MAUSAM, **66**, 1 (January 2015), 93-106

551.508.824

Development of operational near real-time network monitoring and quality control system for implementation at AWS data receiving earth station

M. R. RANALKAR, A. ANJAN, R. P. MISHRA, R. R. MALI and S. KRISHNAIAH

India Meteorological Department, Pune – 411 005, India

(*Received 7 November 2013, Modified 7 March 2014*)

e mail : mr.ranalkar@imd.gov.in; mranalkar@yahoo.com

सार – इस शोध पत्र में भारतीय स्वचालित मौसम केंद्र (AWS) संजाल के स्वचालित निकट वास्तविक समय निगरानी और भू केंद्र से प्राप्त केंद्रीय आँकड़ों की गुणवत्ता नियंत्रण (QC) की कार्य पद्धतियों को विकसित किया गया और किए गए परीक्षण के परिणामों को प्रस्तुत किया गया है। इन कार्य पद्धतियों के अभाव में AWS के आँकड़े प्रचालनात्मक आवश्यकताओं को तो पूरा कर सकते हैं किंतु यह संयोगवश हो सकता है। AWS से प्राप्त मौसम विज्ञान प्राचलों के परिवर्तनशील मानों के व्यापक प्रभाव के कारण उपयोगकर्ताओं का इस कार्य पद्धति से विश्वास उठ सकता है। अत: संजाल प्रबंधकों के लिए यह उचित होगा कि वे उपयोगकर्ताओं को ज्ञात और निश्चित गुणवत्तापूर्ण आँकड़े उपलब्ध कराए।

संजाल निगरानी और लाभप्रद पद्धति की दैनिक रिपोर्टें स्वत: ही तैयार होती हैं और विभिन्न रख-रखाव केंद्रों तथा उपयोगकर्ताओं की निर्धारित ई-मेल आई डी पर स्वत: ही भेजी जाती हैं। QC पद्धति, मूल आँकड़ों की रेंज जाँच (कुल त्रुटि, जलवायविक परिसीमा जाँच), स्टेप जाँच (माप के अंतराल के दौरान अधिकतम और न्यूनतम स्वीकार्य भिन्नता) और अविरोधी जाँच (कालिक, आंतरिक और स्थानिक) पर क्रमिक रूप से लागू होती है। आँकड़ा संकेत के लिए भी नियम विकसित किए गए ताकि AWS आँकड़ों के उपयोगकर्ता वैध रूप से यह जान पाए कि QC की कार्य पद्धतियाँ किस हद तक त्रुटियों को दूर करते हुए आँकड़ों पर प्रयुक्त होंगी। इनसैट AWS प्रयोगशाला में इस कार्य पद्धति को आरम्भ किया गया है और अब उपयोगकर्ता मेटा आँकड़ा व QC संकेत के आँकड़े प्राप्त कर सकेंगे। QC पद्धति के आरंभिक परिणाम उत्साहवर्द्धक रहे तथापि प्रयुक्त किए गए एलगॉरिथ्म के निष्पादन की नियमित रूप से निगरानी की जाएगी और आवश्यकता पड़ने पर उन्नत एलगॉरिथ्म को QC पद्धति के अगले चरण में शामिल किया जाएगा।

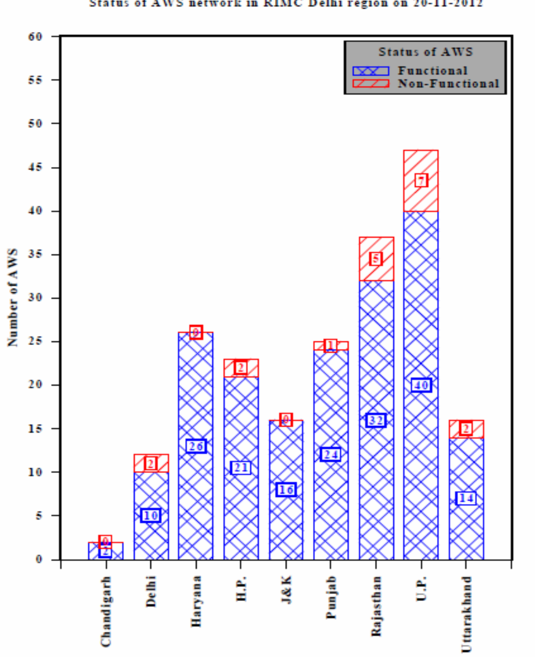
ABSTRACT. The procedures for automatic near real time monitoring of Indian Automatic Weather Station (AWS) network and Quality Control (QC) of data at central data receiving Earth station have been developed and results of test run are presented. In the absence of these procedures, AWS data may meet operational requirements but it would be a result of serendipity. The pervasive influence of erratic values of meteorological parameters received from AWS may cause end users to lose confidence in the system. It therefore behoves network managers to provide data of known and stated quality to end users.

The daily reports of network monitoring and system health are generated automatically and sent automatically by email to various maintenance centres and user defined email ids. The QC system sequentially applies range check (gross error, Climatological limit check), step check (maximum and minimum allowable variation during measurement interval) and consistency check (temporal, internal and spatial) on the raw data. The rules for data flagging are also developed as the users of AWS data may legitimately like to know the extent to which QC procedures applied on the data removed the errors. The system has been made operational at INSAT AWS laboratory and it would now be possible to make the data along with metadata and QC flag available to end users. The preliminary results of QC system are encouraging however, performance of algorithms implemented will be continuously monitored and if required, improved algorithms will be incorporated in next version of QC system.

Key words – AWS, Quality control.

1. Introduction

India Meteorological Department (IMD) has established a network of fixed land based Automatic Weather Stations (AWS) across the country. The network consists of total 675 stations of which a network of 125 AWS was established during 2006-07 (Ranalkar *et al*., 2012) and network of 550 AWS was established during 2009-12 (Ranalkar *et al*., 2014). The transmission techniques called Pseudo Random Burst Sequence (PRBS) is used for a network of 125 AWS (henceforth will be referred as 'PRBS type network') and Time a)



b)

Non-functional TDMA type AWS in RIMC Delhi region.

No.	Station Name	State Name	Make of station
001)	NARELA	Delhi	Astra
002)	SPORTS_COMPLEX	Delhi	Astra
003)	NARKANDA	H.P.	Astra
004)	KUKUMSERI_AGRO	H.P.	Astra
005)	MUKTASAR	Punjab	Astra
006)	KARAULI	Rajasthan	Astra
007)	PRATAPGARH	Rajasthan	Astra
008)	TONK	Rajasthan	Astra
009)	JAIPUR_IMD	Rajasthan	Astra
010)	BHARATPUR_IMD	Rajasthan	Astra
011)	PALIAKALAN_IMD	U.P.	Astra
012)	MAHOBA	U.P.	Astra
013)	FATEHPUR	U.P.	Astra
014)	BHINGA	U.P.	Astra
015)	PILIBHIT	U.P.	Astra
016)	UNNAO_AGRO	U.P.	Astra
017)	AGRA_IMD	U.P.	Astra
018)	PUROLA_IMD	Uttarakhand	Astra
019)	MUKTESHWAR_IMD	Uttarakhand	Astra

TDMA type AWS in RIMC Delhi region with intermittent data reception.

No.	Station Name	State Name	Make of station
001)	CHANDIGARH	Chandigarh	Astra
002)	PITAMPURA	Delhi	Astra
003)	BOPANI	Haryana	Astra
004)	KAUL_AGRO	Haryana	Astra
005)	KURUKSHETRA	Haryana	Astra
006)	MAHENDRAGARH	Haryana	Astra
007)	PANDUPINDARA	Haryana	Astra
008)	SADALPUR	Haryana	Astra
009)	SIRSA	Haryana	Astra
010)	JAGDISHPUR	Haryana	Astra
011)	JALANDHAR	Punjab	Astra
012)	BANSWARA	Rajasthan	Astra
013)	BARMER	Rajasthan	Astra
014)	JALORE	Rajasthan	Astra
015)	MOUNT_ABU	Rajasthan	Astra
016)	ANTA	Rajasthan	Astra
017)	VARANASI_IMD	U.P.	Astra
018)	KARWI_CHITRKOOT	U.P.	Astra
019)	KANNAUJ	U.P.	Astra
020)	MIRDHA	U.P.	Astra
021)	KANPURDEHAT	U.P.	Astra
022)	SULTANPUR_IMD	U.P.	Astra

Figs. 1(a&b). Status of AWS network in RIMC, New Delhi region on 20th November, 2012. (a) Functional status of AWS in each state and (b) List of non-functional and intermittently transmitting stations

Division Multiple Access (TDMA) technique is being used for a network of 550 AWS (henceforth will be referred as 'TDMA type network'). The features of telemetry systems have been finalized by Indian Space Research Organization (ISRO) and are mandatory for all users of Data Relay Transponders (DRT) onboard INSAT series of satellites.

The data are received in the Central Data Receiving Earth Station (henceforth referred to as ES) facility at IMD, Pune where two ES are available for reception of AWS data of PRBS type network and TDMA type network. The transmission techniques incorporate data reduction (scaling down of values of meteorological parameters so that it is accommodated in 10 bits), Error Detection (parity bits, Cyclic Redundancy Check) and Forward Error Correction (FEC) techniques (BCH code and convolution code) so as to check and minimize loss of data during transmission. This is also achieved through repeat transmission of data bursts (twice in case of TDMA type network and thrice for PRBS type network). The multiple bursts are useful not only for ensuring completeness of data but also for quality control of data being received from remote field stations. The data bursts received at ES are stored in raw format in ASCII text files. The raw data are then decoded by the processing software using inverse scale factors and engineering values of parameters are stored in relational database. The measurements are taken at full hour UTC but are transmitted to the ES sequentially at assigned time stamp (in TDMA technique) or time window (in PRBS technique) during following hour prior to next hourly measurement. Thus, data of entire network are available at the ES within one hour of measurement. As the data should be made available immediately to the end users on reception at ES so that its operational utility is not lost, the WMO SYNOP and BUFR code message is generated and disseminated to end users at an interval of every 15 minutes. The values of parameters dew point temperature, mean sea level pressure/geopotential height of nearest isobaric level etc. are derived from basic parameters and are available only in SYNOP and BUFR messages disseminated on GTS.

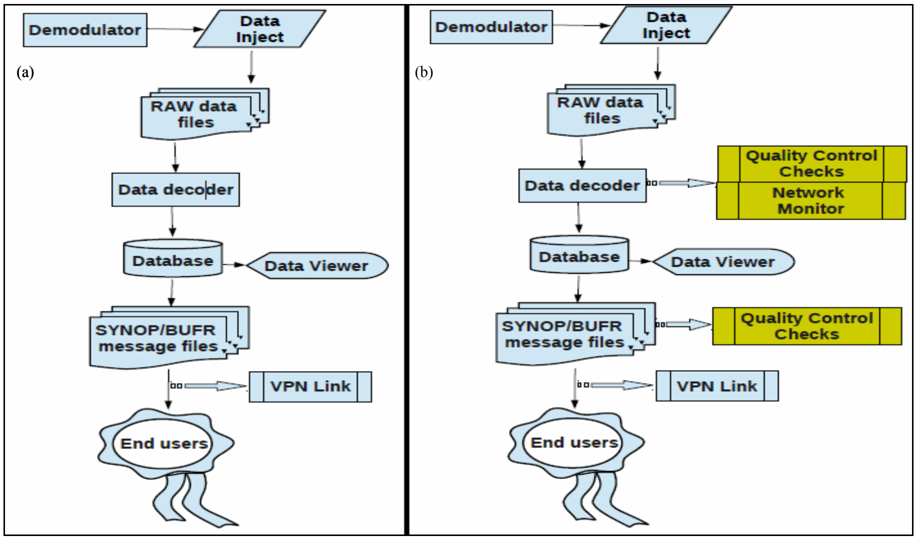
The need to document quality control algorithms for AWS data has been stressed by the WMO Expert team on requirements and implementation of AWS platforms (WMO, 2012). Operational utility of AWS data in India has been documented by Amudha and Raj (2013). However, concerns about quality of data have been raised by some users (personal communication). In the absence of robust quality control of data a bad observation could be transmitted to end users. At times, erroneous but apparently reasonable observations may also reach end users. This accentuates need to have robust quality control system for AWS data. The purpose of quality control is to automatically minimize the number of inaccurate observations and if possible generate missing

Station Name	State		00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	Total hrs with bursts received
Seppa	Arunachal Padesh	Burst reception	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	00
		Battery Voltage	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	
		GPS Lock/Unlock	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	
Deomali_ Agro	Arunachal Padesh	Burst reception	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	24
		Battery Voltage	12.8	13.0	13.5	13.9	15.1	13.9	15.1	15.0	14.1	14.2	13.8	13.4	12.9	12.9	12.9	12.9	12.8	12.9	12.8	12.8	12.8	12.8	12.8	12.7	
		GPS Lock/Unlock	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	
Tezu_ Agro	Arunachal Padesh	Burst reception	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	24
		Battery Voltage	12.6	13.6	14.2	14.5	14.5	14.5	14.5	14.5	14.5	14.5	13.6	13.2	12.9	12.8	12.8	12.7	12.8	12.8	12.7	12.7	12.7	12.7	12.7	12.7	
		GPS Lock/Unlock	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	
Roing	Arunachal Padesh	Burst reception	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	00
		Battery Voltage	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	
		GPS Lock/Unlock	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	×	
Anini_ Agro	Arunachal Padesh	Burst reception	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	24
		Battery Voltage	12.6	13.1	13.7	14.2	14.3	14.3	14.3	14.3	14.4	14.3	14.1	13.5	13.0	12.7	12.7	12.7	12.8	12.7	12.7	12.7	12.6	12.6	12.7	12.7	
		GPS Lock/Unlock	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	
Basar_ Agro	Arunachal Padesh	Burst reception	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	24
		Battery Voltage	12.4	12.6	13.2	14.3	14.3	14.4	14.1	14.3	14.4	14.4	14.4	13.2	12.7	12.6	12.6	12.5	12.5	12.5	12.5	12.5	12.4	12.4	12.4	12.4	
		GPS Lock/Unlock	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	
Lunglei	Mizoram	Burst reception	√	√	√	√	√	√	√	√	√	√	√	×	√	√	√	×	√	×	×	×	×	×	×	×	15
		Battery Voltage	10.5	10.9	11.5	12.7	12.7	14.2	14.0	13.8	13.4	13.0	11.1	×	10.7	10.6	10.6	×	10.5	×	×	×	×	×	×	×	
		GPS Lock/Unlock	L	L	L	L	L	L	L	L	L	L	L	×	L	L	L	×	L	×	×	×	×	×	×	×	
Dhemaji	Assam	Burst reception	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	24
		Battery Voltage	12.8	13.4	14.5	14.5	14.5	14.5	14.5	14.6	14.5	14.6	14.1	13.4	12.9	13.0	13.0	12.9	13.0	12.9	12.9	12.9	12.9	12.9	12.8	12.9	
		GPS Lock/Unlock	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	
Tinsukia	Assam	Burst reception	×	×	×	×	×	×	×	×	×	√	×	√	√	√	√	√	√	√	√	√	×	√	√	√	13
		Battery Voltage	×	×	×	×	×	×	×	×	×	12.9	×	12.5	12.7	12.7	12.7	12.7	12.7	12.7	12.7	12.7	×	12.7	12.7	12.6	
		GPS Lock/Unlock	×	×	×	×	×	×	×	×	×	L	×	L	L	L	L	L	L	L	L	L	×	L	L	L	
Dibrugarh	Assam	Burst reception	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	√	24
		Battery Voltage	12.7	12.7	13.2	13.5	14.4	13.5	13.5	14.4	13.8	13.6	13.4	13.2	12.7	12.7	12.7	12.7	12.7	12.6	12.7	12.7	12.7	12.6	12.7	12.7	
		GPS Lock/Unlock	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	
No. of stations with bursts received			007	007	007	007	007	007	007	007	007	008	007	007	008	008	008	007	008	007	007	007	006	007	007	007	

L - Data logger Real Time Clock (RTC) is synchronized to GPS time
U - Data logger Real Time Clock (RTC) could not be synchronized to GPS time
√ - Data burst for the hour is received at TDMA type Earth station
× - Data burst for the hour is not received at TDMA type Earth station

Fig. 2. Hourly status of AWS Network in RIMC Guwhati region for selected 10 stations on 17th October, 2013

observations by means of suitable software routines (WMO, 2008). Ideally the quality control of data should start at the field site (Level-0 QC) itself. However, the data loggers in IMD network of AWS have no procedures incorporated for quality control of measured value, hence Level-0 QC is not discussed in this paper. In the present



Figs. 3(a&b). Data processing setup at ES, (a) Data flow and process sequence presently implemented and (b) Data flow and process sequence developed in this study

set up, on receipt of data at ES, near real time quality control procedures (Level-1 QC) can be implemented prior to generation of WMO SYNOP and BUFR messages without introducing significant delay in onward transmission of data over GTS. It should, however, be borne in mind that QC checks are last line of defence in preventing erroneous data from reaching the end users. As Howell *et al*. (1984) points out, many important measures such as site exposure conditions, selection of instruments including sensors, sampling and measurement intervals, schedule of maintenance and calibration of equipment, self-diagnosis by acquisition system etc. should be taken before applying QC algorithms.

In this paper, an attempt has been made to develop automatic network monitoring and pragmatic quality control and data flagging system suitable for meteorological observations received from AWS at the ES. The algorithms include range check (gross error check and check for Climatological limit), time consistency check, internal consistency check and spatial consistency check. The procedures presented in this paper have been tested and are primarily meant for automatic near real time QC of the data received at ES. The QC system is written in FORTRAN-95 and network monitoring procedures have been developed using FORTRAN interface to Python and graphics library called DISLIN (http://www.mps.mpg.de/dislin/) developed by Max Planck Institute for Solar System Research, Germany. It is a challenge to provide data of known and stated quality to the end users in near real time; however, the task is not beyond realms of possibility.

2. Network monitoring

Inherently, network monitoring is done with an objective to facilitate improvement in network availability. The monitoring of network of such gigantic proportions by human is practically impossible. An effective automatic monitoring of the network of AWS is possible through a combination of software system and manual intervention where scanning of records and generation of statistics is done by software based on which decisions are taken by maintenance centres. The raw data file of previous day is scanned daily at 2340 UTC (or at defined time) to check status of each AWS based on number of data bursts received during 0000 UTC to 2300 UTC. If no data burst is received in past 24 hours the station is classified as non functional station. If data for less than 20 hours in a day (*i.e*., less than 80% of data in a day) is received from an AWS at the ES then such station is considered to be intermittently transmitting station.

Based on the statistics, a state-wise report on network monitoring comprising of graphical representation of number of non-functional stations, list of non-functional and intermittently transmitting stations is generated for each Regional Instruments Maintenance Centre (RIMC) region. A Python script is developed at the ES to automatically send these reports to all concerned authorities using Simple Mail Transfer Protocol (SMTP) client session object. A typical output of FORTRAN program generating daily network status of AWS installed in the region under the control of RIMC, Delhi is shown in Fig. 1(a) and list of non-functional and intermittently

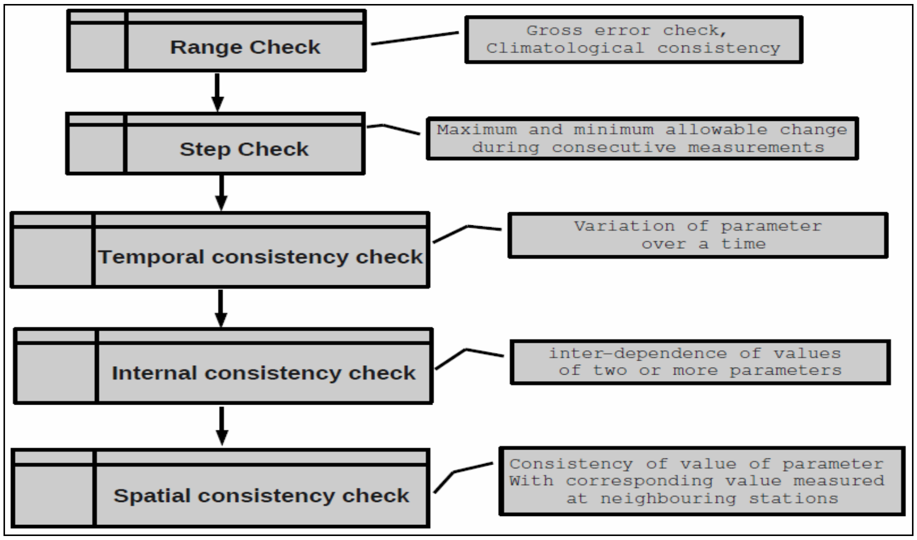


Fig. 4. Sequence of quality control procedures applied on the data

transmitting stations in this region on a day is presented in Fig. 1(b) respectively.

The report as shown in Figs. 1(a&b) provides the status of AWS in each state to maintenance centres along with detailed list of stations that warrants urgent maintenance. However, this report does not provide any quantitative information on the health of the system. Before undertaking maintenance tours to the stations listed out in the report, the maintenance centres would like to know information on battery voltage and synchronization of data logger RTC to GPS timing signal. In order to address this issue, a separate report on hourly battery voltage and GPS lock information for each station is generated. The output of FORTRAN program for selected 10 stations in RIMC Guwahati region is shown in Fig. 2. It can be seen from the report that information on deteriorating voltage and failure in RTC update using GPS signals is crucial information for maintenance of stations.

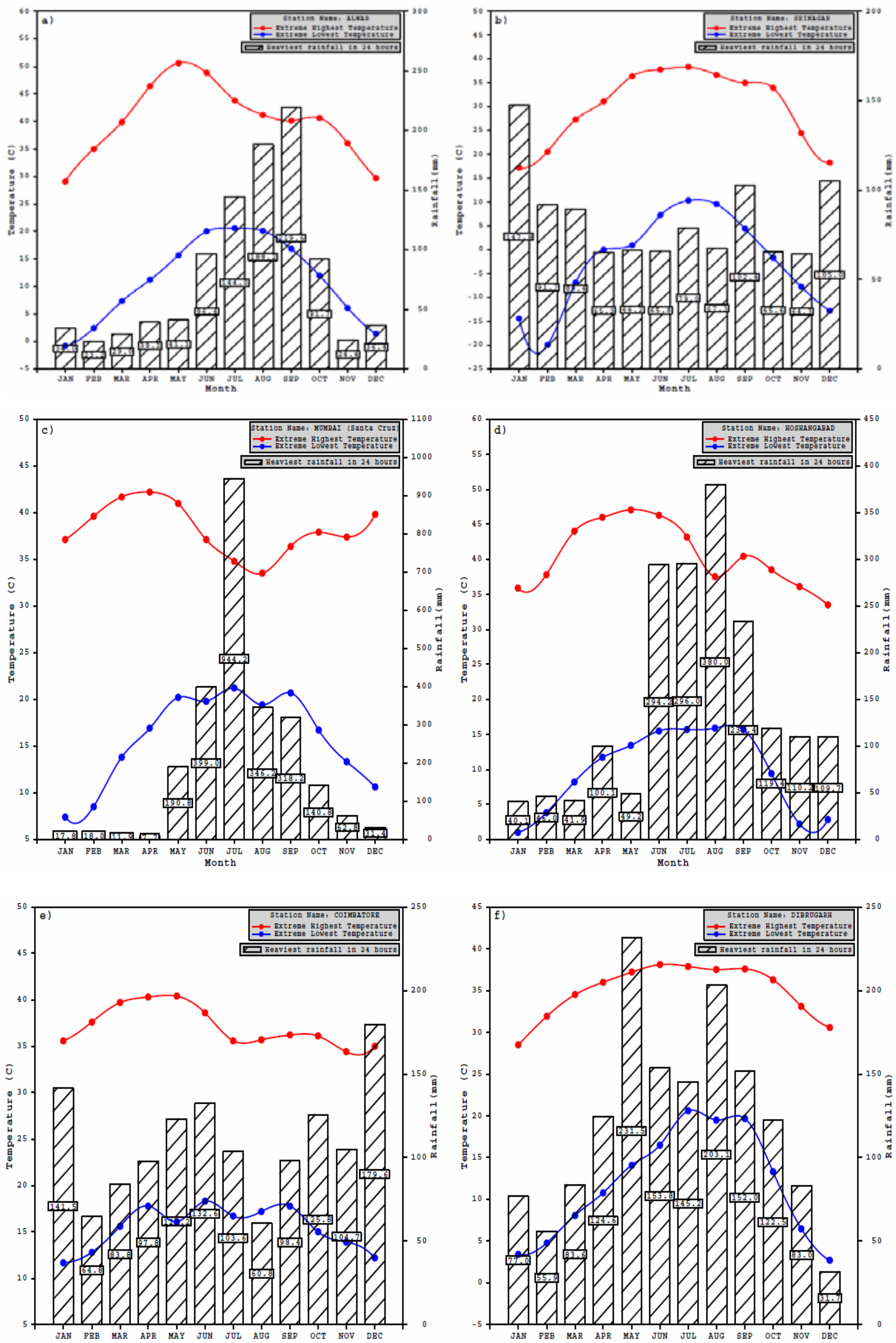
3. QC of data

Every data burst in raw form received at ES from AWS of TDMA type network (PRBS type network) is stored in daily (hourly) text files. Ideally, every hour two (three) data bursts (containing same hourly data) should be received at ES from AWS in TDMA type network (PRBS type network). The repeat transmission of data bursts is done to minimize transmission losses especially due to collision with other data bursts. The current raw data file is scanned at the ES at defined interval (usually 5-10 minutes) and data bursts received in the file since previous scan are decoded into engineering values of parameters. These values are then flushed to operational database. The WMO SYNOP and BUFR message is generated from the values in the database and are disseminated to the end users. Though some QC checks are implemented in transmission techniques to minimize loss of data during transmission the present system lacks QC of values of meteorological parameters before making them available to end users for operational utilization. It is pertinent to apply QC procedures at the stage of decoding raw data to obtain engineering values of parameters. This approach ensures that only quality controlled data is archived in the database. The parameters such as dew point temperature, mean sea level pressure, daily maximum and daily minimum temperature etc. are derived from basic meteorological parameters and encoded only in WMO messages. Thus, QC procedures should also be applied on database so as to generate messages free of erratic values. Fig. 3 is a schematic depicting current data processing sequence and that developed in this study.

In following sections, for the sake of brevity, analysis and discussion will be restricted to TDMA type network. However, it can be extended to PRBS type network with minor modifications. The QC procedures proposed and developed in this study are applied sequentially as shown in Fig. 4.

3.1. *Range check*

The range check primarily screens out the outliers in the data of single station. Generally, sensor range and Climatological extreme dictates rules for screening the data. O'Brien and Keefer (1985) have proposed such rules based on high/low range limit values and continuous no observed change in the value of a parameter. The range check is efficient in eliminating gross errors in the data.



Figs. 5(a-f). Climatological extremes of representative surface observatory stations used for range and Climatological consistency check. (*Source* : Climatological Tables, IMD, 2010 and www.imdpune.gov.in)

These errors may occur due to satellite transmission problems (parity errors), sensor malfunction, poor battery health and broken cable connections. The range check therefore detects values that are certainly wrong or impossible either meteorologically or physically. We have used sensor range and Climatological extremes to screen the data. For stations that do not have Climatological records a heuristic approach is taken to decide screening rules.

In this check, value of a parameter is first checked with the range of sensor. If the value is beyond the sensor

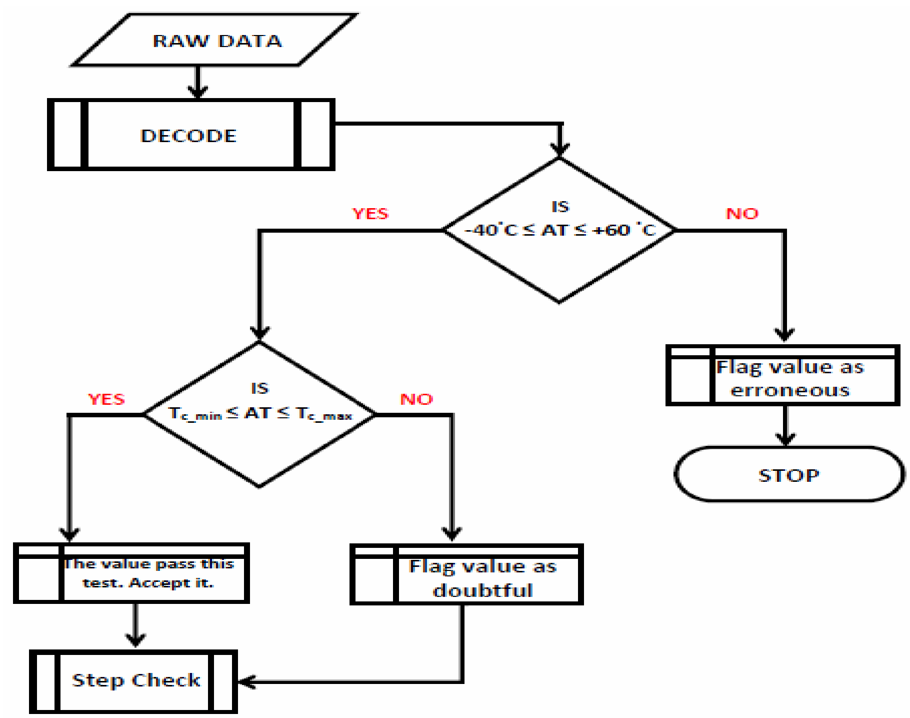


Fig. 6. Algorithm for range and Climatological consistency check of air temperature

range (or physically impossible) then it is discarded. Very often, the value of a parameter may be in error though it is within sensor range (or physical limit). It is therefore, reasonable to compare the value of parameter with monthly Climatological extremes so that unusual values are flagged as either erroneous or doubtful. The monthly extreme values also account for seasonal variations in the parameters.

The monthly Climatological extremes of temperature and heaviest 24 hour rainfall were obtained from Climatological Tables (IMD, 2010) (also available at www.imdpune.gov.in) and kept in a look up table. These extreme values for representative surface observatory stations are shown in Fig. 5. The extreme highest temperature of 50.6 °C was recorded at Alwar (Rajasthan) observatory on 10th May, 1956 and extreme lowest temperature of -20 °C was recorded at Srinagar (Jammu and Kashmir) observatory on 6th February, 1895. These are highest and lowest ever recorded temperature (in known meteorological history) over Indian landmass. The monthly extreme values of individual stations are expected to lie within this range, however, aberrations cannot be ruled out if another extreme event occurs. The screening widow is then shrunk to monthly extreme values of stations. However, most of the AWS have been installed at sites for which Climatological records are not available. Hence, Climatological consistency of a group of AWS is checked against extremes of nearby representative observatory.

The algorithm for air temperature tested on raw data bursts is shown in Fig. 6. The sensor range is - 40 to +60 °C, hence, temperature less than -40 °C and greater than +60 °C would certainly be an erroneous value. The data received at the Earth Station on 5th and 6th June, 2013 from Hoshangabad AWS is presented in Fig. 7. At 2200 UTC of 5th June, 2013 and 0000 UTC of 6th June, 2013, an erroneous value of -120 °C was received at the ES. The value is not only beyond sensor range but also meteorologically impossible and hence would be rejected. At 2300 UTC of 5th June, 2013, temperature of 50.4 °C was received at the ES which is within sensor range but exceeds the Climatological extreme ever recorded at Hoshangabad in the month of June as shown in Fig. 5. Similarly, an obtrusively erroneous value of 0 hPa for station level pressure received from Churu AWS in Rajasthan and hourly rainfall of 1023 mm received from Hoshangabad AWS are shown in Figs. 8(a&b) respectively.

3.2. *Time consistency check*

The temporal consistency of the datum is checked by subjecting the value to step check. This check is efficient in isolating the outliers in the time series. In a step check the value of parameter '*X*' at time '*t*' is checked with the value at time '*t-h*'. It is based on the principle of maximum allowable variation (*k*) in a parameter within certain period of time. The values may have wide variation for different stations depending on diurnal range of parameter

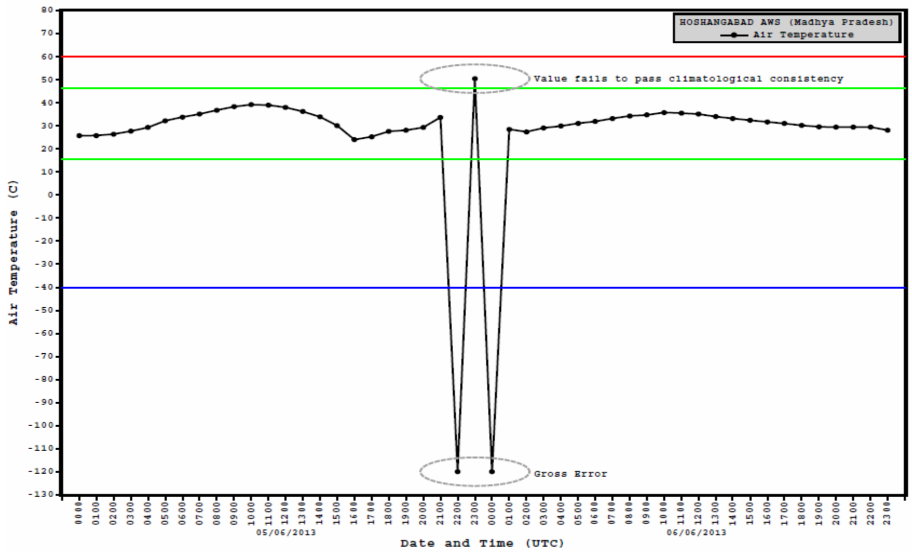


Fig. 7. Air temperature data received at ES on 5th and 6th June, 2013 from Hoshangabad AWS. (Blue and Red colour lines delineates sensor range and Green line demarcates Climatological extreme for the month of June)

TABLE 1

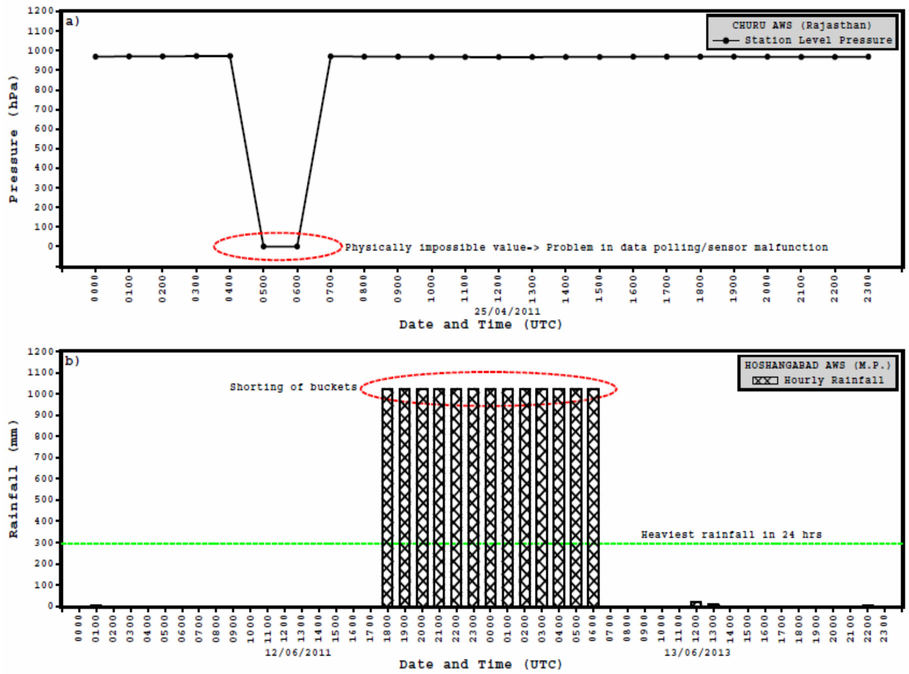
Limits of step check for air temperature and pressure for Adilabad AWS

Parameter	Period	Flag	J	F	M	A	M	J	J	A	S	O	N	D
Pressure	1 hr	2	2.0	2.0	1.5	2.5	3.5	3.5	2.0	2.0	2.0	3.0	2.0	2.0
	3 hrs	3	3.0	4.0	4.0	5.0	5.0	5.5	4.0	4.0	4.0	5.0	4.0	3.0
Temperature	1 hr	2	8.0	7.0	7 .0	7.0	8.5	9.0	7.0	6.0	5.0	7.0	6.0	7.5
	3 hrs	3	10.0	10.0	11.0	12.0	14.0	14.0	10.0	8.0	7.0	10.0	10.0	10.0

being checked and local climatic conditions. If $\left|X(t-h)-X(t)\right|>k$, where h is an integer and for hourly data we chose h = 1 and 3 respectively then $X(t)$ is either erroneous or doubtful depending on the value of 'k'. Based on the data for the year 2011 and 2012, the limits of step check for air temperature and pressure for Adilabad AWS are given in Table 1. Such limits for selected stations can be used for group of stations. The step check is efficient in flagging suspicious values before dissemination. However, the drawback of this check is that it does not prevent correct value from being flagged as suspicious. Fig. 9 shows air temperature recorded at Adilabad AWS that fails to pass step check and flagged as suspicious.

For any geophysical parameter with continuous distribution and constant sampling rate, an outlier can be found using a dip-test (Øgland, 1993). In this test a positive real number δ is chosen for a parameter $x(t)$ to check acceptable dip in value of parameter. The value x_i which satisfies the condition $T=(x_{i-1}-x_i)(x_{i+1}-x_i)>\delta^2$ is regarded as suspicious and may be rejected. The dip test is associated with implicit hyperbola $xy=\delta^2$ scaled with a factor δ^2. For a test T of type where an observation x_i is rejected due to $T(x_{i-1}-x_i, x_{i+1}-x_i)$ being greater than some constant δ^2 independent of x_i, the area $A=\{(x,y):T(x,y)>\delta^2\}$ is defined as rejection area for the test. The dip test can be used in near real time and is useful in taking decision for regeneration of missing or erroneous values. A variant of dip test as proposed in the WMO report of expert team on surface technology and measurement techniques WMO (2004) requires knowledge of standard deviation of parameter calculated from at least past few measurements. If there is a sudden kink in the time series, it is possible to obtain incorrect data value (provided the parameter has continuous distribution) by applying interpolation/extrapolation technique as shown in Figs. 10(a&b). The kinks may occur if the value fails to pass range check or step check. However, it becomes increasingly difficult to reconstruct the time series, if consecutive values are either in error or missing.

For highly variable parameter such as rainfall with discrete distribution the step check is not useful. However,



Figs. 8(a&b). Gross error in the data received from AWS. (a) Physically impossible pressure value (0 hPa) received from Churu AWS and (b) Erroneous hourly rainfall (1023 mm) received from Hoshangabad AWS

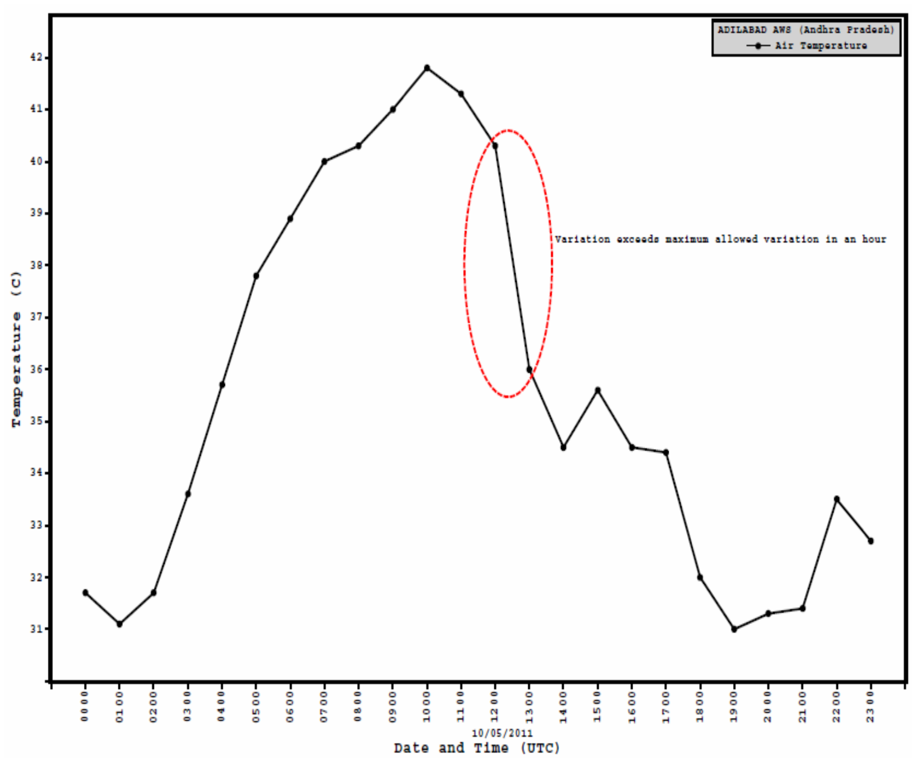
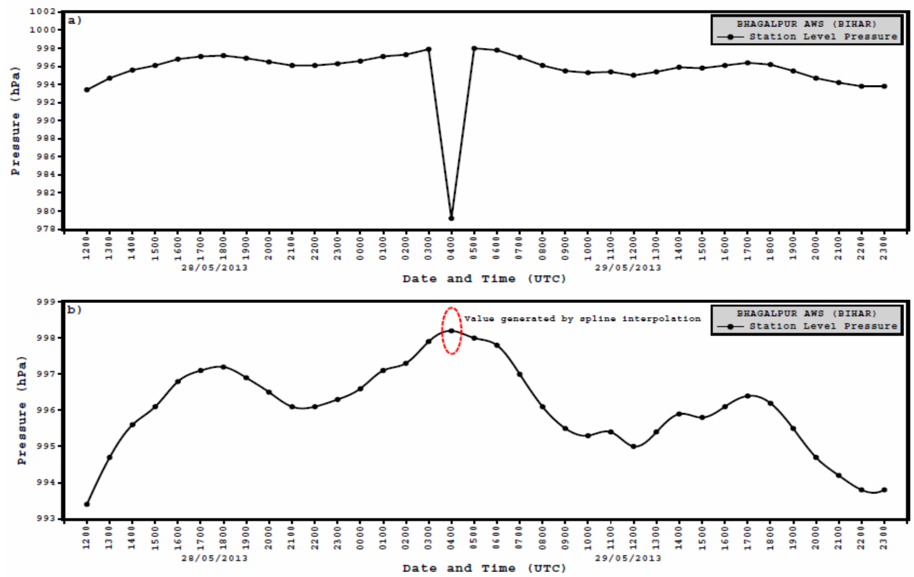


Fig. 9. Variation of air temperature at Adilabad AWS on 10th May, 2011

time consistency of accumulated rainfall can be checked using $RF_t < RF_{t-h}$, where RF_t and RF_{t-h} are accumulated rainfall from past 0300 UTC to '*t*' and '*t-h*' UTC respectively where *t* may range from 0300 UTC of current



Figs. 10(a&b). Scope for correcting erroneous values. (a) Kink in pressure value received from Bhagalpur AWS and (b) Pressure value generated by spline interpolation

day to 0500 UTC of previous day and '*t-h*' ranges from 0200 UTC of current day to 0400 UTC of previous day. Thus '*h*' may have an integer value 1 to 23.

3.3. *Internal consistency check*

The internal consistency checks are straightforward and can be easily implemented in automatic QC system. These checks are useful in trapping certain errors as well as possible error (WMO, 1993). For example, if relative humidity is low say 40% and at the same time rain is reported by station then either humidity or rainfall is wrong. Some obvious consistency checks for a given time of observation implemented at ES are given below.

T (Air temperature) < T_d (Dew point temperature) - γ

where, γ is positive real number (say 0.2) independent of T and T_d.

T (Air temperature) > $T_{\max}$ (Hourly maximum temp.)

T (Air temperature) < $T_{\min}$ (Hourly minimum temp.)

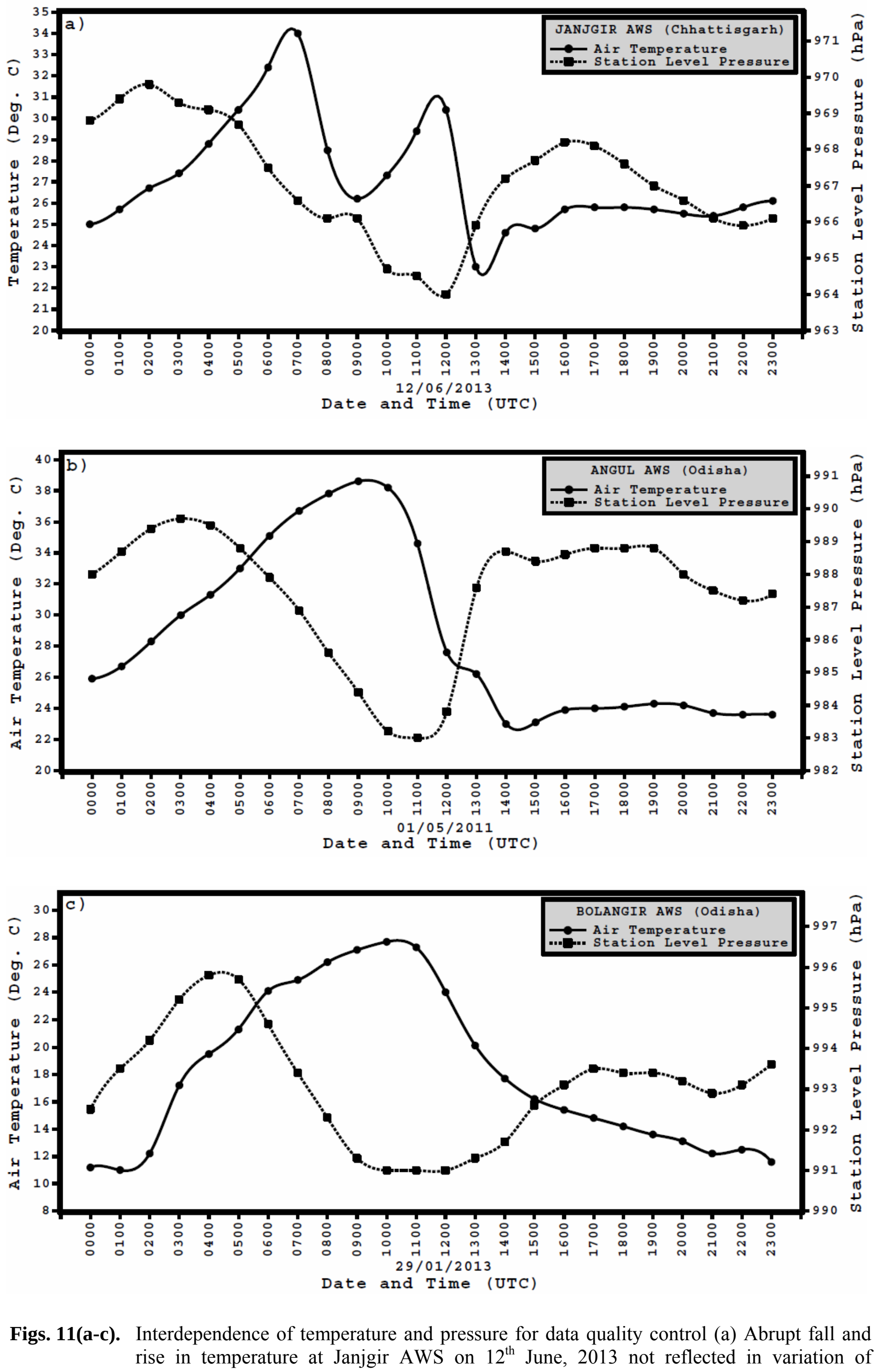
$T_{\max} - T > k$ and $T - T_{\min} > k$

$RF_t > RF_{t-1}$ and $rf_t = 0$, where t is current UTC hour, RF_t is accumulated rainfall since previous 0300 UTC and rf_t is hourly rainfall from 't-1' to 't' recorded at a station on particular day. In addition to above, internal consistency checks on SYNOP codes as mentioned in WMO (1993) have been implemented.

In general, daily maximum temperature at a station occurs between 0900 UTC and 1000 UTC. It can be seen from Fig. 11(a) that on 12th June, 2013 hourly air temperature at Janjgir AWS was maximum at 0700 UTC. It decreased till 0900 UTC and then again increased till 1200 UTC. However, station level pressure decreased steadily from 0200 UTC to 1200 UTC. From Fig. 11(a), we therefore infer that either temperature or pressure values are doubtful/erroneous during 0700 UTC to 1100 UTC as fall in pressure is not reflected in rise of temperature. The abrupt fall (rise) in temperature during evening hours (morning hours) at stations may often fail to pass maximum allowable variation in an hour though such fall (rise) could be realistic during summer season as depicted in Fig. 11(b) for Angul AWS in Odisha on 1 May, 2011. The steep fall in temperature between 1100 UTC to 1200 UTC is reflected in rise of pressure between 1200 UTC to 1300 UTC. In fair weather conditions, the diurnal variation of temperature and pressure as shown in Fig. 11 (c) for Bolangir AWS in Odisha turns out to be a handy tool for quality control of data.

3.4. *Spatial consistency check*

The Spatial Regression Test (SRT) as explained by You *et al*. (2008) is employed to check spatial consistency of measurement. This test checks whether the measurement falls within a confidence interval formed from surrounding station data during a time period of



Figs. 11(a-c). Interdependence of temperature and pressure for data quality control (a) Abrupt fall and rise in temperature at Janjgir AWS on 12th June, 2013 not reflected in variation of pressure, (b) Fall in temperature exceeding maximum allowable variation in an hour reflected in rise of pressure at Angul AWS in Odisha and (c) Diurnal variation of pressure and temperature at Bolangir AWS in Odisha on 29th January, 2013

length ‘*n*’. All stations (M) lying within a grid box of size 2° latitude × 2° longitude, centred on station of interest are considered and linear regression is performed for each station in the grid box paired with station of interest. For hourly AWS data we chose the time period n = 25 hours centred on the datum of interest. This accounts for diurnal variation of parameter having continuous distribution. A linear regression based estimate $x_i = a_i + b_i y_i$ is obtained for each neighbouring station, where y_i is measurement at i^{th} neighbouring station, x_i is regressed estimate of station of interest based on y_i and a_i and b_i are parameters of regression line. The weighted estimate x' is derived using standard error of estimate (s) also called Root Mean Square Error (RMSE) in the weighting function. $x' = \frac{\sum_{i=1}^{N} x_i s_i^{-2}}{\sum_{i=1}^{N} s_i^{-2}}$, where N is number of stations in the grid box with explained variance R^2 greater than 0.5. The SRT is executed only if at least 2 stations satisfying this condition are available in the grid box. The weighted standard error of estimate (s') is given by $\frac{1}{s'^2} = \frac{1}{N}\sum_{i=1}^{N} s_i^{-2}$. As time of measurement is fixed for all AWS, the systematic time shift as mentioned in You *et al*. (2008) would not result in the dataset. Hence, regression estimate (and standard error of estimate) for lagged time series were not computed. The estimated confidence intervals are based on s' and we check whether or not datum falls within the confidence intervals $x' - ks' \le x \le x' + ks'$. Increasing the value of 'k' decreases the chance of flagging the extreme value. The value of 'k' is set to 3. If the condition is satisfied, the datum passes spatial regression test.

It is mentioned above that time consistency check may flag correct value to be suspicious. As shown in Fig. 9 the value of 1300 UTC temperature (35.5 °C) recorded at Adilabad on 10^{th} May, 2011 fails time consistency check and is flagged as suspicious. It would be prudent to check if the value is spatially consistent. Four stations, *viz*., Karimnagar, Warangal, Jagtiyal and Rudrur lie in 2° latitude × 2° longitude grid box centred on Adilabad. The data of 25 hours centred on 1300 UTC of 10^{th} May, 2011 is considered for each station. The spatial regression test explained above yields 36.9 – 3 × 1.35 < 35.5 < 36.9 + 3 × 1.35, *i.e*., 32.9 < 35.5 < 41.0. As the datum lies in the confidence interval, it passes the SRT and hence accepted.

4. Sensitivity of algorithms for derived parameters

The dew point temperature and Mean Sea Level Pressure (MSLP) are derived at the ES from basic meteorological parameters. The algorithms used at ES to derive these parameters have been explained by Ranalkar *et al*. (2012). The accuracy of these parameters depends on sensitivity of the algorithms to variations in the parameters involved. We present the sensitivity of algorithm for derivation of MSLP. All National Meteorological and Hydrological Services use International Meteorological Tables (WMO, 1966) to reduce Station Level Pressure (SLP) to Mean Sea Level Pressure (MSLP). The table is based on a Laplace's formula (WMO, 1964) given below.

$$Z_p = \mathrm{K}\left(1+\alpha\theta_m\right)\left(\frac{1}{1-k\cos 2\phi}\right)\left(1+\frac{Z_P}{r}\right)\log\frac{P_o}{P_S}$$

$$Z_p = \mathrm{K}\left(1+\alpha\theta_m\right)\left(1+E\right)\left(1+\frac{Z_P}{r}\right)\log\frac{P_o}{P_S}$$

where,

Z_p = Geometric altitude corresponding to the station elevation

K = 18400 is Hypsometric Constant

α = 0.00367 °C $^{-1}$ Coefficient of expansion of air at STP

θ_m = Aie temperature + $\Gamma Z_p/2$ is mean temperature of air column, Γ = 0.0065 °C m^{-1}

k = 0.00259

$E = k \cos 2\varphi$

φ = Latitude of station

r = Mean terrestrial radius = 6371104 m

P_o = Pressure at mean sea level

P_s = Pressure at station level

Applying correction for humidity, equation (1) can be written as

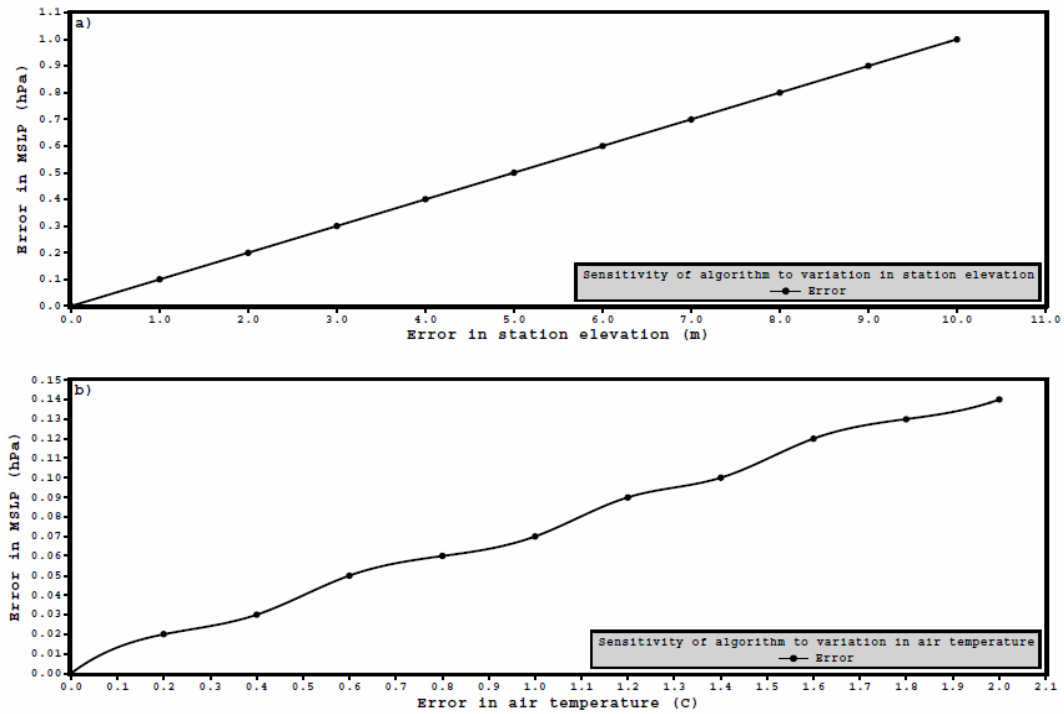
$$Z_p = K\left(1+\alpha\theta_m\right)\left(1+E\right)\left(\frac{1}{1-\beta}\right)\left(1+\frac{Z_p}{r}\right)\log\frac{P_o}{P_S}$$

where, $\beta = 0.378\frac{V_p}{P_m}$ and $P_m = \frac{B_s + B_o}{2}$

V_p = Mean vapour pressure of the air column,

P_m = Mean pressure of the air column between SLP and MSLP

B_s = Height of mercury column (Barometric height) at station level (mm)



Figs. 12(a&b). Sensitivity of algorithm for MSLP. (a) Influence of error in station elevation on error in MSLP and (b) Influence of error in air temperature on error in MSLP

B_o = Height of mercury column (Barometric height) at mean sea level (mm)

In India, the factor $\left(1+\frac{Z_p}{r}\right)$ is omitted. Hence,

$$Z_p = 18400\left[1+0.00367\left(AT+\frac{0.0065Z_p}{2}\right)\right]$$

$$(1+0.00259\cos 2\phi)\left(\frac{1}{1-0.378\frac{V_p}{P_m}}\right)\log\frac{P_o}{P_s}$$

$$Z_p = ABC\log\frac{P_o}{P_s}$$

$$P_o = P_s 10^{\frac{Z_p}{ABC}}$$

It is clear from above formula that station elevation, air temperature, mean vapour pressure of the air column and mean pressure of the air column between SLP and MSLP have bearing on the value of MSLP. Sensitivity of the algorithm to variation in station elevation and air temperature after holding all other parameters constant is depicted in Figs. 12(a&b). It is evident from Fig. 12(a) that as the error in station elevation increases the error in MSLP also increases and an error of 10 m in station elevation leads to an error of 1 hPa in MSLP. Similarly, an error of 2 °C in air temperature introduces an error of 0.14 hPa in MSLP as shown in Fig. 12(b). This underlines need to have accurate metadata of all stations so as to obtain accurate derived parameters.

5. Data flagging

The value of a parameter when subjected to QC system may be rejected, accepted, or corrected/ interpolated, hence, with automatic handling of data it is necessary to define flag which is attached to the value to indicate its status. Such explicit flag serves to provide data of known and stated quality to the end users. The flags can be framed based on the need of end users and extensive flagging of data can be used for internal use at data processing centre. It is desired that flags are embedded as metadata in the data being supplied to the end users so that they know whether the observation is 'good' or 'doubtful' and 'original' or 'corrected'. Manual inspection of quality of hourly data values at ES is practically impossible. All QC procedures at the ES are automatic and result of each

TABLE 2

Quality Control flags for data values

Flag number	Description
0	QC Not Done
1	Good
2	Doubtful
3	Erroneous
4	Changed/Interpolated or extrapolated
5	Missing Value

QC procedure is a flag number which is stored as an attribute of datum. It is planned to make data along with QC flags as mentioned in Table 2 available to the end users through appropriate BUFR descriptor so that user can make a decision on whether to use the value or not.

6. Summary

It is to be noted that in general, QC procedures are statistical in nature, hence; no such QC system can guarantee that erroneous data value will not pop up in the quality controlled data. Nevertheless, it helps to either reject the data with gross errors or attach suitable flag to the datum for its quality. This in turn helps in building confidence for utility of data. The results of QC system and network monitoring often traps problems related to sensor malfunction, health of AWS, transmission time management. This is very useful information for planning maintenance of stations. The results of automated QC system also highlights the need to have more sensor specific algorithms, rigorous QC of data at the national data centre before final archival and compliment it with manual intervention. The non-availability of Climatological data for many stations pose problem for QC system. It is however, reasonable to consider climatology of nearest neighbour in same climatic zone with comparable elevations. It is experienced that even with implementation of only range check and time consistency check quality of data significantly improves. The QC system being implemented at AWS laboratory is still evolving with a scope for improvement of algorithms.

Acknowledgement

This work is a part of in-house development of Network monitoring and QC system at INSAT AWS Laboratory. These procedures are now being implemented as GUI software linked with relational database along with additional features. The authors wish to express their gratitude towards Dr. L. S. Rathore, Director General of Meteorology, for encouraging this activity and being enduring source of inspiration. The first author desires to express his gratitude towards Dr. Igor Zahumansky, Director WMI, WMO HQ, Geneva for invaluable discussions on the topic, Dr. Jitze P. Van der Meulen of KNMI Weather Research, Netherlands for the help rendered in getting the algorithm for reduction of SLP to MSLP and Ms. Laetitia Rey of WMO Library at Geneva for providing WMO documents.

References

Amudha, B. and Raj, Y. E. A., 2013, "Operational weather forecasting using data from Automatic Weather Stations and other modern observing systems- Case study of tropical cyclone Jal 2010", *Mausam*, **64**, 437-456.

Howell, T. A., Meek, D. W., Phene, C. J., Davis, K. R. and McCormick, R. L., 1984, "Automated weather data collection for research on irrigation scheduling", *Transcript of the American Society of Agricultural Engineers*, **27**, 2, 386-391.

India Meteorological Department, 2010, "Climatological Normals (1961-1990)", published by O/o the ADGM (R), IMD, Pune.

O'Brien, K. J. and Keefer, T. N., 1985, "Real-time data verification", Proc. ASCE Special Conf., Buffalo, NY, American Society of Civil Engineers, 764-770.

Øgland, P., 1993, "Theoretical analysis of the dip-test in quality control of geophysical observations", DNMI Report No. 24/93 KLIMA.

Ranalkar, M.R., Mishra, R.P., Anjan, A. and Krishnaiah, S., 2012, "Network of Automatic Weather Stations: Pseudo Random Burst Sequence Type", *Mausam*, **63**, 4, 587-606.

Ranalkar, M. R., Gupta, M. K., Mishra, R. P., Anjan, A. and Krishnaiah, S., 2014, "Network of Automatic Weather Stations: Time Division Multiple Access Type", *Mausam*, **65**, 3, 393-406.

World Meteorological Organization, 1964, "Note on the standardization of pressure reduction methods in the international network of synoptic stations", WMO No. 154.TP.74, 1st Edn.

World Meteorological Organization, 1966, "International Meteorological Tables", WMO No. 188.TP. 94, 2nd Edn.

World Meteorological Organization, 1993, "Guide on the global data processing system", Chapter - 6, WMO No. 305, 2nd Edn.

World Meteorological Organization, 2004, "Guidelines on quality control procedures for data from Automatic Weather Stations", CIMO/OPAG-SURFACE/ET ST&MT-1/Doc.6.1(2).

World Meteorological Organization, 2008, "Guide to Meteorological Instruments and Methods of Observation", Part-I, Chapter- 1, WMO No. 8, 7th Edn.

World Meteorological Organization, 2012, "Expert Team on Requirements and Implementation of AWS Platforms (ET-AWS)", Seventh Session, CBS ET-AWS-7.

You, J., Hubbard, K. G. and Goddard, S., 2008, "Comparison of methods for spatially estimating station temperatures in a quality control system", *Int. J. Climatol.*, **28**, 777-787.