Assuring the quality of observations

Karolina Stanisławska, Hálfdán Ágústsson

Seychelles, 24 February 2016

Outline

- Problem definition
- Automatic quality assurance system
- Overview of Belgingur QA system
- A case study

Overview

- Measurement devices can cause multiple types of errors
- Erroneous observations can lead to wrong assessments
- Important to identify malfunctioning gauges or periods of observations that can't be trusted

Sources of problem

- A measurement device stopped working, battery is off, memory was full
- Improperly calibrated devices
- Sensor was stuck for some reason (icing...)
- Lightning activity

Spotting the problem

- Are the values within their assumed range?
- Do the observations vary with time? Don't they vary too much?
- Is the data series smooth or does it have spikes? Are there any single missing values from the time series?
- Are the observations from neighbouring stations similar?

- Manual inspection

Automatic Quality Assurance

- Shafer, M. A., C. A. Fiebrich and D. S. Arndt (1999). Quality assurance procedures in the Oklahoma network. J. Atm. Ocean. Tech., 17(4):474–494.

- Perform **four quality tests** on **daily portions** of data and assign one of **six quality indicators** for each test

Quality checks

- Range check
- Step check
- Persistence check
- Spatial check

QA: range check

- Compare each observation datum, separately for each variable, to typical sensor performance and the expected climatological extremes
- Do all the variables fall within their ranges?
- Detect flaws such as oversaturation, incorrectly calibrated instruments

QA: step check

- Is the timeseries continuous?
- Aren't the jumps between consecutive observations too large?

QA: persistence check

- Works on data from a whole day at a time
- Detect too small changes in the standard deviation of the observed variable
- Detect too short span of the observed variable within the day
- Marks the whole day for given variable according to the result

QA: spatial check

- How does the result for this station correspond with neighbourhood ones?
- At least six stations must be analyzed concurrently

QA flags

0 – pass

1 – suspect

2 – warning

3 – failed

8 – not tested

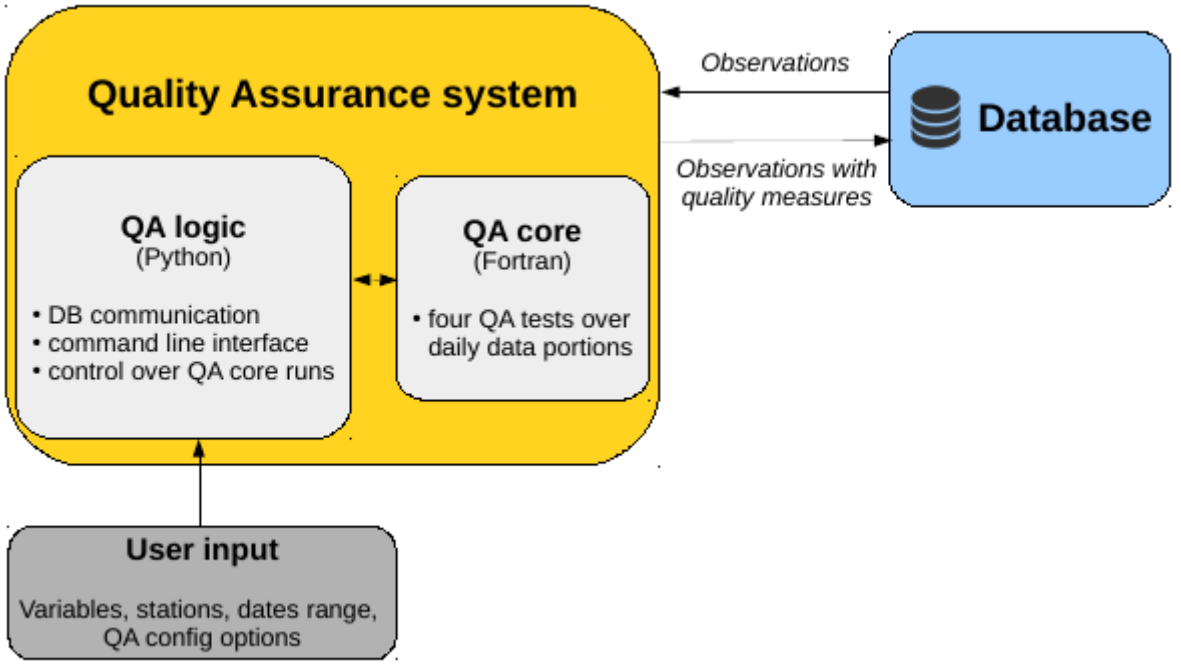
9 – missing

QA results

- 0 – pass, 1 – suspect, 2 – warning,
  3 – failed, 8 – not tested, 9 – missing

time	station_id	variable	value	q_range	q_step	q_persistence	q_spatial
2010-07-02 11:00:00+00	908	rel_hum	106	3	8	8	8
2010-07-02 11:00:00+00	908	temp	3	0	0	0	0
2010-07-02 11:00:00+00	908	wind_dir	132	0	0	0	0
2010-07-02 11:00:00+00	908	wind_speed	3.2	0	0	0	0
2010-07-02 12:00:00+00	908	rel_hum	106	3	8	8	8
2010-07-02 12:00:00+00	908	temp	3.3	0	0	0	0
2010-07-02 12:00:00+00	908	wind_dir	183	0	0	0	0
2010-07-02 12:00:00+00	908	wind_speed	1.7	0	0	0	0
2010-07-02 13:00:00+00	908	rel_hum	103	0	8	0	0
2010-07-02 13:00:00+00	908	temp	3.4	0	0	0	0
2010-07-02 13:00:00+00	908	wind_dir	156	0	0	0	0
2010-07-02 13:00:00+00	908	wind_speed	3	0	0	0	0

The QA system of Belgingur



- ...first, populate the database with data

Defining metadata

- Create a setup file describing your stations and use our metadata importer to get it to a database

```
provider: {name: Vedurstofan, ref: vi.is}
stations:
 - {active: true, has: 52.0, lat: 64.1275, lon: -21.9028, manual: true, name: Reykjavík, ref: vi.is.rvk, ref_provider:
 - {active: false, has: 20.0, lat: 64.417, lon: -21.833, manual: true, name: Leirá, ref: vi.is.lra, ref_provider: 96,
 - {active: true, has: 14.0, lat: 64.6419, lon: -21.590799999999998, manual: true, name: Stafholtsey, ref: vi.is.stey,
 - {active: false, has: 46.0, lat: 64.667, lon: -21.617, manual: true, name: Hamraendar í Stafholtstungum, ref: vi.is.h
 - {active: false, has: 45.0, lat: 64.81700000000001, lon: -22.267000000000003, manual: true, name: Haukatunga, ref: vi
 - {active: false, has: 80.0, lat: 64.80000000000001, lon: -23.033, manual: true, name: Neðri-Hóll, ref: vi.is.ndrh, re
 - {active: false, has: 12.0, lat: 64.81700000000001, lon: -23.133, manual: true, name: Garðar, ref: vi.is.grdr, ref_pr
 - {active: true, has: 13.0, lat: 64.8393, lon: -23.3003, manual: true, name: Bláfeldur, ref: vi.is.blfl, ref_provider
 - {active: false, has: 70.0, lat: 64.9, lon: -23.933, manual: true, name: Gufuskálar, ref: vi.is.gfsk, ref_provider: 1
 - {active: false, has: 12.0, lat: 64.917, lon: -23.883000000000003, manual: true, name: Hellissandur, ref: vi.is.hsd,
```

Importing observations to a database

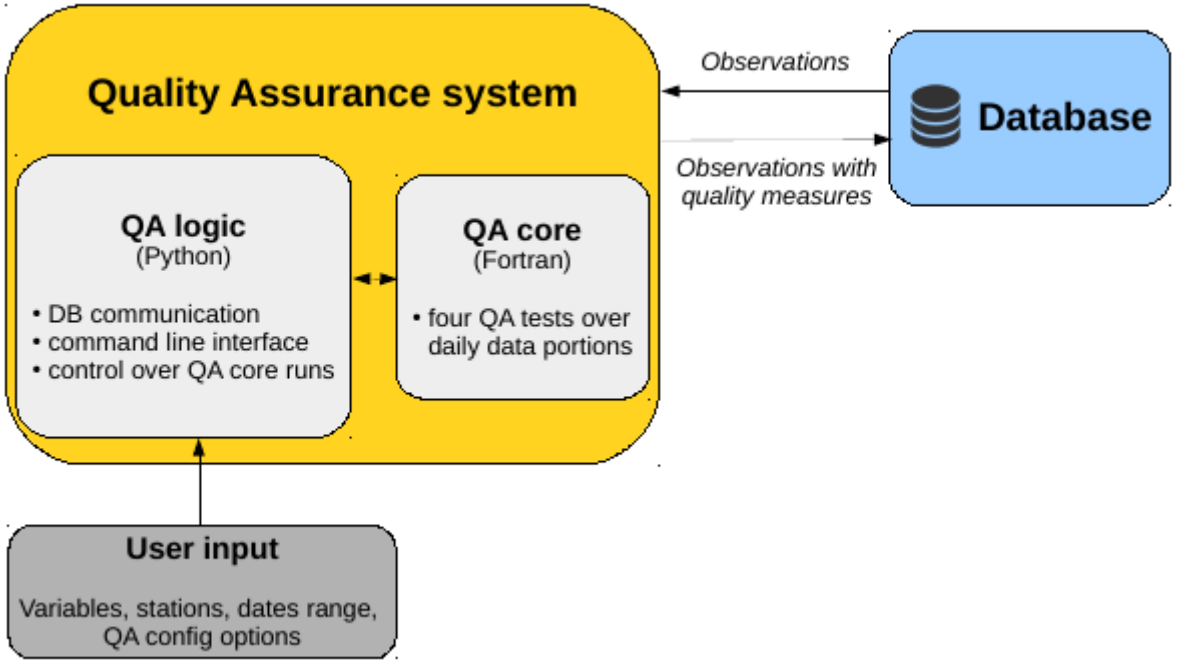
- Transform your data to this format...

```
#station_ref: vi.is
#provider_ref: 96
time, temp, wind_speed, rel_hum
2000-01-01T00:00:00, 11.1, 11, 91
2000-01-01T00:10:00, 11.2, 12, 92
2000-01-01T00:20:00, 11.3, 13, 93
2000-01-01T00:30:00, 11.4, 14, 94
2000-01-01T00:40:00, 11.5, 15, 95
```

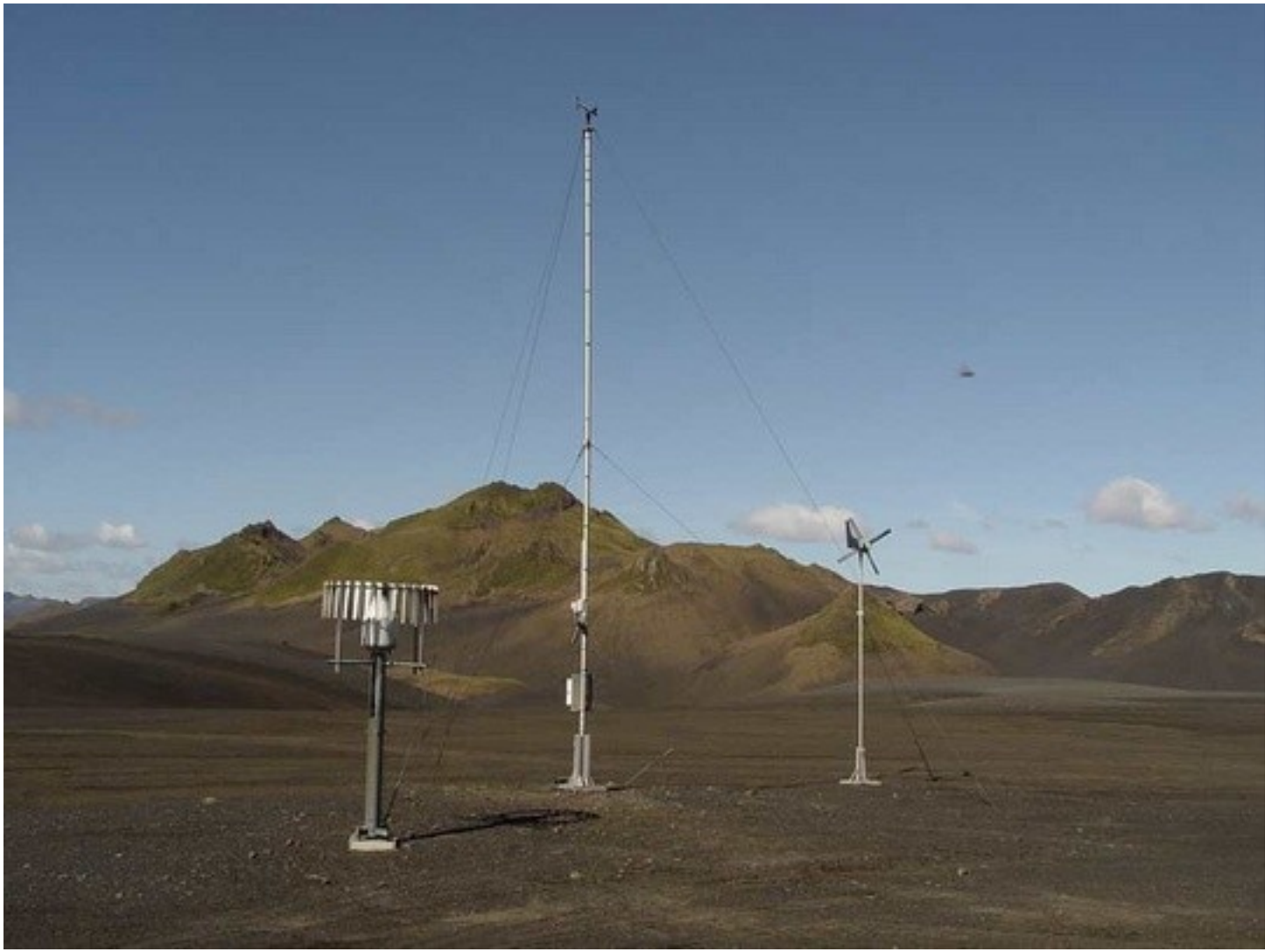
- ...or write your own customized parser providing your data to our system

```
class ExampleObservationParser(SeparatedTextObservationParser):

    def __init__(self):
        super(ExampleObservationParser, self).__init__(separator=',', headers='[^,]*[a-zA-Z]+[^,]*(,.*[a-zA-Z]+.*)*$')
        self.meta['provider_ref'] = 'met.sc'

    def parse_header_line(self, line):
        if not super(ExampleObservationParser, self).parse_header_line(line):
            return False
        if line != '':
            parts = self.separator.split(line.strip())
            self.field_list = [('time', parse_time, parts.index('DD/MM/YYYY HH:MM')), ('station_ref', str, parts.index('Site'))]
            for key_in, key_out in FIELD_ALIASES.iteritems():
                if key_in in parts:
                    self.field_list.append((key_out, float_or_no_data, parts.index(key_in)))
        return True

    def parse_data_line(self, line):
        return super(ExampleObservationParser, self).parse_data_line(line.strip())
```

The QA system of Belgingur



QA system performance for observations from an automatic weather station in Iceland

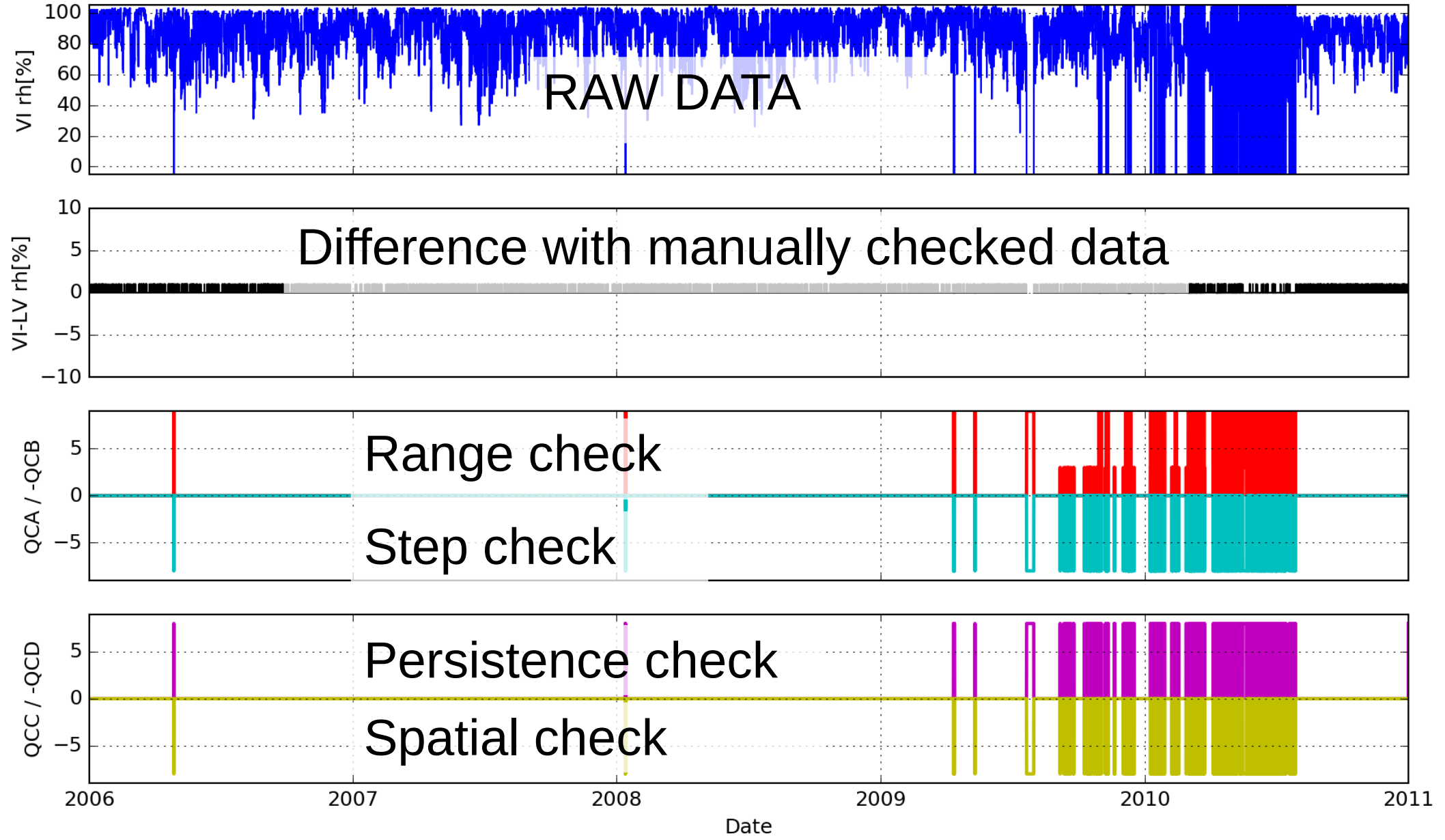


QA system performance

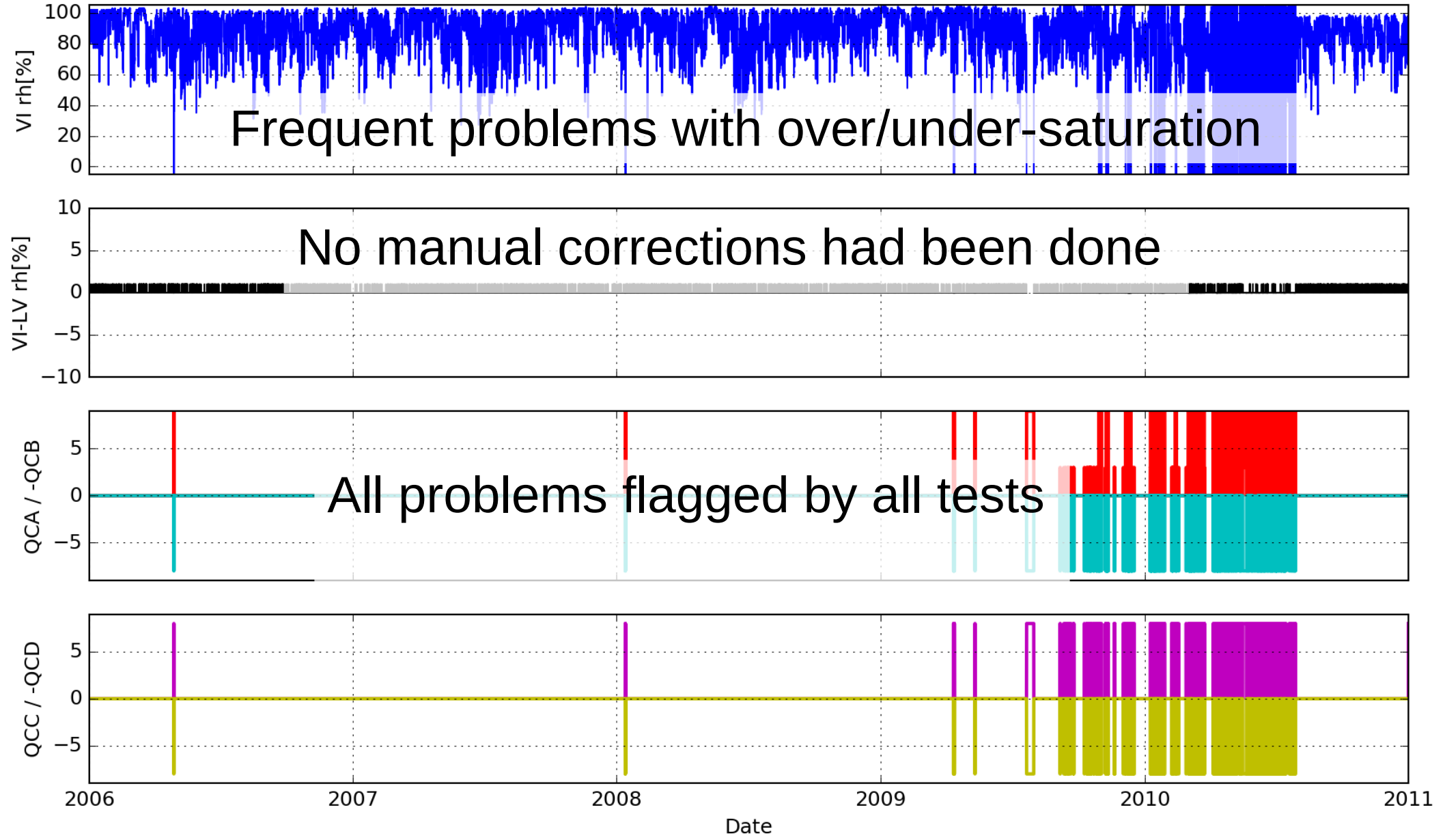
Typical problems in observed data:

- Large spikes in the input data, i.e. in wind speed during lightning activity and bad weather.
- Icing problems on anemometers and thermometers.
- NANs in observed datums, i.e. due to incorrectly registered observations.
- Measurements which are out of the range of physical values, typically associated with poor instruments measurement technique.
- Faulty or failing instruments, causing an error or a drift in the measured variable.

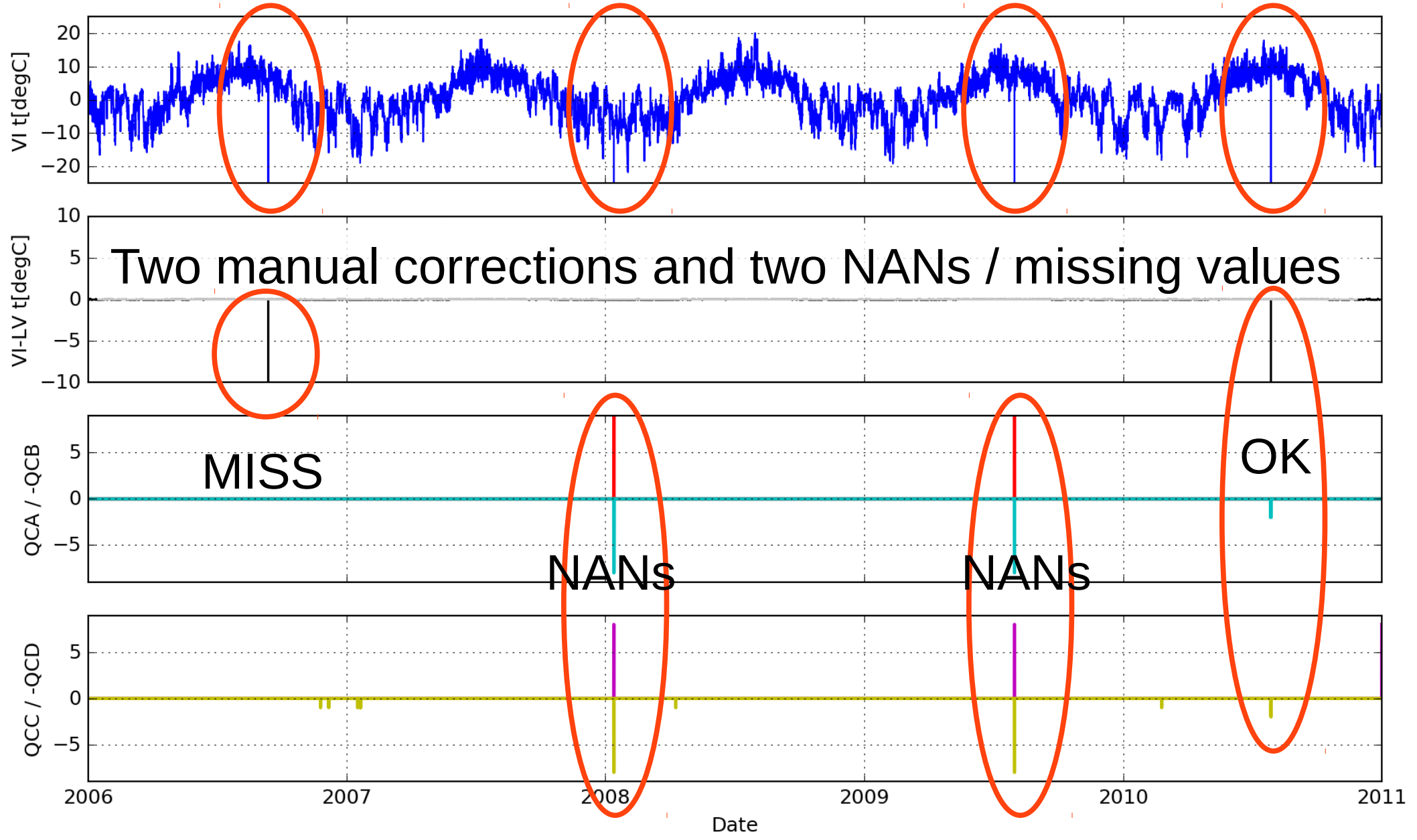
QA system


RAW DATA
Difference with manually checked data
Range check
Step check
Persistence check
Spatial check
VI rh[%]
VI-LV rh[%]
QCA / -QCB
QCC / -QCD
100
80
60
40
20
0
10
5
0
−5
−10
2006
2007
2008
2009
2010
2011
Date

QA system – Relative humidity

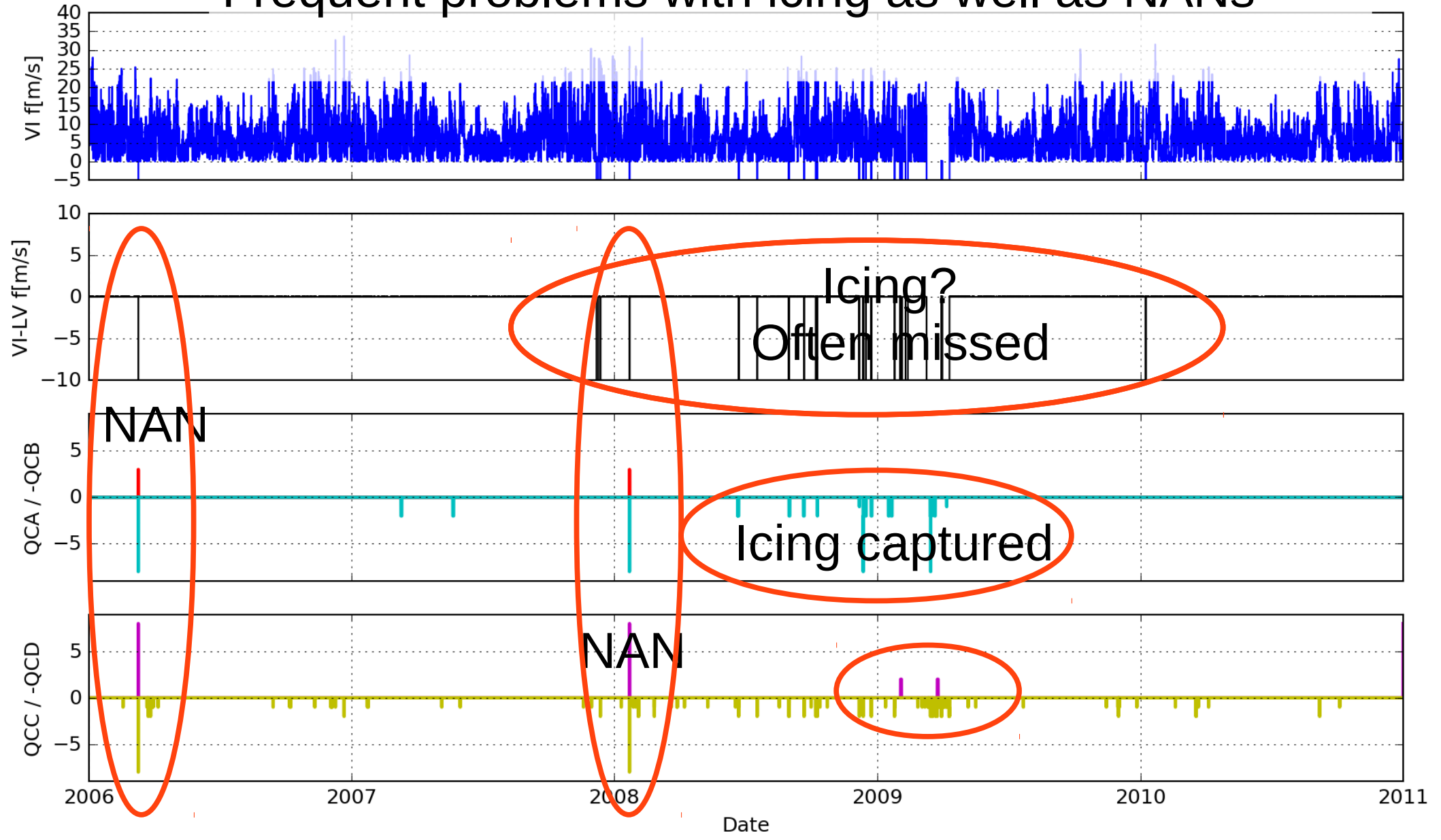


QA system - Temperature

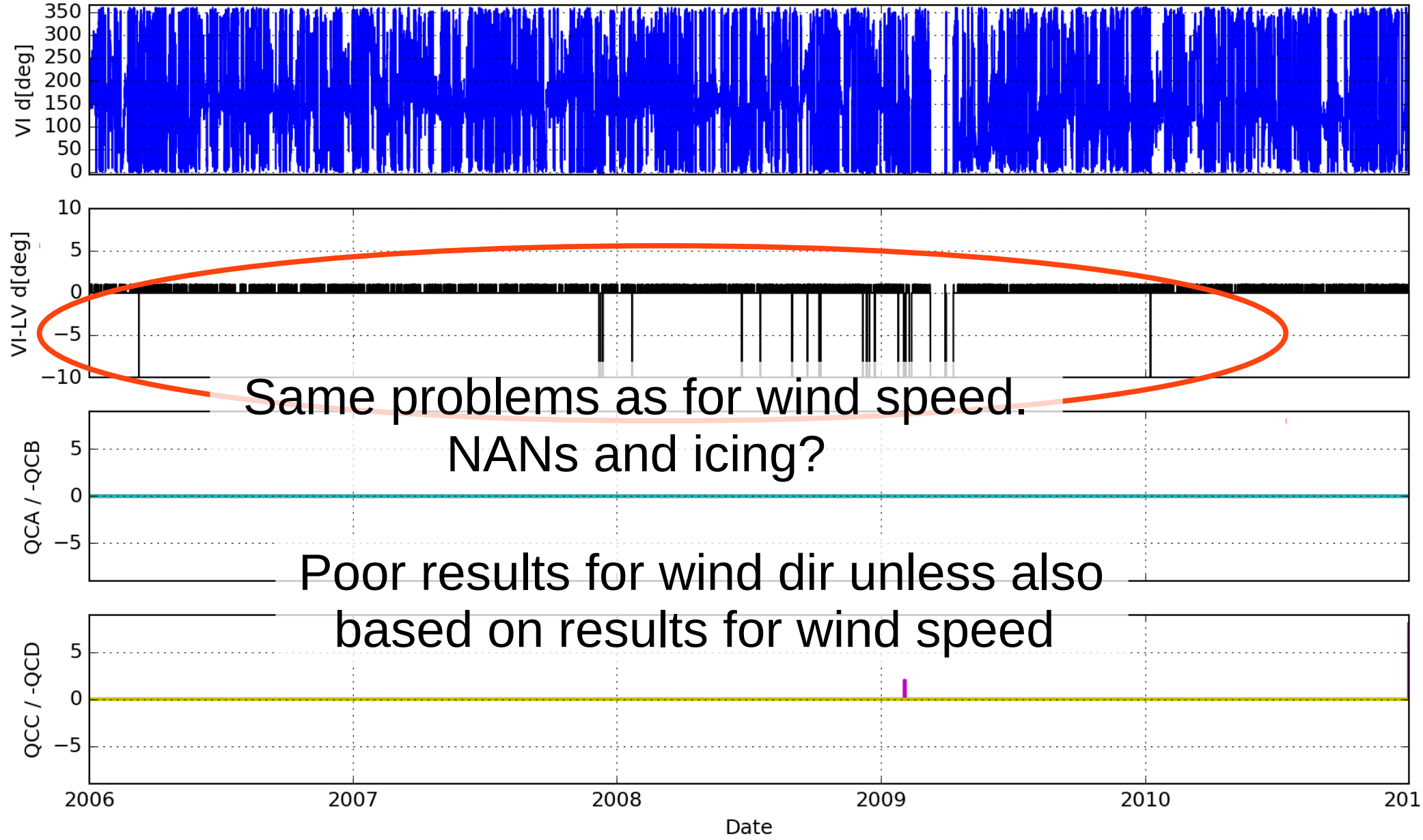


QA system – Wind speed

Frequent problems with icing as well as NANs



QA-system – Wind direction



Same problems as for wind speed.
NANs and icing?

Poor results for wind dir unless also based on results for wind speed

QA system in Iceland

Performance of QA-system was checked for +10 stations in Iceland by comparison of QA-results with manually checked data:

- Most potential errors are caught by QA-system.
- Necessary improvements to persistence check, with regard to anemometer problems (icing), have been identified.
- Wind direction analysis must take into account QA-results for wind speed.
- Spatial test often fails in complex terrain or in regions with sparse stations.
- Missing records (no-observation) can not be flagged.
- Regional and climatologically relevant criteria/thresholds should be found and used in tests.

Further development

- Implement additional means of assessment:
  - analyze wind direction dependent on the wind speed result
  - more flexibility for the persistence check
  - the original system was designed mainly for hourly data, what happens if we put 10-minute data?
  - more conditions on the spatial check

Potential use in Africa

- Speed up the manual inspection of the data
- Spot malfunctioning gauges with the spatial check
- Use in the linear regression point forecast procedure (negatively flagged observations won't be used as input for the method)

Thank you!