

AWS Data Quality and Control

Isaac Mugume

Makerere University, Uganda
University of Bergen, Norway

August 30, 2016



- Necessity of using accurate data
- reduction of errors in initial conditions
- obs could have errors due to hardware/software prob, siting & exposure prob
- So
- data quality control can help check inconsistencies in data
- note: bad data may replace good data is robust QMS is not in place



- use of principles (i.e. dew–point temp. can't be greater than dry bulb temp)
- are we registering unheard/unseen values (i.e. temp $>$ 50 deg.C in Ug.)
- is data defying known variations



- missing data (complete day, hour, minutes)
- multiple sends in a given minute
- repetition (duplicating same data)
-



- how our sensors perform at night? Do they resolve radiational cooling?
- how do they perform during rainy events
- the wind sensors: how do they resolve abrupt wind gusts? do they have acceptable hysteresis?
- can our sensors give alerts when weather is coming to un acceptable values

Acceptable data quality considerations



- a detailed manual for each component
- component durability
- easy to fix/maintenance
- adaptable, user-friendly softwares incl. interfaces
- adequate protection from external environment (lightening, dust, rodents, weather parameters that can affect another)
- should recover from extremes
- self-checking diagnostic; should send self-diagnostic report at regular intervals
- siting & exposure of AWS
- reproduction of acceptable known constants (i.e. triple point, calm conditions)
- consideration of acceptable range / uncertainty
- ability to recover from disaster
- time-lag/response
- missing records assigned a value (e.g. -9999)



- how is the life of the sensor (i.e. deterioration rate)
- how 'elastic' is the sensor in terms of data capture?
- operational models of AWS (i.e. if sending fails, is possible to recover data)
- manual of potential probs and how to handle them
- suspect data should not just be discarded.
- what is acceptable time resolution (1min real-time analysis, how to take care of other standard reports: METAR/SPECI, SYNOP)
- regular monitoring and logging (i.e. even AWS working normally is good rather than missing it): what is acceptable real-time monitoring/logging? (15min, 1hr)