%	0.0%	53.3%
	mean	9.9%	15.3%	**31.4%**	18.5%	9.5%	**14.4%**	14.1%	14.7%	3.6%	**30.4%**
0.7											
	5cm	6.2%	6.3%	11.0%	0.1%	6.6%	6.1%	7.9%	8.4%	1.9%	19.2%
	10cm	6.1%	0.5%	27.1%	15.7%	6.0%	8.9%	8.9%	2.7%	4.6%	26.3%
	20cm	7.6%	8.7%	0.0%	0.0%	8.7%	6.3%	3.1%	8.6%	5.2%	31.0%
	40cm	9.0%	16.2%	0.0%	0.0%	9.8%	10.4%	9.6%	15.8%	6.3%	19.5%
	80cm	16.6%	11.5%	0.0%	0.0%	15.4%	17.2%	12.5%	26.5%	0.0%	53.3%
	mean	9.1%	8.6%	**19.0%**	7.9%	9.3%	**9.8%**	8.4%	12.4%	3.6%	**29.9%**

Table 9 illustrates the comparison between different soil types, soil organic matters and elevations. The soil moisture of Tibet-Obs are easily influenced by elevation showed that huge difference between soil moisture of different elevations. Comparing with elevation, the results between flags of sandy type and silt type only appeared slight difference. For soil organic matters, soil moisture in soil organic matter of 15 to 20 g/100g gives higher percentage than other types. More soil moisture was flagged at greater soil organic matters over shallow layers, while more soil moisture of low organic matters was flagged at deep layers.

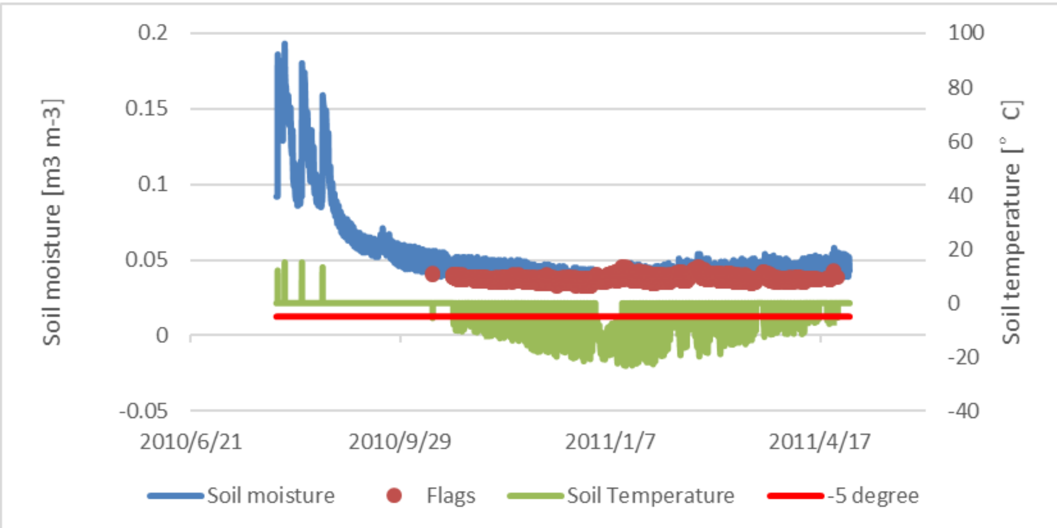
Even though 0.7 m^3m^{-3} as a threshold enhanced much in soil moisture flagging. A large amount of soil moisture from humid area still beyond this threshold which means that this threshold is infeasible for all stations. Because there exist stations with very high level of soil moisture, like NST04 station of Maqu Network, which located under a slope and close to a river. Moreover, the Maqu network is easily affected by monsoon season, leading to very wet condition at any locations. To appropriately define the threshold for specific station like NST04 whose soil moisture is very high, Hubbard et al., 2005 developed threshold standard based on historical values by calculating the standard deviation could successfully flagging spurious values. This method could be introduced into this quality control at local scale over all vertical depth.

5.3. Temperature

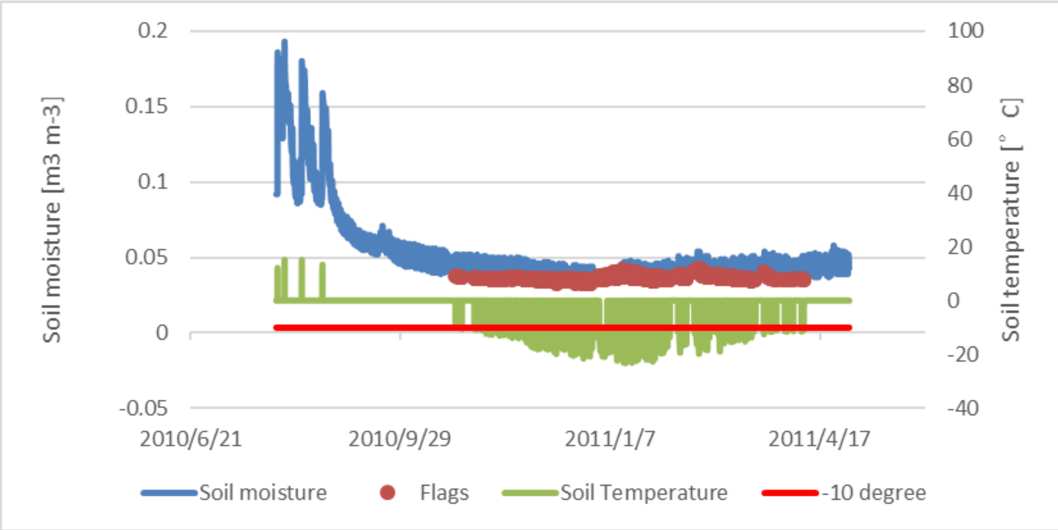
The temperature threshold checking was analysed by four levels, including 0 degree, -5 degree, -10 degree and -15 degree. Conventionally studies consider that there is no soil moisture when the soil temperature is below zero. However, Kurylyk et al., 2013 found that there still exist soil moisture with a value of 0.01 when soil temperature is below -5 degree by experimentally using a linear regression of soil moisture associated with soil temperature. These temperature thresholds were obtained through analysing the appearance of the relationship between soil moisture and soil temperature. The result shows that the temperature threshold affected large amount of soil moisture during winter season in Tibet plateau. Four examples of results of different temperature thresholds are exhibited at sq03 station of Ngari Network in an arid climate.



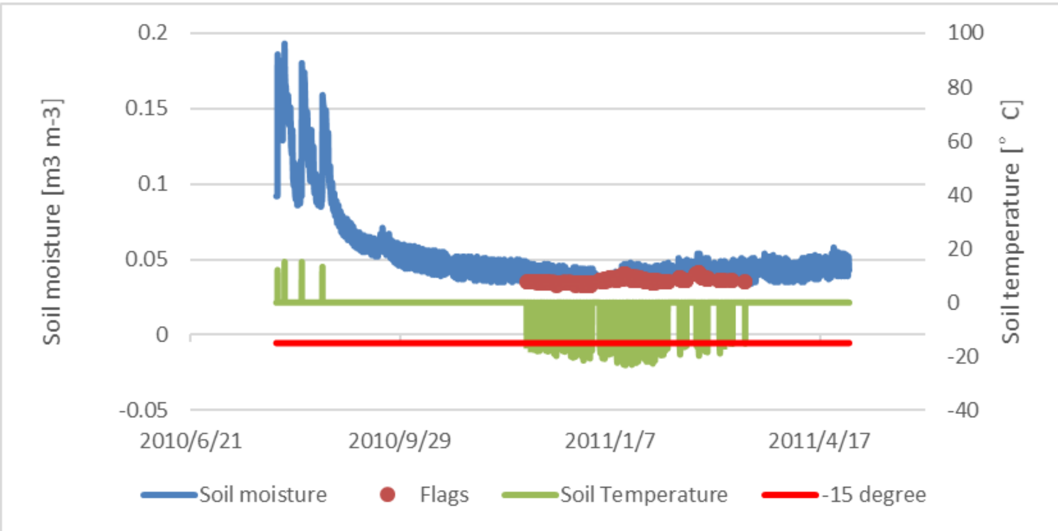
(a)



(b)



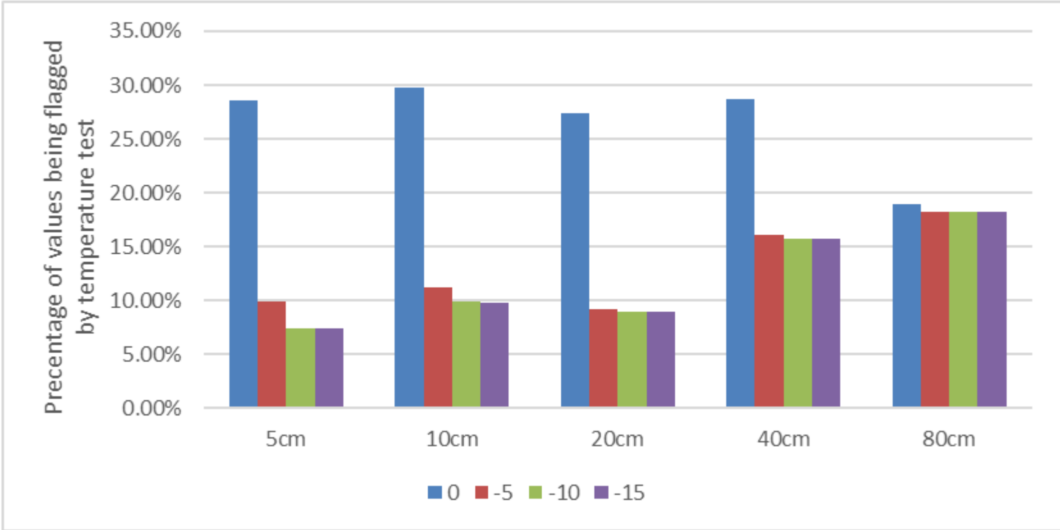
(c)



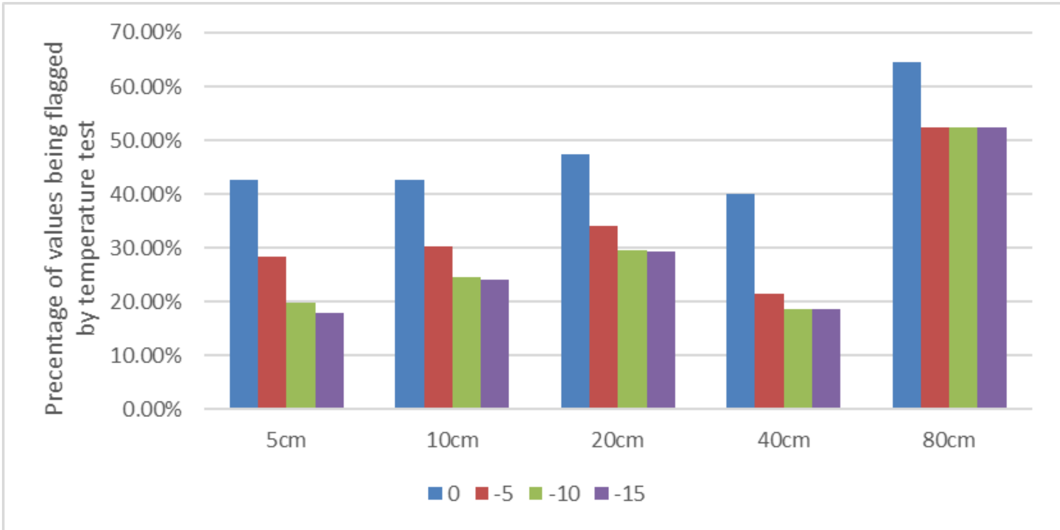
(d)

Figure 22, (a) an example of flags with temperature threshold at 0 degree, (b) an example of flags with temperature threshold at -5 degree, (c) an example of flags with temperature threshold at -10 degree, (d) an example of flags with temperature threshold at -15 degree.

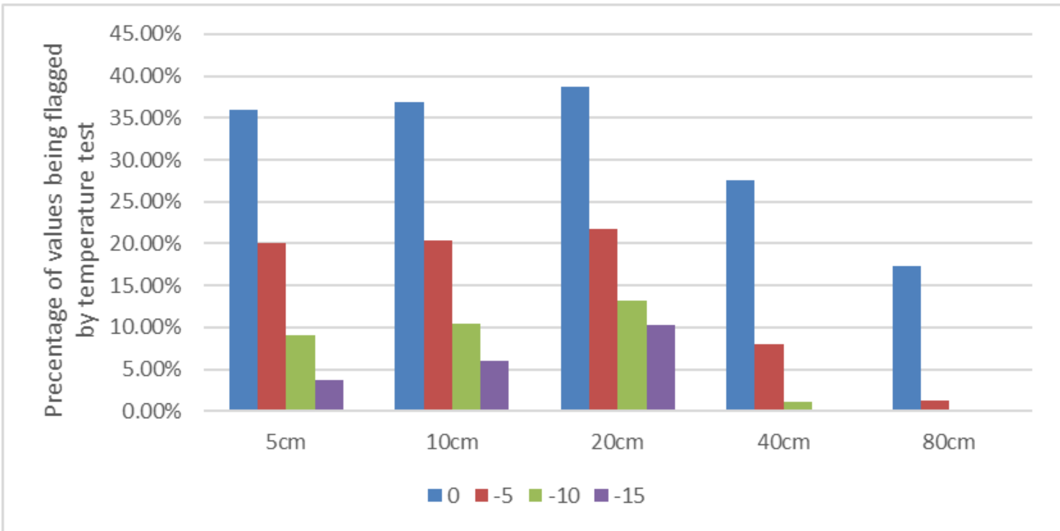
As we can see from the four examples of various temperature thresholds, the soil moisture being flagged against temperature threshold are decreased by lower the temperature standard for detection. Among these climatic zones, the flagged soil moisture by temperature threshold at 0 degree in Semiarid has the highest percentage than that of arid and that of humid showed in Figure 23. On the contrary, after lower the temperature threshold to -15 degree, the decrease of Arid is the highest at an average percentage of 30% compare with that of humid is 20% and that of semiarid is about 15%. The most obvious decrease happens after adjust the threshold from 0 degree to -5 degree, showing that most of stations from multiple climatic regions at various depths declined at least 10% except soil moisture of humid area at 80cm. The adjustment of temperature from -5 degree to -10 degree also provides clearly fall in flag fraction in both semiarid and arid regions over all soil depths.



(a)



(b)



(c)

Figure 23, Comparison of temperature checking over three climatic zones of Tibet-Obs networks, (a) Comparison of flags of Humid area, (b) Comparison of flags over Semiarid Area, (c) Comparison for Arid regions.

Figure 23a illustrated the relationship between each modification in humid area. When the temperature threshold was 0, the soil moisture of humid flagged against temperature were about 22% at layers of 5cm, 10cm, 20cm and 40cm. Whereas the soil moisture of 80cm being flagged is only 18%. From Figure 23a we can see that, after lowering the temperature threshold levels from 0 degree to -5 degree, only soil moisture of 80cm depth decreased by just 1%, while over other depths soil moisture have declined by at least 10%. A reason for this case is that the soil moisture of deep layers is much lower than that of shallow layers and the fluctuation of soil moisture in deep layers is also more stable. The ratio of decrease over all layers were declined with the decrease of temperature threshold. For example, after temperature threshold being installed as -10 degree, only soil moisture of 5cm layers and 10cm witness slightly fall at about 2% compare with the decrease of 10% by transform temperature threshold from 0 to -5 degree. The number of flags were not changed when the threshold change from -10 degree to -15 degree.

Over the result of temperature threshold in semiarid, showing in Figure 23b, flags at 80cm is different from that of soil moisture of 80cm at humid, its flags are greater than that of humid at about 60%. After the adjustment to threshold from 0 degree to -5 degree, the flagged soil moisture of shallow layers including 5cm and 10cm declined from 40% to 30% at 10%. The change for 20cm was above 15% and the variation of 40cm dropped from 40% to 20%. Compared with the variation of flags in humid at 80cm, the alteration of flags at 80cm in semiarid is 15% which is significantly larger. The alterations of temperature threshold from -5 degree to -10 degree are smaller than the change from 0 degree to -5 degree.

As showed in Figure 23c, the flagged soil moisture of arid differs from those of flags from humid and semiarid. It is obvious that soil moisture of 20cm has highest percentage being flagged, while the lowest percentage of flagged soil moisture is 80cm. The flags of shallow soil are much higher than those of deep soil. After modifying the temperature threshold from 0 to -5 degree, the flag friction decreases with depths. Moreover, after the temperature threshold adjust to -15 degree, neither soil moisture of 40cm layer nor that of 80cm layer is being flagged, only soil moisture of shallow layers are being flagged.

Table 10, A summary of flagged percentage categorized to soil organic matters, soil types and elevations for temperature threshold. Note that, the value in blod text present the highest percentage between each comparison of mean value of various layers.

TEMPERATURE		*SOIL ORGANIC MATTERS(G/100G)*				*SOIL TYPE(SAND,SILT,CLAY)*		*ELEVATION(M)*			
0 ℃		0-5	5-10	15-20	>20	SAND	SILT	<3500	3500-4000	4000-4500	>4500
	5cm	34.6%	27.0%	26.0%	20.4%	34.9%	30.2%	28.4%	29.1%	35.9%	42.7%
	10cm	35.1%	23.7%	34.7%	28.6%	33.5%	33.9%	29.0%	31.5%	36.9%	42.6%
	20cm	36.6%	27.8%	0.0%	0.0%	37.7%	32.2%	27.1%	27.8%	38.8%	47.4%
	40cm	29.2%	29.4%	0.0%	0.0%	28.5%	30.3%	28.2%	29.2%	27.5%	40.0%
	80cm	27.3%	13.5%	0.0%	0.0%	29.0%	16.9%	12.1%	26.5%	17.3%	64.4%
	mean	**32.6%**	24.3%	30.3%	24.5%	**32.7%**	28.7%	25.0%	28.8%	31.3%	**47.4%**
-5 ℃											

	5cm	17.7%	9.0%	6.2%	3.5%	18.2%	12.5%	10.3%	9.0%	20.1%	28.3%
	10cm	18.5%	2.3%	14.1%	15.9%	15.7%	17.4%	10.4%	13.1%	20.5%	30.4%
	20cm	19.7%	9.2%	0.0%	0.0%	20.8%	14.7%	9.6%	8.6%	21.8%	34.1%
	40cm	11.5%	16.2%	0.0%	0.0%	9.5%	16.1%	15.8%	15.8%	7.5%	21.5%
	80cm	16.7%	11.5%	0.0%	0.0%	16.0%	16.0%	10.8%	26.5%	1.3%	52.4%
	mean	**16.8%**	9.7%	10.2%	9.7%	**16.0%**	15.3%	11.4%	14.6%	14.2%	**33.3%**
-10 ℃											
	5cm	10.4%	5.3%	3.8%	0.3%	9.9%	8.1%	7.7%	6.9%	9.1%	19.8%
	10cm	12.2%	0.1%	13.2%	15.6%	8.5%	14.0%	8.7%	12.2%	10.5%	24.5%
	20cm	13.8%	8.7%	0.0%	0.0%	14.7%	11.0%	9.2%	8.6%	13.2%	29.5%
	40cm	7.5%	16.2%	0.0%	0.0%	5.0%	14.1%	15.8%	15.8%	1.2%	18.7%
	80cm	16.0%	11.5%	0.0%	0.0%	15.2%	16.0%	10.8%	26.5%	0.0%	52.3%
	mean	**12.0%**	8.4%	8.5%	8.0%	10.7%	**12.6%**	10.4%	14.0%	6.8%	**29.0%**
-15 ℃											
	5cm	7.4%	5.2%	3.8%	0.0%	6.6%	6.7%	7.6%	6.9%	3.8%	17.9%
	10cm	9.8%	0.1%	13.2%	15.6%	6.1%	12.7%	8.7%	12.2%	6.0%	24.0%
	20cm	12.0%	8.7%	0.0%	0.0%	13.1%	9.4%	9.2%	8.6%	10.3%	29.4%
	40cm	6.8%	16.2%	0.0%	0.0%	4.5%	13.6%	15.8%	15.8%	0.1%	18.7%
	80cm	16.0%	11.5%	0.0%	0.0%	15.2%	16.0%	10.8%	26.5%	0.0%	52.3%
	mean	**10.4%**	8.3%	8.5%	7.8%	9.1%	**11.7%**	10.4%	14.0%	4.0%	**28.5%**

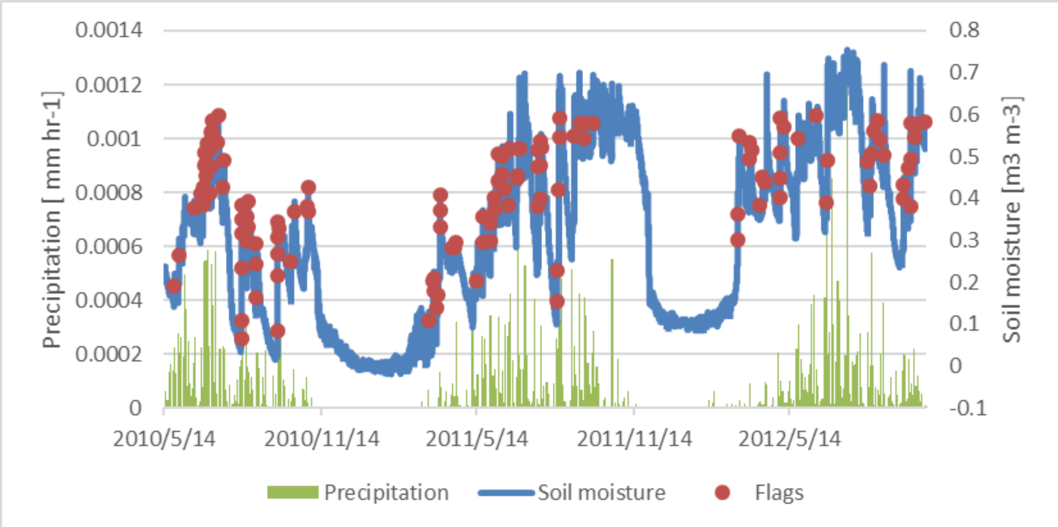
Soil temperature and soil types determine the ratio of liquid water content in frozen soil (Kurylyk et al., 2013). Table 10 shows how the flag fraction was affected by soil organic content, soil types and elevations, with different soil temperature thresholds. The red cells in table 10 present the highest percentage for certain group (e.g. soil organic content, soil types or elevation). By comparing the difference of the mean value from each group we can see that flag fraction was affected by elevations significantly, while only lightly influenced by soil types. Among them, the flag fraction of soil organic matter group shows similar magnitude as the soil type group. Soil moisture of higher elevation are more vulnerable to be influenced by temperature threshold. Moreover, the highest change between various layers is soil of 5cm means that soil moisture of shallow layers in high elevation are easily impacted by temperature.

5.4. Precipitation

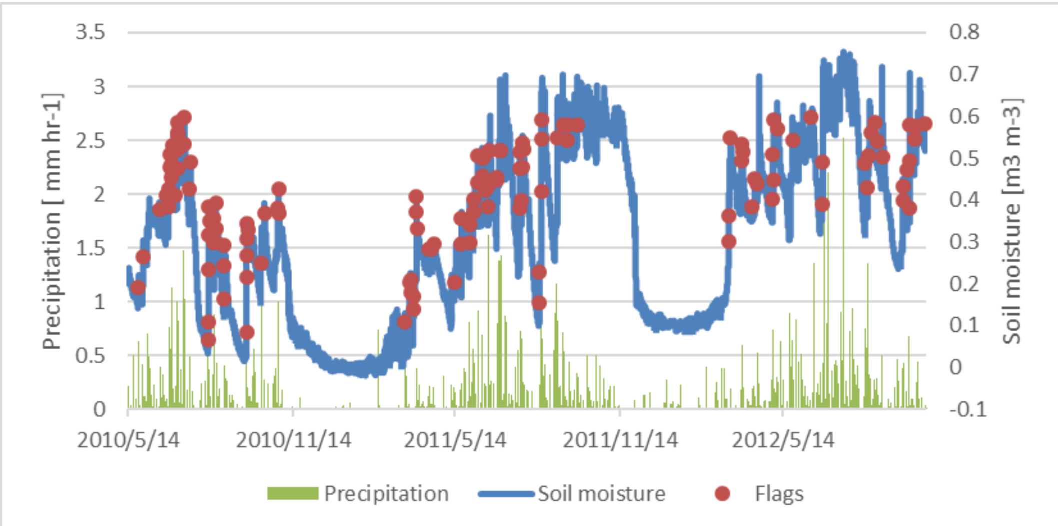
Two precipitation data sets were proposed to use in the quality control program as geophysical checking for precipitation, GLDAS and CMFD respectively. Precipitation significantly affects the rises in time series of soil moisture. However, to keep the consistency between precipitation and temperature, two correlative temperature datasets are also employed, air temperature of GLDAS and CMFD. Thus, four group were created by GPGT (GLDAS precipitation with air temperature), GPCT (GLDAS precipitation with CMFD air temperature), CPCT (CMFD precipitation with air temperature) and CPGT (CMFD precipitation with GLDAS precipitation). Figure 24 below demonstrate examples of those four groups by showing flags in soil moisture from NST_02 station (Maqu networks) at 10 cm depth.

Three bar chart showed in Figure 25 statistically illustrate alteration between these groups over multiple climatic zones at various depths. Overall, soil moisture flagged by the this consistency check has the greatest percentage in semiarid over all layers with an average of flag fraction at least 40%, while the humid area have lowest flagging percentage with just 16% averagely for all layers. The arid area have a percentage of 25% being flagged. Except for the soil moisture of humid areas, there is no obvious change between each group over various layers in semiarid or arid regions.

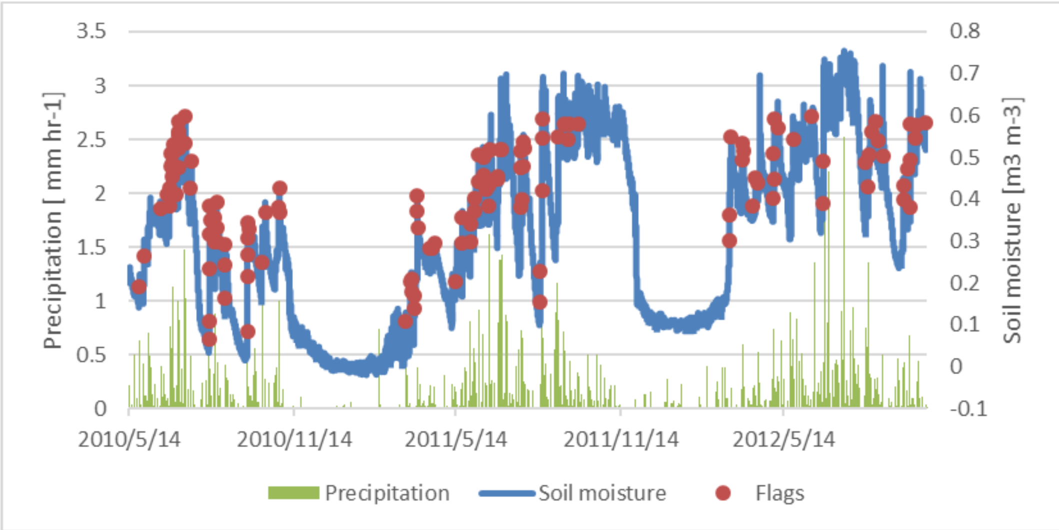
For soil moisture in humid regions, the combination of GLDAS precipitation with CMFD air temperature raise the most significant flag fraction, when compared with other groups over layers of 5cm, 10cm, 20cm and 40cm. Nevertheless, it is envisaged that the inconsistency of GLDAS/CMFD precipitation with the local situation will definitely lead to the fault flagging of soil moisture, which can be seen clearly for the Maqu network, which is a humid region.



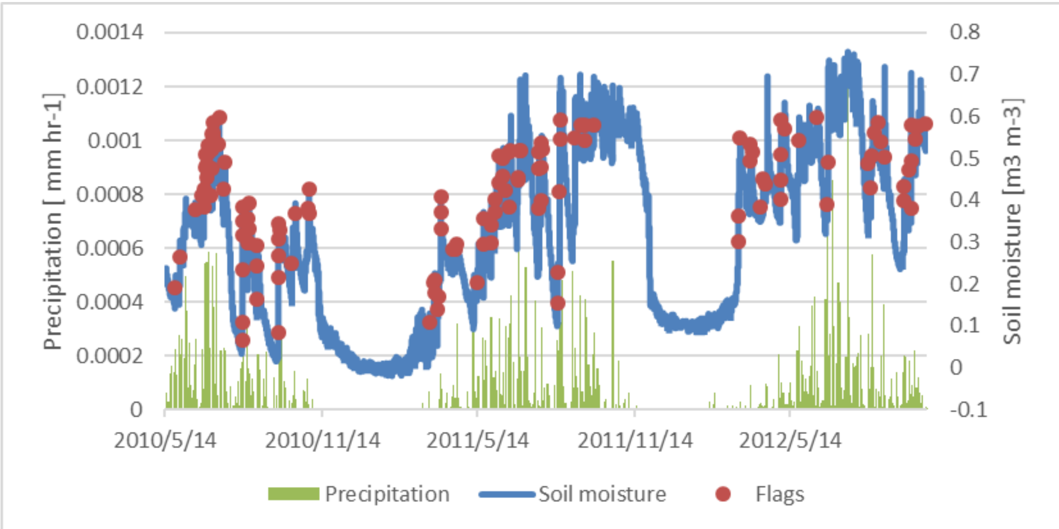
(a)



(b)



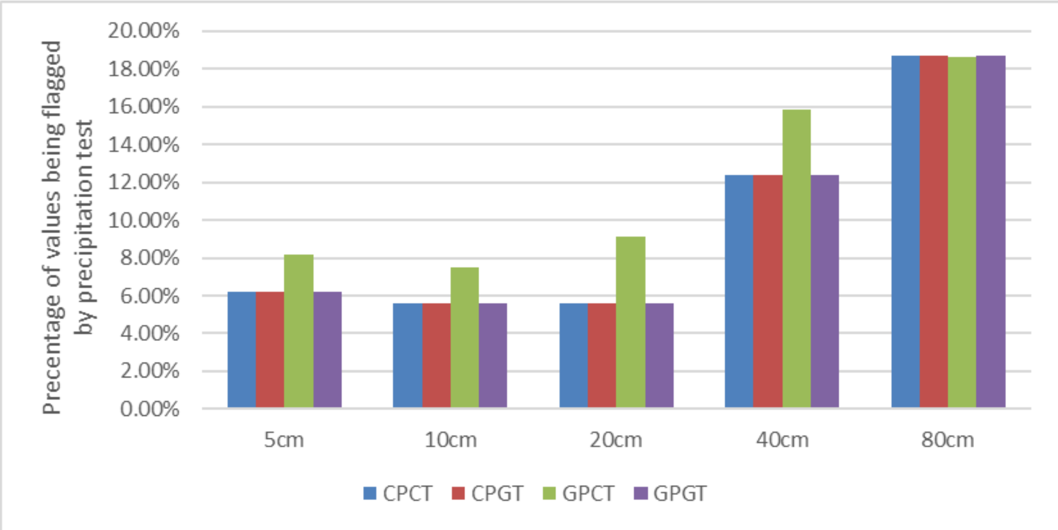
(c)



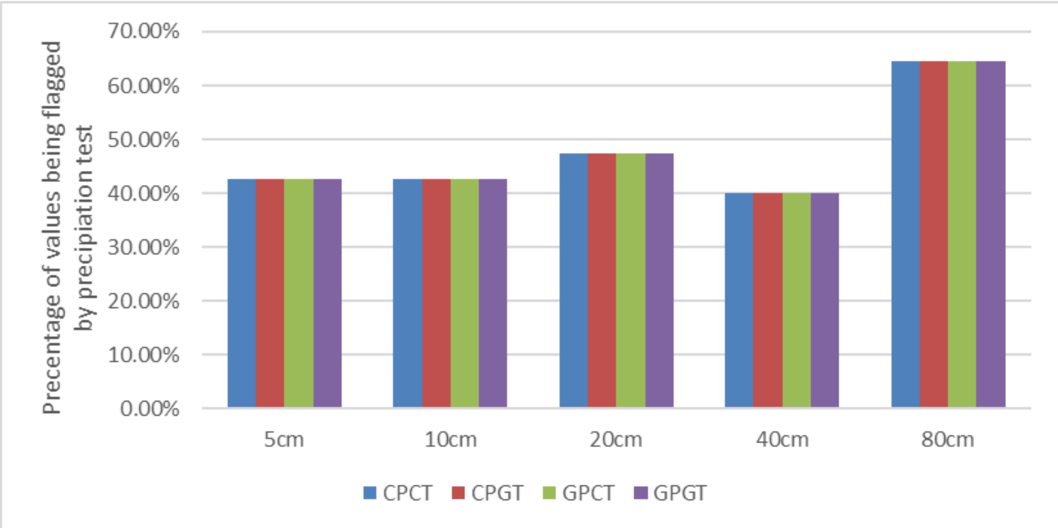
(d)

Figure 24, (a) an example of flags with precipitation checking by GLDAS precipitation and CMFD air temperature, (b) an example of flags with precipitation checking by CMFD precipitation and CMFD air temperature, (c) an example of flags with precipitation checking by CMFD precipitation and CMFD air temperature, (d) an example of flags with precipitation checking by GLDAS precipitation and GLDAS air temperature

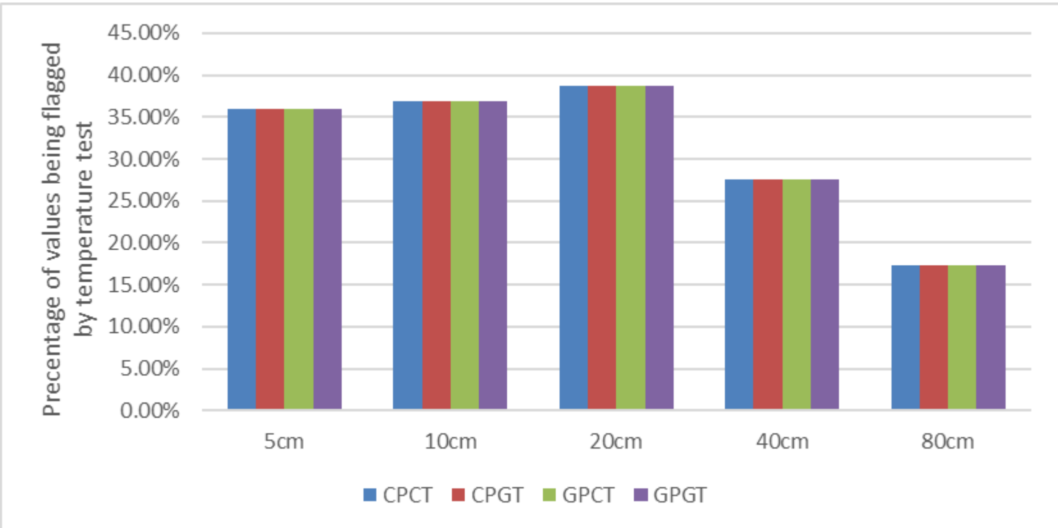
However, in semiarid and arid areas, there was no obvious difference between these four groups. In semiarid area, soil moisture of 20cm and 80cm have more flags than other layers. For arid area, it shows a large uncertainty in shallow soil, including 5cm, 10cm and 20cm. But flags of deep layers are smaller than those of shallow soil.



(a)



(b)



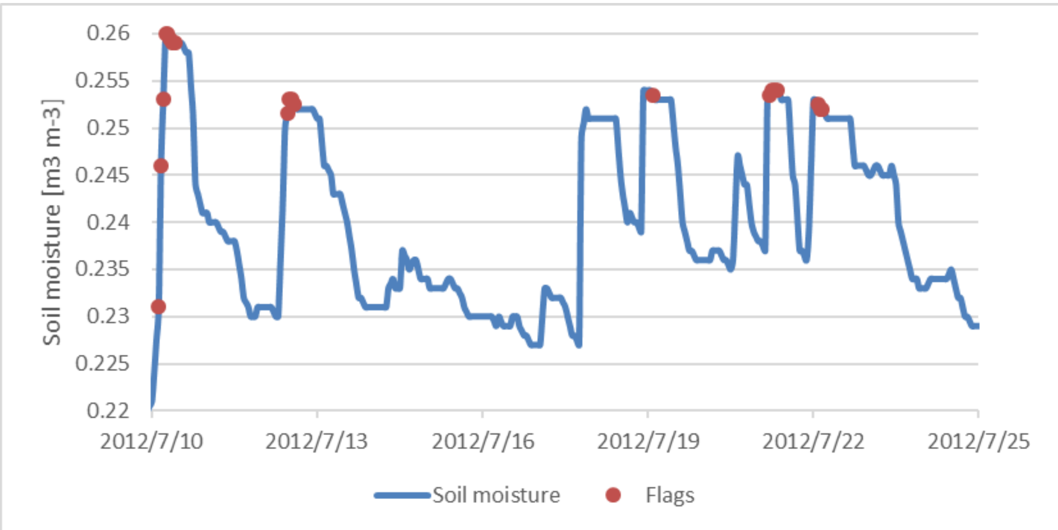
(c)

Figure 25, Comparison of precipitation over three climatic zones of Tibet-Obs networks, (a) Comparison of flags of Humid area, (b) Comparison of flags over Semiarid Area, (c) Comparison for Arid regions.

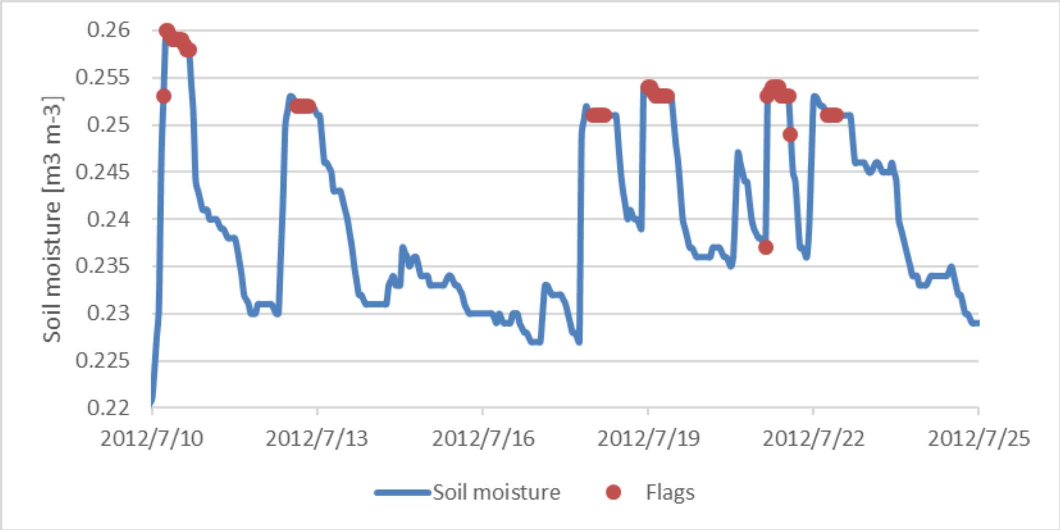
Overall, soil moistures are flagged where there is no precipitation or small amount of rainfall but the soil moisture still seen a rise. The above results show that it seems inappropriately to use either GLDAS/CMFD to identify spurious values. Most flagged soil moisture were not having quality issue. The main reason to be responsible for this case is that precipitation products employed in this research is not appropriate to be applied for the whole Tibet-Obs. The soil moisture record in this research has a temporal scale of 1hour, however, the temporal resolution of these precipitation is only 3 hour. Moreover, disaggregation of precipitation to achieve the same temporal resolution brought unpredictable inconsistency between soil moisture and precipitation. We recommend to use a finer precipitation data to enhance the reliability of this quality control approach.

5.5. Constant value detection

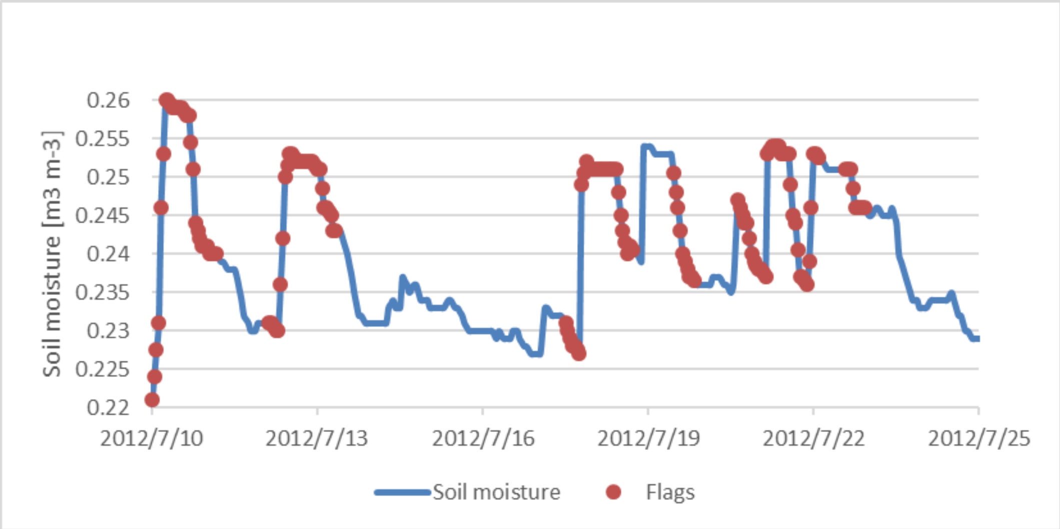
Before the code being modification, constant value detection played a critical part within the quality control by showing that almost 90% soil moisture were encountered with this quality issue. However, this situation was incorrect. After the modification, the result dramatically improved by decreasing more than 80% of the flag fraction. The proposed time window for doing this constant value check is 12 hours. This was determined by checking the sensitivity of the length of time window on the flag fraction (i.e.: 12 hours, 24 hours, 48 hours, 72 hours and 96 hours as showed below in Figure 26). The examples in Figure 26 show that the soil moisture was flagged by the issue of having constant value. The station here is North station in Naqu network from a semiarid area at 40cm depth.



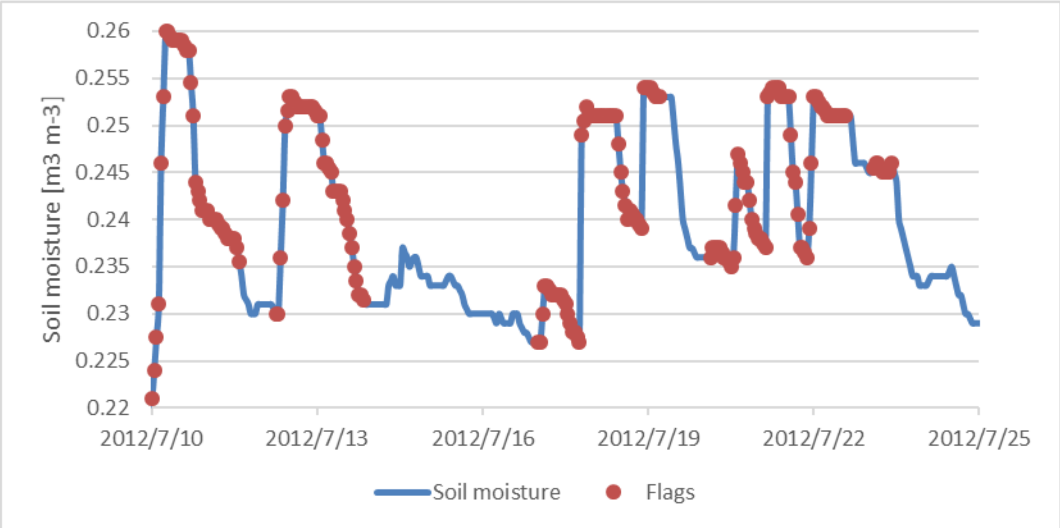
(a)


Soil moisture [m3 m-3]
0.26
0.255
0.25
0.245
0.24
0.235
0.23
0.225
0.22
2012/7/10
2012/7/13
2012/7/16
2012/7/19
2012/7/22
2012/7/25
Soil moisture
Flags

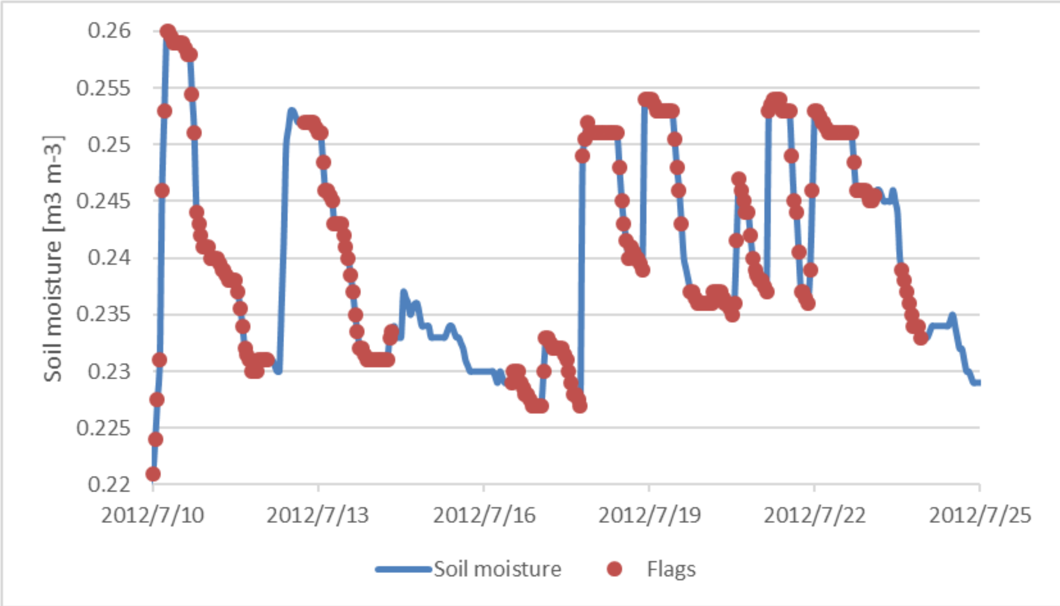
(b)


Soil moisture [m3 m-3]
0.26
0.255
0.25
0.245
0.24
0.235
0.23
0.225
0.22
2012/7/10
2012/7/13
2012/7/16
2012/7/19
2012/7/22
2012/7/25
Soil moisture
Flags

(c)


Soil moisture [m3 m-3]
0.26
0.255
0.25
0.245
0.24
0.235
0.23
0.225
0.22
2012/7/10
2012/7/13
2012/7/16
2012/7/19
2012/7/22
2012/7/25
Soil moisture
Flags

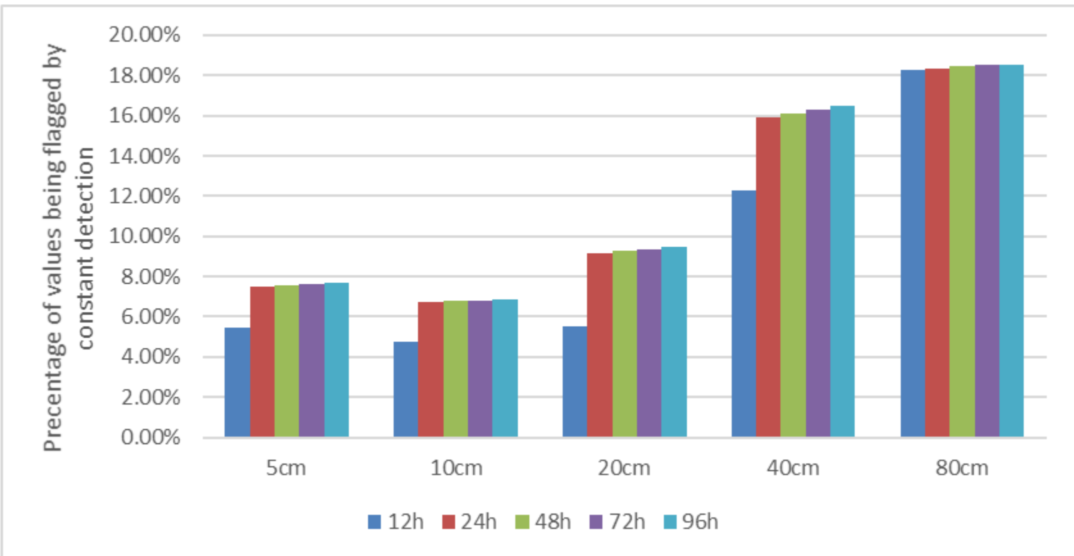
(d)



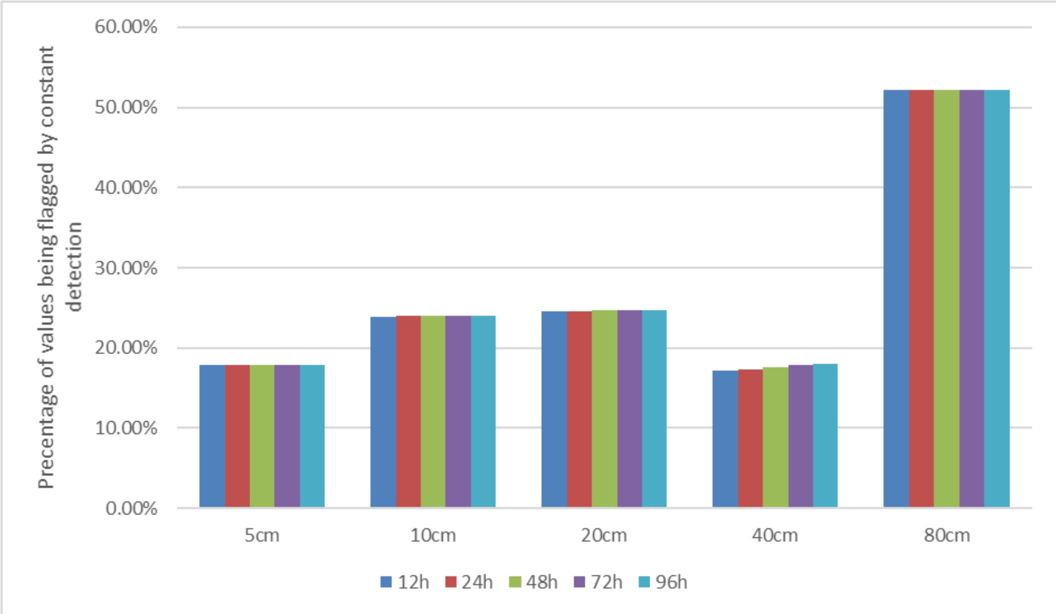
(e)

Figure 26, (a) an example of flags with constant detection over 12 hours, (b) an example of flags with constant detection over 24 hours, (c) an example of flags with constant detection over 48 hours, (d) an example of flags with constant detection over 72 hours (e) an example of flags with constant detection over 72 hours.

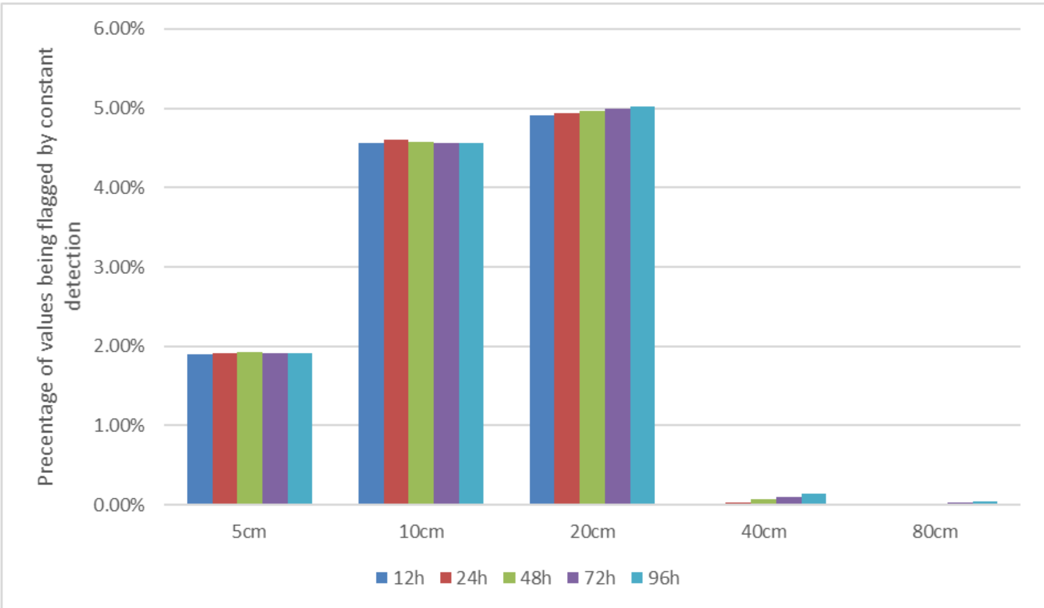
Three comparisons of diverse climatic zones were presented below in Figure 27. The soil moisture of semiarid possess the maximal percentage of constant flagged value. Compare to semiarid, flags of arid area hold the minimal percentage of constant values over multiple layers. Flagged soil moisture in humid area is notably greater in deep layers at 40cm and 80cm.



(a)



(b)



(c)

Figure 27, Comparison of constant value over three climatic zones of Tibet-Obs networks, (a) Comparison of flags of Humid area, (b) Comparison of flags over Semiarid Area, (c) Comparison for Arid regions.

The comparison in Figure 27 clearly illustrate the relationship of different constant detection standard at three climatic zones over entire soil profile. In the Figure 27a, soil moisture flagged as constant value at 12 hours show the best result compared with flags detected following the time window of 24hours, 48 hours, 72 hours and 96 hours.

5.6. Optimum matching and its validation

Based on the sensitive analyse of variables for spectrum-based methods, several variables are adjusted to facilitate calculating the best flagging results for Tibet-Obs networks. Threshold as a critical part flagged large portion of soil moisture by initial value at 0.5, was adjusted to 0.7 for avoiding overflagging. The temperature condition employed in the method was modified from 0 degree to -15 degree. Precipitation from GLDAS was introduced for flagging, with which GLDAS air temperature also conducted to keep the consistency with precipitation data. The time window to conduct constant value checking was limited to 12 hours. The matrix below evaluated the performance of flagging result for the best matching(Hubbard et al., 2005). The matrix validating the result of optimum matching by comparing the flags with the visual checked flags. All soil moisture, including values identified as errors or right, will be tested to ensure the correctness of detection. The erroneous measurement verifies the flags are errors or are incorrectly identified. The correct measurement was used to confirm whether there is soil moisture have not been flagged by quality control program. This validation applied for 6 variables, threshold, temperature, precipitation, spikes, breaks and constant values. The validation shows the quality control successfully detected erroneous values except spike detection and breaks.

Flagging result	*Erroneous measurement*		*Correct measurement*		*Flag Observation*
	Erroneous Detection ok	Correct Fault detection	Erroneous Fault detection	Correct Detection ok	
Threshold	100%	0%	0%	100%	16.6%
Temperature testing	98.2%	1.8%	0.1%	99.9%	15.09%
Precipitation-based flagging	87%	13%	3.6%	96.4%	0.72%
Spikes	50.3%	49.7%	14.7%	85.3%	0.01%
Break	51.9%	48.1%	34%	76%	0.67%
Constant value	92.2%	7.8%	15.5%	84.5%	0.03%

Table 11, Summary of the flagging performance based on optimum matching

6. CONCLUSION

Spectrum-based quality control is applicable to detect spurious soil moisture over all station in Tibet-Obs networks at various depth. The original research conducted by Dorigo et al., 2013 was aim to developed an automated quality control at universal scale. However, the initial evaluation exclusively applied to soil moisture at shallow layers which is less than 10cm. But, soil moisture of Tibet-Obs are at local scale with 5 layers extending to 80cm. For different climatic zones, the appearance of time series of soil moisture is totally different over various depth.

The general threshold for entire stations does not work appropriately. 0.7 m3 m-3 is proposed to implement for checking over humid area at numerous layers except 20cm and 40cm. The real challenge to flag spurious records in humid area is to determine the threshold. For some stations in the humid climatic zones, like NST_04, the soil moisture value can reach 0.8 m3 m-3. On the other hand, according to the characteristics of the soil moisture of semiarid and arid regions, 0.4 m3m-3 provide a more rational standard for arid area. Moreover, the difference of established thresholds for semiarid between shallow layers and deep layers is that 0.5 m3 m-3 as conventional threshold is applicable to all layers. Flags were categorized into groups based on soil types, soil organic matters and elevation, and showed that these factors should be considered for identifying quality issues. Porosity as a useful variable corresponded with soil moisture, but only employed in the methods for precipitation checking, and used to define the upper boundary. Establishing a porosity based upper boundary thresholds improve the application of current QC code.

Due to the high elevation of Tibet plateau, temperature is severely affected by the elevation, resulting in much lower temperature compare with other regions. The performance of quality control is criterial influenced by temperature threshold. In the current QC code, the seasonal condition has not been considered. Seasonal typical temperature threshold should be taken into consideration for the QC of soil moisture data.

The main problem for precipitation detection maybe the inconsistency between employed precipitation and air temperature. Moreover, in CMFD precipitation data and its air temperature data, some stations of Tibet-Obs could not acquire a value due to the location of these station are out of boundary. However, both of GLDAS precipitation and CMFD precipitation employed at a temporal resolution of 3 hours. The disaggregation of precipitation data is used by convert 3 hourly data to hourly resolution giving intrinsic instability to proper identify spurious data. But if a local precipitation with finer temporal resolution can be provide, it is able to more accurately detect erroneous values.

For constant value detection, setting time interval with 12 hours seems revealed more acceptable result than other time intervals. The possible reason responsible for this issue is the inappropriately defined positive constant value which should be improved by clearly defining the flags between the start and the end. The original purpose of extending the time interval from 12 hours to 96 hours is to minimize the overflagging. Unfortunately, enhancing the time interval form 12 hours to 96 hours lead to more incorrect flags. The appearance of constant flags of different climatic zones are diverse, showing that positive constant normally appeared in humid area while negative constant appeared frequently in semiarid and arid by sensor failure. The other difference between positive constant and negative constant over entire Tibet-Obs network is that positive constant often lasts short period while negative constant normally appeared long time negative values.

Overall, the spectrum-based method provides acceptable solution for quality control of in situ soil moisture over Tibet-Obs networks. By appropriately defining several restrict conditions, the incorrected flags could be minimized. Moreover, seasonal difference for quality control is proposed as a future research area over Tibet-Obs network.

LIST OF REFERENCES

Dorigo, W. A., Xaver, A., Vreugdenhil, M., Gruber, A., Hegyiová, A., Sanchis-Dufau, A. D., … Drusch, M. (2013). Global Automated Quality Control of In Situ Soil Moisture Data from the International Soil Moisture Network. *Vadose Zone Journal*, *12*(3), 0. https://doi.org/10.2136/vzj2012.0097

Eischeid, J. K., Bruce Baker, C., Karl, T. R., Diaz, H. F., Eischeid, J. K., Baker, C. B., … Diaz, H. F. (1995). The Quality Control of Long-Term Climatological Data Using Objective Data Analysis. *Journal of Applied Meteorology*, *34*(12), 2787–2795. https://doi.org/10.1175/1520-0450(1995)034<2787:TQCOLT>2.0.CO;2

Geiger, M., Diabaté, L., Ménard, L., & Wald, L. (2002). A web service for controlling the quality of measurements of global solar irradiation. *Solar Energy*, *73*(6), 475–480. https://doi.org/10.1016/S0038-092X(02)00121-4

Gruber, A., Dorigo, W. A., Zwieback, S., Xaver, A., & Wagner, W. (2013). Characterizing Coarse-Scale Representativeness of in situ Soil Moisture Measurements from the International Soil Moisture Network. *Vadose Zone Journal*, *12*(2), 0. https://doi.org/10.2136/vzj2012.0170

Hubbard, K. G., Goddard, S., Sorensen, W. D., Wells, N., & Osugi, T. T. (2005). Performance of Quality Assurance Procedures for an Applied Climate Information System. *Journal of Atmospheric and Oceanic Technology*, *22*(1), 105–112. https://doi.org/10.1175/JTECH-1657.1

Inan, Z. H., Kuntalp, M., Vreugdenhil, M., Gruber, A., Hegyiová, A., Sanchis-Dufau, A. D., … Decharme, B. (2007). A study on fuzzy C-means clustering-based systems in automatic spike detection. *Computers in Biology and Medicine*, *37*(8), 1160–6. https://doi.org/10.1016/j.compbiomed.2006.10.010

Jie, H., & Kun, Y. (2011). China Meteorological Forcing Dataset. *Cold and Arid Regions Science Data Center at Lanzhou*. https://doi.org/doi:10.3972/westdc.002.2014.db

Journée, M., & Bertrand, C. (2011). Quality control of solar radiation data within the RMIB solar measurements network. *Solar Energy*, *85*(1), 72–86. https://doi.org/10.1016/j.solener.2010.10.021

Kurylyk, B. L., & Watanabe, K. (2013). The mathematical representation of freezing and thawing processes in variably-saturated, non-deformable soils. *Advances in Water Resources*, *60*, 160–177. https://doi.org/10.1016/j.advwatres.2013.07.016

Kutzbach, J. E., Prell, W. L., & Ruddiman, W. F. (1993). Sensitivity of Eurasian Climate to Surface Uplift of the Tibetan Plateau. *The Journal of Geology*, *101*(2), 177–190. Retrieved from http://www.jstor.org/stable/30081146

Malyshev, A. B., & Sudakova, I. B. (1994). Algorithm for the spikes detection in the cosmic ray intensity measurements from the data of satellite Meteor. *GEOMAGNETISM AND AERONOMY C/C OF GEOMAGNETIZM I AERONOMIIA*, *34*, 578–580.

May, W., Meier, A., Rummukainen, M., Berg, A., Chéruy, F., & Hagemann, S. (2015). Contributions of soil moisture interactions to climate change in the tropics in the GLACE–CMIP5 experiment. *Climate Dynamics*, *45*(11), 3275–3297. https://doi.org/10.1007/s00382-015-2538-9

MEEK, D., & HATFIELD, J. (1994). Data quality checking for single station meteorological databases☆. *Agricultural and Forest Meteorology*, *69*(1–2), 85–109. https://doi.org/10.1016/0168-1923(94)90083-3

Ormsbee, L. E. (1989). Rainfall Disaggregation Model for Continuous Hydrologic Modeling. *Journal of Hydraulic Engineering*, *115*(4), 507–525. https://doi.org/10.1061/(ASCE)0733-9429(1989)115:4(507)

Rajib, M. A., Merwade, V., & Yu, Z. (2016). Multi-objective calibration of a hydrologic model using spatially distributed remotely sensed/in-situ soil moisture. *Journal of Hydrology*, *536*, 192–207. https://doi.org/10.1016/j.jhydrol.2016.02.037

Schmugge, T. J. (1983). Remote Sensing of Soil Moisture: Recent Advances. *IEEE Transactions on Geoscience and Remote Sensing*, *GE-21*(3), 336–344. https://doi.org/10.1109/TGRS.1983.350563

Shafer, M. A., Fiebrich, C. A., Arndt, D. S., Fredrickson, S. E., Hughes, T. W., Shafer, M. A., … Hughes, T. W. (2000). Quality Assurance Procedures in the Oklahoma Mesonetwork. *Journal of Atmospheric and Oceanic Technology*, *17*(4), 474–494. https://doi.org/10.1175/1520-0426(2000)017<0474:QAPITO>2.0.CO;2

Su, Z., de Rosnay, P., Wen, J., Wang, L., & Zeng, Y. (2013). Evaluation of ECMWF's soil moisture analyses using observations on the Tibetan Plateau. *Journal of Geophysical Research: Atmospheres*, *118*(11), 5304–5318. https://doi.org/10.1002/jgrd.50468

Su, Z., Wen, J., Dente, L., van der Velde, R., Wang, L., Ma, Y., … Hu, Z. (2011). The Tibetan Plateau observatory of plateau scale soil moisture and soil temperature (Tibet-Obs) for quantifying uncertainties in coarse resolution satellite and model products. *Hydrology and Earth System Sciences*,

15(7), 2303–2316. https://doi.org/10.5194/hess-15-2303-2011
WMO. (2004). Guidelines on Quality Control Procedures for Data from Automatic Weather Stations. Retrieved from https://www.wmo.int/pages/prog/www/IMOP/meetings/Surface/ET-STMT1_Geneva2004/Doc6.1(2).pdf quality control guideline automated
You, J., Hubbard, K. G., Mahmood, R., Sridhar, V., & Todey, D. (2010). Quality Control of Soil Water Data in Applied Climate Information System—Case Study in Nebraska. *Journal of Hydrologic Engineering*, *15*(3), 200–209. https://doi.org/10.1061/(ASCE)HE.1943-5584.0000174
You, J., Hubbard, K. G., You, J., & Hubbard, K. G. (2006). Quality Control of Weather Data during Extreme Events. *Journal of Atmospheric and Oceanic Technology*, *23*(2), 184–197. https://doi.org/10.1175/JTECH1851.1
Zeng, Y., Su, Z., van der Velde, R., Wang, L., Xu, K., Wang, X., & Wen, J. (2016). Blending Satellite Observed, Model Simulated, and in Situ Measured Soil Moisture over Tibetan Plateau. *Remote Sensing*, *8*(3), 268. https://doi.org/10.3390/rs8030268

7. APPENDIX

A summary for collating data from multiple stations

Station Name	Lat	Lon		Processing
NGARI				
SQ01	32.48	80.07		Interpolation
SQ02	32.50	80.02		Interpolation
SQ03	32.50	79.97		Interpolation
SQ04	32.50	79.95		Interpolation
SQ05	32.50	79.92		None
SQ06	32.50	79.87		Interpolation
SQ07	32.52	79.83		Interpolation
SQ08	32.55	79.83		Division
SQ09	32.45	80.05		Interpolation
SQ10	32.42	80.00		None
SQ11	32.45	79.97		Interpolation
SQ12	32.45	79.93		Interpolation
SQ13	32.43	79.90		None
SQ14	32.45	80.17		Deletion
SQ15	32.43	80.18		
SQ16	32.43	80.07		None
AL01	33.43	79.73		Interpolation
AL02	33.45	79.62		Division
AL03	33.45	79.62		Interpolation
NAQU				
Naqu Station	31.37	91.88		Deletion
East Station	31.37	91.92		None
North Station	31.37	91.87		None
South Station	31.32	91.87		None
West Station	31.33	91.82		None
MAQU				
CST_01	33.88	102.13		Interpolation

CST_02	33.67	102.13		Interpolation
CST_03	33.87	101.97		Division
CST_04	33.77	101.72		Deletion
CST_05	33.67	101.88		Division
NST_01	33.88	102.13		Interpolation
NST_02	33.88	102.13		Deletion
NST_03	33.77	102.13		Deletion
NST_04	33.62	102.05		Interpolation
NST_05	33.63	102.05		Deletion
NST_06	34.00	102.27		Deletion
NST_07	33.98	102.35		Deletion
NST_08	33.97	102.60		Deletion
NST_09	33.90	102.55		Deletion
NST_10	33.85	102.57		Deletion
NST_11	33.68	102.47		Deletion
NST_12	33.62	102.47		Deletion
NST_13	34.02	101.93		Deletion
NST_14	33.92	102.12		None
NST_15	33.85	101.88		Deletion

A summary of the visual inspection of flags in soil moisture of three subnetworks over multiple layers, 1 represent there exist this type of flag.

Maqu	RA	PREC	TEMP	SPIKE	BREAK	CONSTANT
CST01_10cm	1		1	1		
CST01_20cm	1		1	1		
CST01_40cm	1		1	1	1	
CST01_5cm	1		1	1		
CST01_80cm	1	1	1	1	1	1
CST02_10cm	1	1	1	1		
CST02_20cm	1		1	1		
CST02_40cm	1	1	1	1		1
CST02_5cm	1	1	1	1	1	
CST02_80cm	1	1	1	1	1	1
CST03_10cm	1	1	1	1	1	1
CST03_20cm			1			
CST03_40cm		1	1			
CST03_5cm	1	1	1	1		
CST03_80cm	1		1	1	1	1
CST04_10cm	1		1	1		

CST04_20cm			1	1		
CST04_40cm	1		1	1		
CST04_5cm	1		1	1	1	1
CST04_80cm	1		1	1	1	1
CST05_10cm	1	1	1	1	1	1
CST05_20cm	1		1	1	1	1
CST05_40cm	1		1	1	1	1
CST05_5cm	1	1	1	1	1	1
CST05_80cm	1		1	1	1	1
NST01_10cm	1		1		1	
NST01_20cm	1		1		1	1
NST01_40cm	1		1	1	1	1
NST01_5cm	1	1	1	1	1	1
NST01_80cm	1	1	1	1	1	1
NST02_10cm	1	1	1	1		
NST02_5cm	1	1	1	1		
NST03_10cm	1	1	1	1		
NST03_5cm	1	1	1	1		
NST04_10cm	1	1	1	1	1	1
NST04_5cm	1	1	1	1	1	1
NST05_10cm	1		1	1		
NST05_20cm	1	1	1			
NST05_40cm	1		1	1	1	1
NST05_5cm	1		1	1		
NST06_10cm		1	1	1		
NST06_20cm			1	1		
NST06_40cm			1	1	1	
NST06_5cm		1	1	1		
NST07_10cm	1		1	1		
NST07_5cm	1		1	1		
NST08_10cm			1	1	1	1
NST08_5cm			1	1		
NST09_10cm		1	1	1		
NST09_5cm		1	1	1		
NST10_10cm	1	1	1	1		
NST10_20cm	1		1	1	1	1
NST10_40cm	1		1	1		1
NST10_5cm	1		1	1	1	1
NST11_10cm	1	1	1	1		
NST11_5cm	1	1	1	1		
NST12_10cm		1	1			
NST12_20cm			1			
NST12_40cm	1		1		1	1
NST12_5cm		1	1	1		
NST12_80cm		1	1			

NST13_10cm		1	1	1		
NST13_20cm			1	1		
NST13_40cm	1	1	1	1	1	1
NST13_5cm	1	1	1	1		
NST14_10cm	1	1	1	1		1
NST14_5cm	1	1	1	1	1	1
NST15_10cm	1		1	1	1	
NST15_5cm	1	1	1	1	1	1
Ngari	RA	PREC	TEMP	SPIKE	BREAK	CONSTANT
AL_01_10cm			1		1	
AL_01_20cm		1	1		1	
AL_01_40cm			1		1	
AL_01_5cm		1	1	1		
AL_01_80cm		1	1		1	
AL_02_10cm			1			
AL_02_20cm			1			
AL_02_40cm			1			
AL_02_5cm	1		1		1	1
AL_02_80cm			1			
AL_03_10cm			1			
AL_03_20cm			1			
AL_03_40cm			1			
AL_03_5cm			1	1		
AL_03_80cm			1			
Sq_05_20cm		1	1	1		
Sq_05_40cm			1			
Sq_05_5cm		1	1	1		
Sq_06_10cm	1		1	1	1	1
Sq_06_20cm			1			
Sq_06_40cm			1			
Sq_06_5cm		1	1	1		
Sq_06_80cm			1			
Sq_07_10cm		1	1			
Sq_07_20cm		1	1	1		
Sq_07_40cm		1	1	1		1
Sq_07_5cm		1	1			
Sq_08_10cm		1	1	1		
Sq_08_20cm		1	1	1		
Sq_08_40cm			1			
Sq_08_5cm			1			
Sq_08_80cm			1			
Sq_09_10cm		1	1			
Sq_09_20cm			1			
Sq_09_40cm			1			
Sq_09_5cm		1	1			

Sq_10_10cm			1			
Sq_10_20cm			1			
Sq_10_40cm		1	1			
Sq_10_5cm		1	1	1		
Sq_10_80cm		1	1			
Sq_11_10cm	1	1	1	1	1	1
Sq_11_20cm			1			
Sq_11_40cm		1	1			
Sq_11_5cm		1	1	1		
Sq_11_80cm			1			
Sq_12_10cm		1	1	1		
Sq_12_20cm	1		1		1	1
Sq_12_40cm		1	1			
Sq_12_5cm		1	1	1		
Sq_12_80cm		1	1		1	
Sq_13_10cm			1			
Sq_13_20cm	1		1	1	1	1
Sq_13_40cm			1			
Sq_13_5cm		1	1			
Sq_13_80cm			1			
Sq_14_10cm	1		1		1	1
Sq_14_20cm	1		1	1	1	1
Sq_14_40cm			1			
Sq_14_5cm			1			
Sq_16_10cm			1			
Sq_16_20cm			1	1		
Sq_16_40cm			1			
Sq_16_5cm	1		1	1	1	1
Sq_16_80cm			1			
Naqu	RA	PREC	TEMP	SPIKE	BREAK	CONSTANT
CENT_10cm	1		1	1		
CENT_20cm			1	1		1
CENT_40cm			1	1		
CENT_5cm	1	1	1	1	1	1
CENT_80cm			1	1		
EAST_10cm	1		1	1		
EAST_20cm	1	1	1	1		
EAST_40cm	1		1		1	1
EAST_5cm	1		1	1	1	1
EAST_80cm	1		1	1	1	1
NOTH_20cm	1	1	1	1	1	1
NOTH_40cm	1		1	1		
NOTH_5cm			1	1	1	1
NOTH_80cm	1		1			
SOTH_10cm	1	1	1	1		

SOTH_20cm			1	1		
SOTH_40cm			1			
SOTH_5cm	1		1	1		
SOTH_80cm			1			
WEST_10cm	1		1	1		
WEST_20cm	1	1	1	1		
WEST_40cm	1		1	1		
WEST_5cm	1	1	1	1	1	
WEST_80cm	1		1	1		