# Data flagging workshop

Online – 24 January 2025

ACTRIS CiGas

# Aims

- Homogenization of flags and local event definition within the different NFs

- Suggestion of guidelines (taking into account the specificities of each station)

# Valid or invalid data?

The three main types of flags in EBAS are:

- **V (valid measurement):** Indicate a special condition at the station or while analysing the sample. But instrument has operated correctly, the data are therefore considered valid.

- **I (invalid measurement):** The measurement is invalid and should be excluded from all further data use or processing. Should only be used in data level 0. In data level 1 those data should be excluded

- **M (missing measurement):** Data are missing, either not measured or excluded through data processing (from lev0 to lev1). Those flags may only be used when the value is MISSING.

# NOx Flagging lev 0

| NOx lev0 | | |
|---|---|---|
| 000 | V | |
| 111 | V | Irregular data checked and accepted by data originator. Valid measurement |
| 147 | V | Below theoretical detection limit or formal Q/A limit, but a value has been measured and reported and is considered valid |
| 559 | V | Unspecified contamination or local influence, but considered valid |
| | | |
| 686 | I | Invalid due to zero check. Used for Level 0. |
| 687 | I | Invalid due to span check. Used for Level 0 |
| 699 | I | Mechanical problem, unspecified reason |
| | | |
| 999 | M | Missing measurement, unspecified reason |

Harmonization EBAS flags

Suggestion short list of flags by DC:
- **111** (V) in all levels for **irregular** data checked and accepted by data originator (example: wildfire) for all templates in EBAS. *Open discussion at the end*

- **699** (I; mechanical problem, unspecified reason) for level 0

https://ebas-submit.nilu.no/templates/NOx/lev0
https://ebas-submit.nilu.no/templates/NOx/lev1
https://ebas-submit.nilu.no/templates/NOx/lev2

# NOx Flagging lev 2

| NOx lev2 | | |
|---|---|---|
| 000 | V | |
| 111 | V | Irregular data checked and accepted by data originator. Valid measurement |
| 147 | V | Below theoretical detection limit or formal Q/A limit, but a value has been measured and reported and is considered valid |
| 390 | V | Data completeness less than 50% |
| 559 | V | Unspecified contamination or local influence, but considered valid |
| | | |
| 999 | M | Missing measurement, unspecified reason |

# Flags in EBAS commonly used in CiGas / VOC

| VOC, NMHC_lev0, OVOC_lev0 | | |
|---|---|---|
| 000 | V | |
| Add 111? (new) | | Irregular data checked and accepted by data originator. Valid measurement |
| 147 | V | Below theoretical detection limit or formal Q/A limit, but a value has been measured and reported and is considered valid |
| 559 | V | Unspecified contamination or local influence, but considered valid |
| 653 | V | Sampling period shorter than normal, considered representative. Observed values reported |
| 654 | V | Sampling period longer than normal, considered representative. Observed values reported |
| 683 | I | Invalid due to calibration. Used for Level 0 |
| 684 | I | Invalid due to zero/span check. Used for Level 0. |
| 685 | I | Invalid due to secondary standard gas measurement. Used for Level 0. |
| Add 699? | | Mechanical problem, unspecified reason |
| 999 | M | Missing measurement, unspecified reason |

https://ebas-submit.nilu.no/templates/VOC/NMHC_lev0
https://ebas-submit.nilu.no/templates/VOC/NMHC_lev2

Harmonization EBAS flags between NOx and VOCs templates (*under discussion*):

683: regular calibrations points with working standard
684 -> 686 (NOx template) zero air measurement
685: laboratory standard, not used as calibration
687: Regular span checks (NOx), target gas

Add 111 (V)? *Discussion at the end*

Add **699** (I; mechanical problem, unspecified reason) for level 0?

# Flags in EBAS commonly used in CiGas / VOC

| VOC, NMHC_lev2, OVOC_lev2 | | |
|---|---|---|
| 000 | V | |
| Add 111? (new) | | Irregular data checked and accepted by data originator. Valid measurement |
| 147 | V | Below theoretical detection limit or formal Q/A limit, but a value has been measured and reported and is considered valid |
| 559 | V | Unspecified contamination or local influence, but considered valid |
| 653 | V | Sampling period shorter than normal, considered representative. Observed values reported |
| 654 | V | Sampling period longer than normal, considered representative. Observed values reported |
| | | |
| | | |
| | | |
| | | |
| 999 | M | Missing measurement, unspecified reason |

https://ebas-submit.nilu.no/templates/VOC/NMHC_lev0
https://ebas-submit.nilu.no/templates/VOC/NMHC_lev2

# CAMS NRT Project: new VOC lev0 template

https://ebas-submit.nilu.no/templates/VOC/new_NMHC_lev0

Flag 686 (Invalid due to zero check. Used for Level 0) and
Flag 687 (Invalid due to span check. Used for Level 0.)
Flag 699 (Invalid due to mechanical problem)

For next submission for NOx, correct flags should be used for calibration, target tank, blank (under testing for VOCs for level 0 submission)

# Flag discussion within CiGas community

- **999** – M – Missing
- **000** – V – Valid
- **559** – V – Local event

# What/how do we flag at the moment?

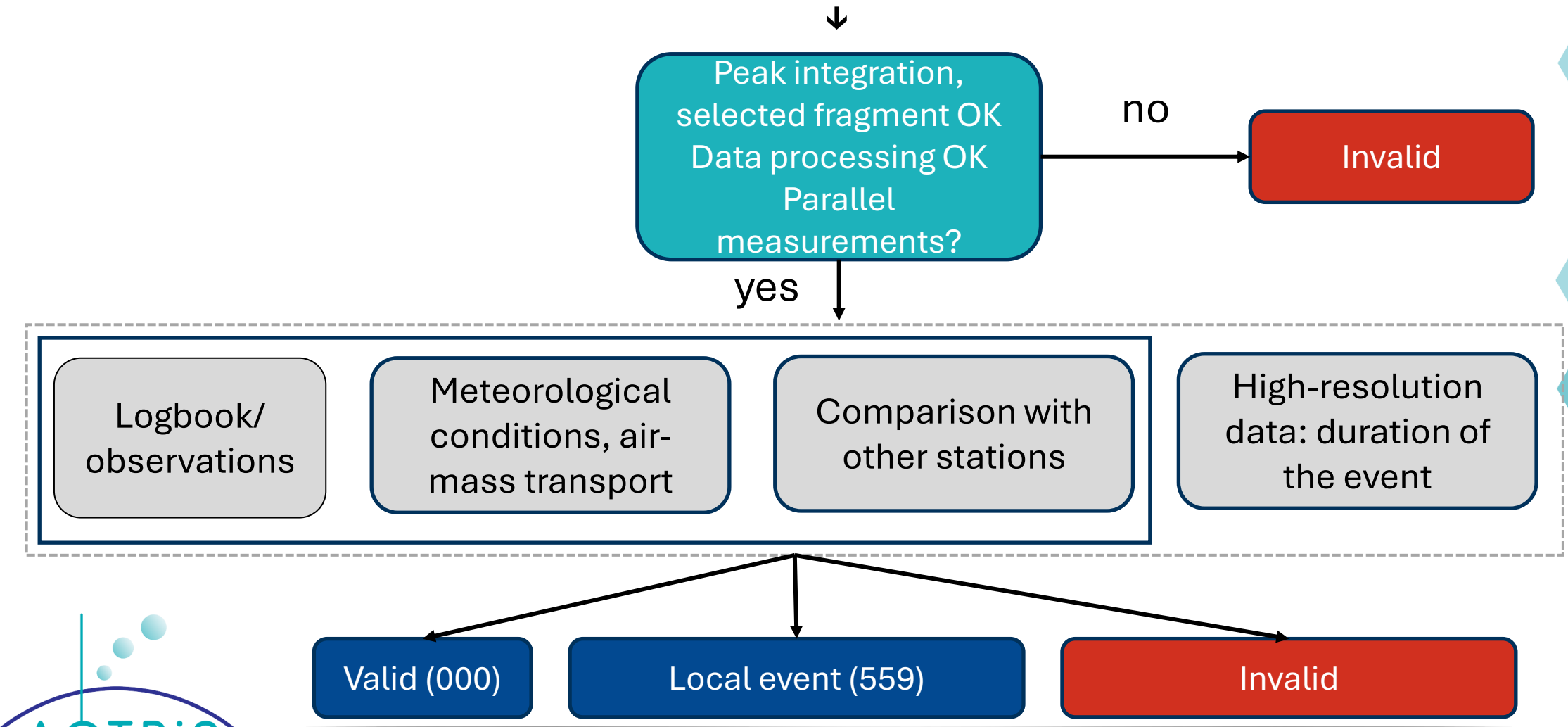
Mostly only apparent/visible outliers (**subjective** aspect, up to the operator)

- Measurement and data processing **OK**
- **Logbook** entries as indication
- Indications with **meteorological conditions**, **air-mass transport**: indicators with $O_3$, wind speed/direction, footprints
- Flag the concerned substances or all substances (if an event can be determined, compounds with **similar sources** also involved? **Contamination**?) *open discussion at the end*
- **Distance** from the station (local ≠ regional event)
- Outlier statistic/peak event tool for high-resolution data (1-min NOx/PTR) under testing (HPB)
- Flag as less as possible, if for a peak event, measurement OK, contamination not identified, use 000

→ At the end, the operator has the last word

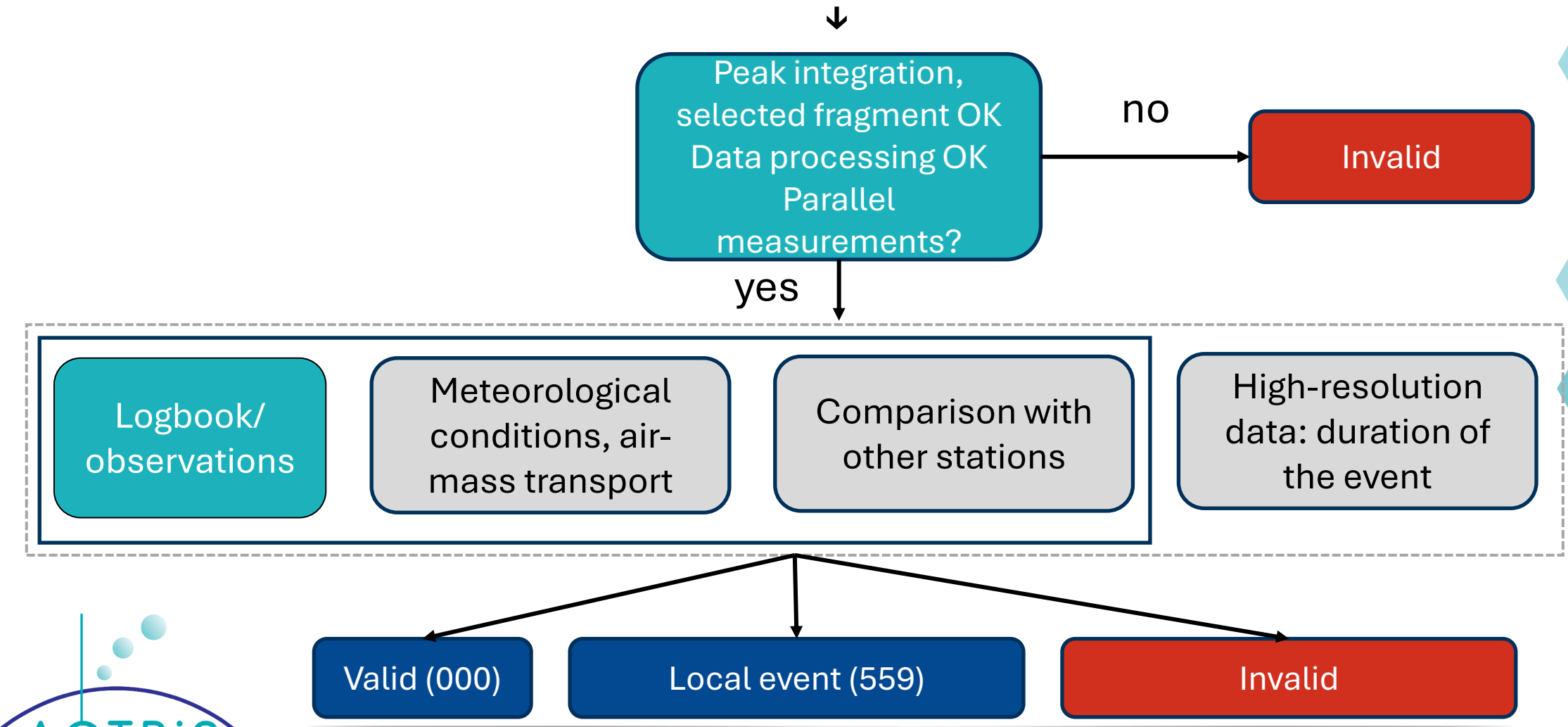
# Draft for guidelines on local event flagging

Unexpected **high concentration** for a VOC over a short period of time

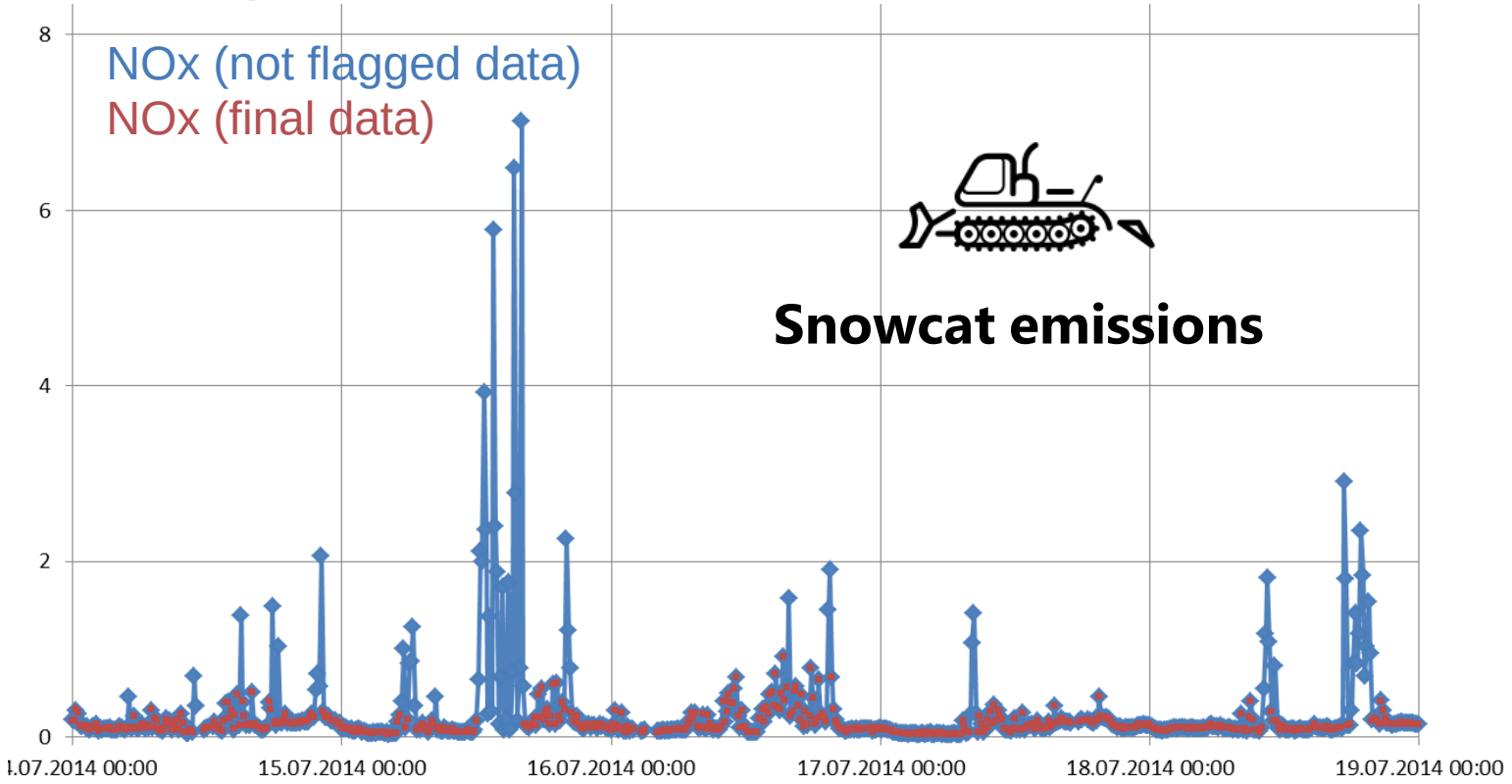


# Draft for guidelines on local event flagging

Unexpected **high concentration** for a VOC over a short period of time



# Logbook/observations: one-time known event (Jungfraujoch)






**Snowcat emissions**

→ Chromatograms OK

→ Known event

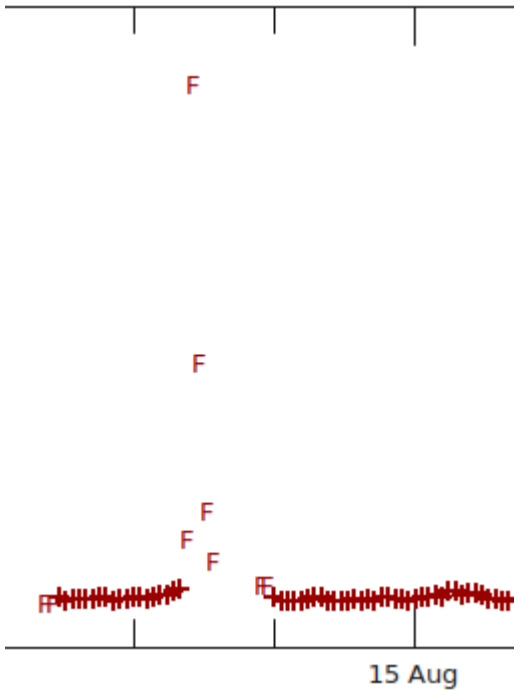
Decision from data provider: flagged as **invalid** as it is not representative for the station



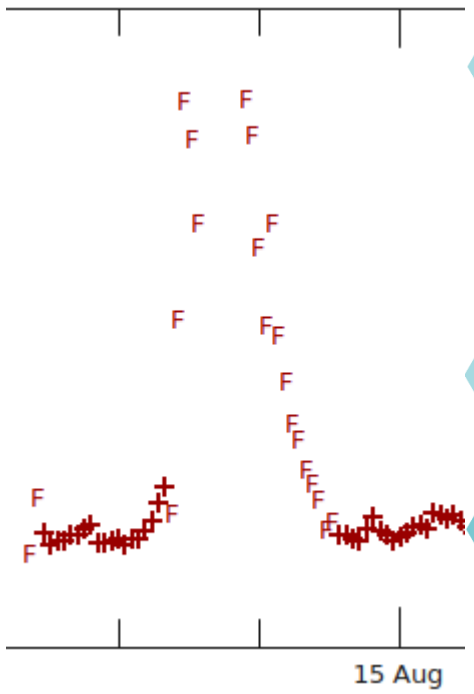
*Tennis show match (2014)*

# Logbook/observations: one-time known event

Fire extinguisher training at Zürich Kaserne (Aug 2024)




Ethylbenzene
100-fold higher




methyl-tert-butyl-ether
40-fold higher

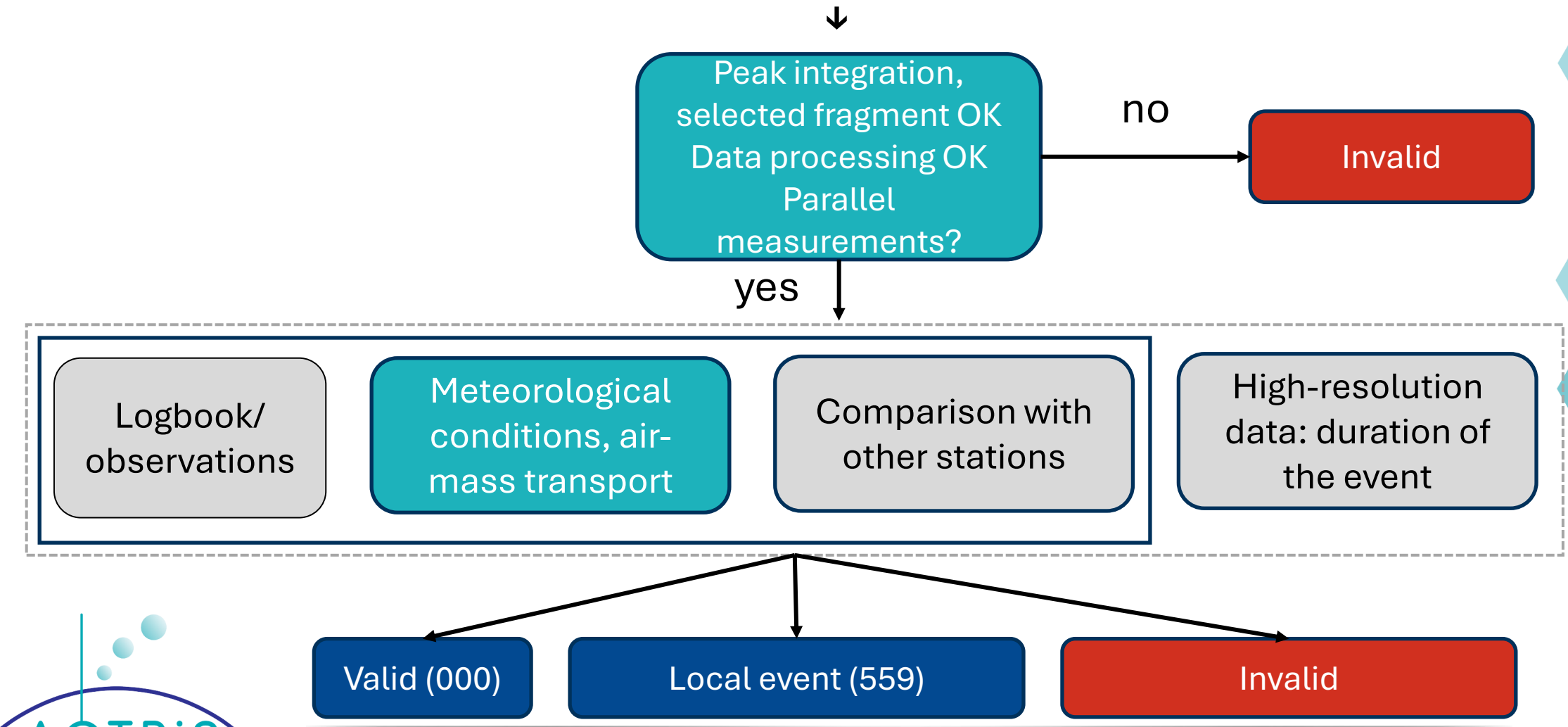
→ Unexpected high concentrations coupled with high CO values

→ Contamination identified

Outcome: flagged as **invalid**, considered as contamination by data provider

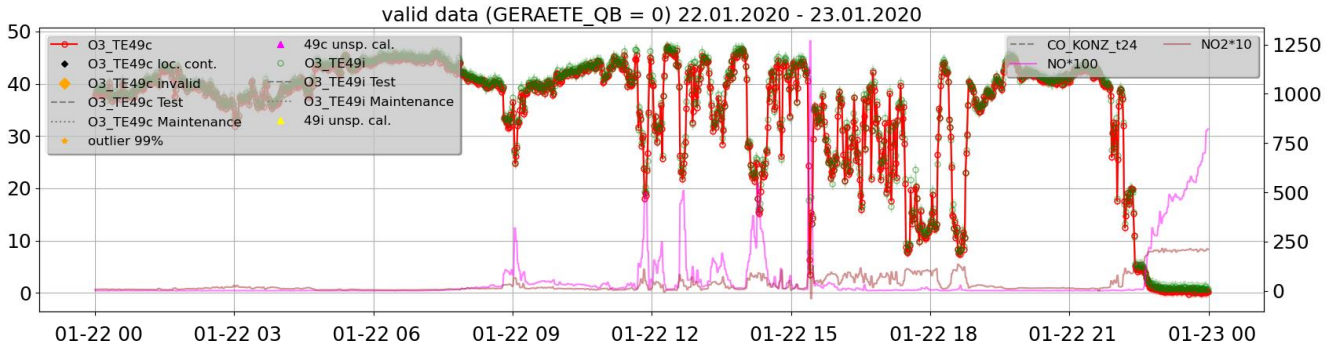
# Draft for guidelines on local event flagging

Unexpected **high concentration** for a VOC over a short period of time



# Typical reasons and occurences of NOx and/or O3 peaks at Meteorological Observatory Hohenpeißenberg, flagged as 559?

- ➔ peaks caused by changing wind directions / air masses **No**
- ➔ height of boundary layer oscillating at altitude of observatory **No**
- ➔ ... and lasting „dipping“ into the boundary layer **No**





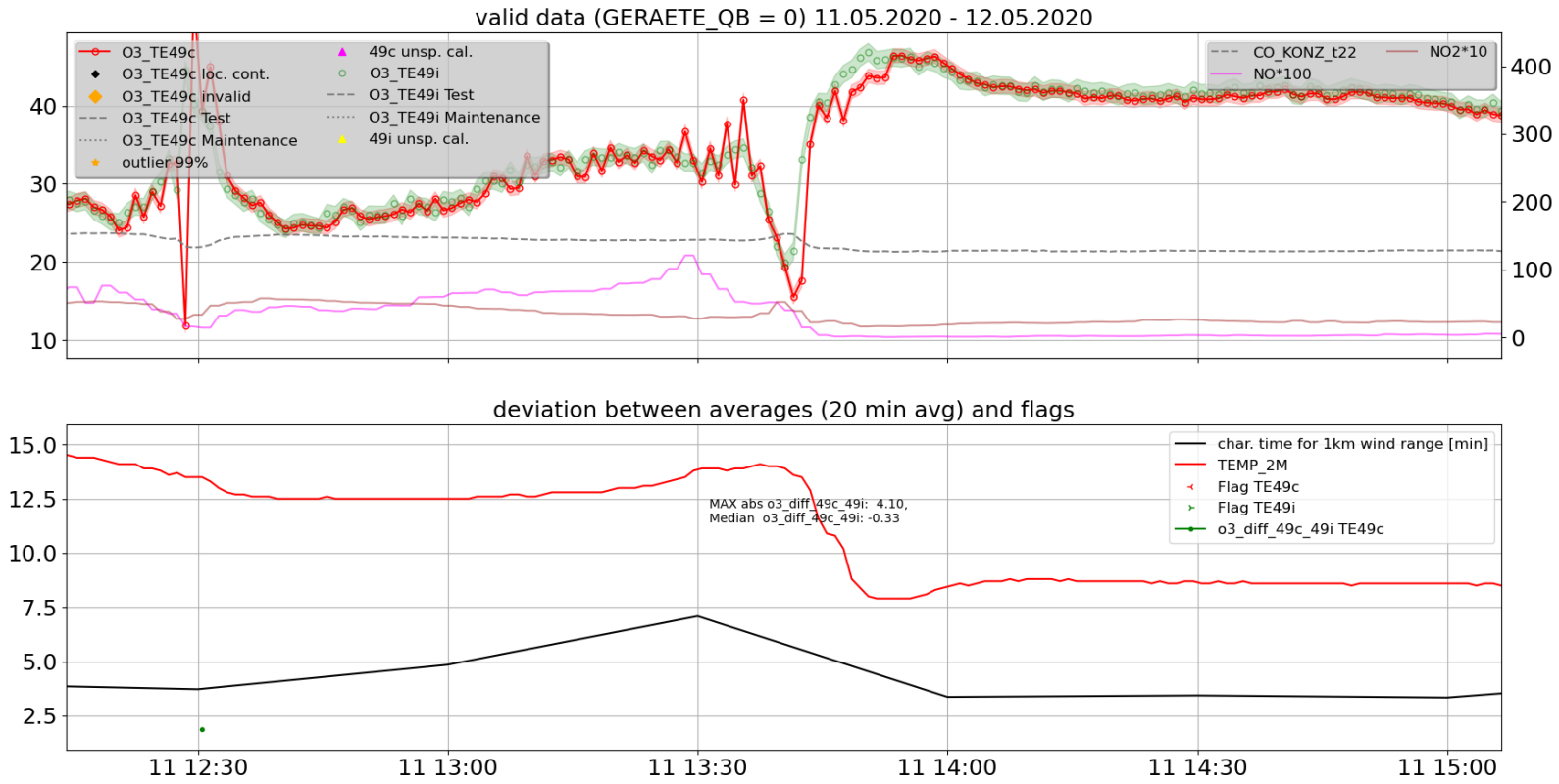
- ➔ accumulation during low wind speeds, *„broad“* peaks
- ➔ single or groups of *„narrow“* peaks **Yes**

**depends, if local sources/peaks can be identified**

- ➔ local emissions which can be attributed to documented/observed sources **Yes**

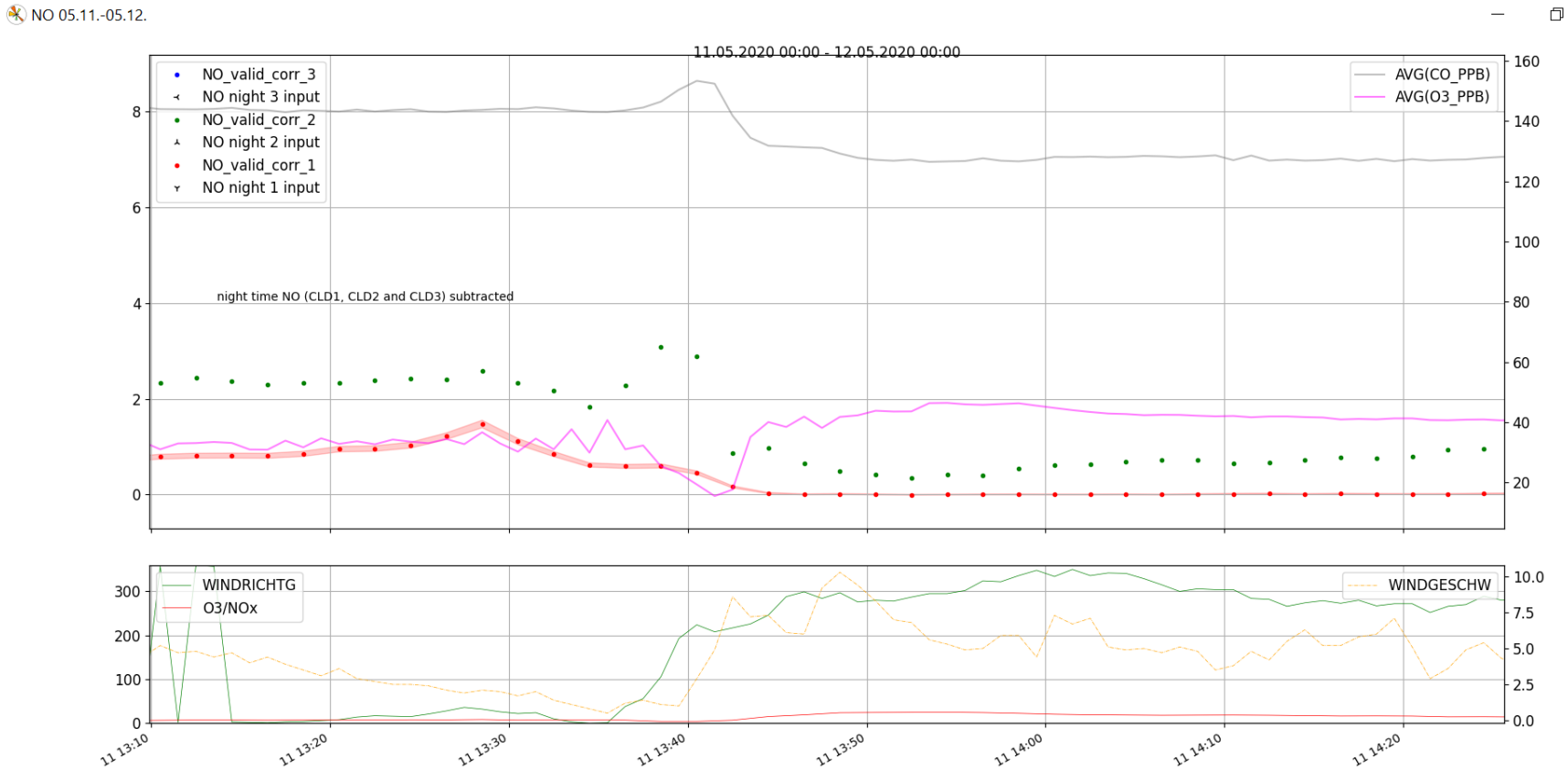
*definition of „broad“ and „narrow“ see below*

# example negative $O_3$ peak: more probably change of air masses than local influence



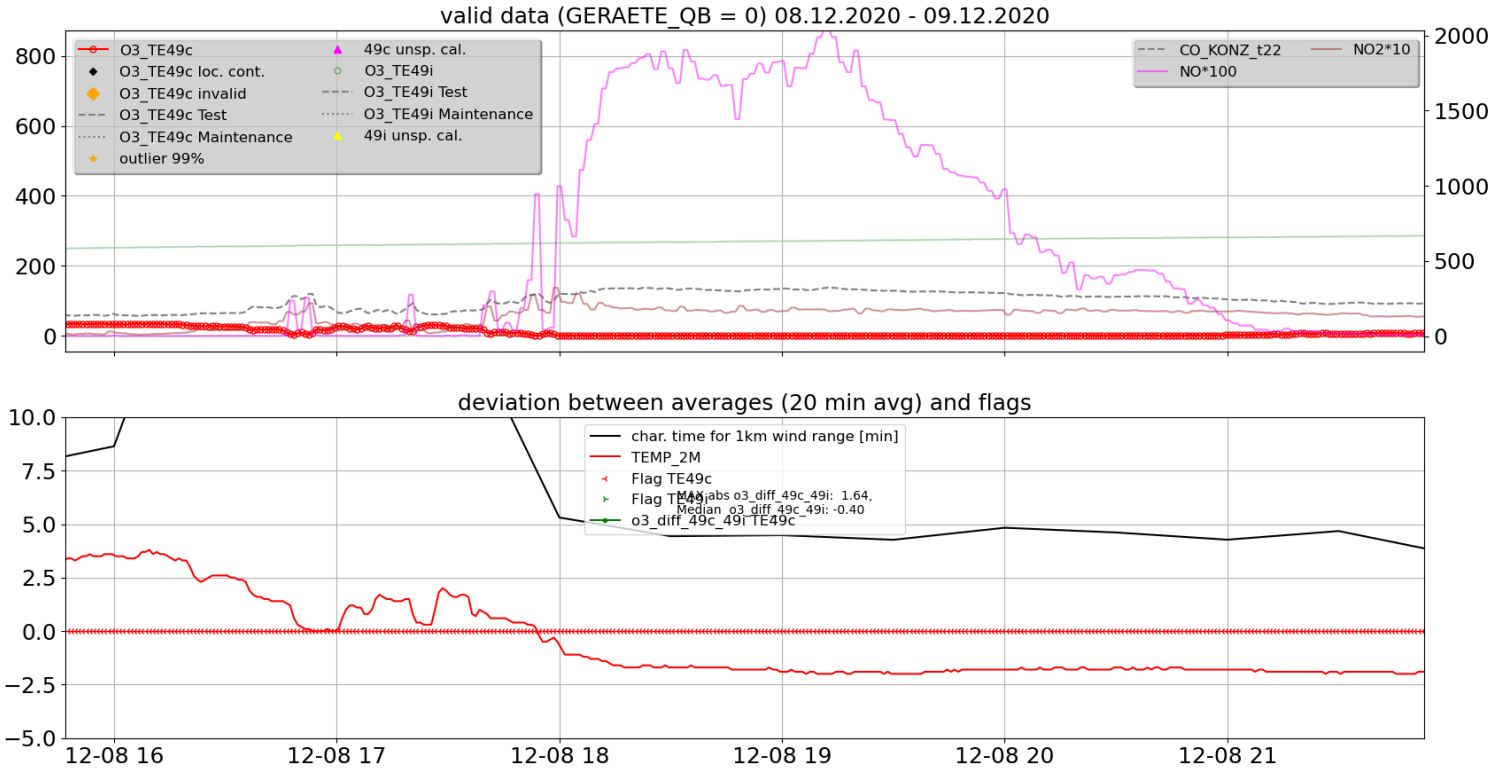

- upper panel: decreasing $O_3$ (red line) and increasing $NO_x$ (magenta, grey)
- lower panel, temperature changes (red line)

# example negative $O_3$ peak: more probably change of air masses than local influence




- same event, different visualization with wind direction (lower panel, green line)
(in this case negative) peak is caused by alternating winds → **no flag 559**

# example persistent „dipping“ into the inversion layer

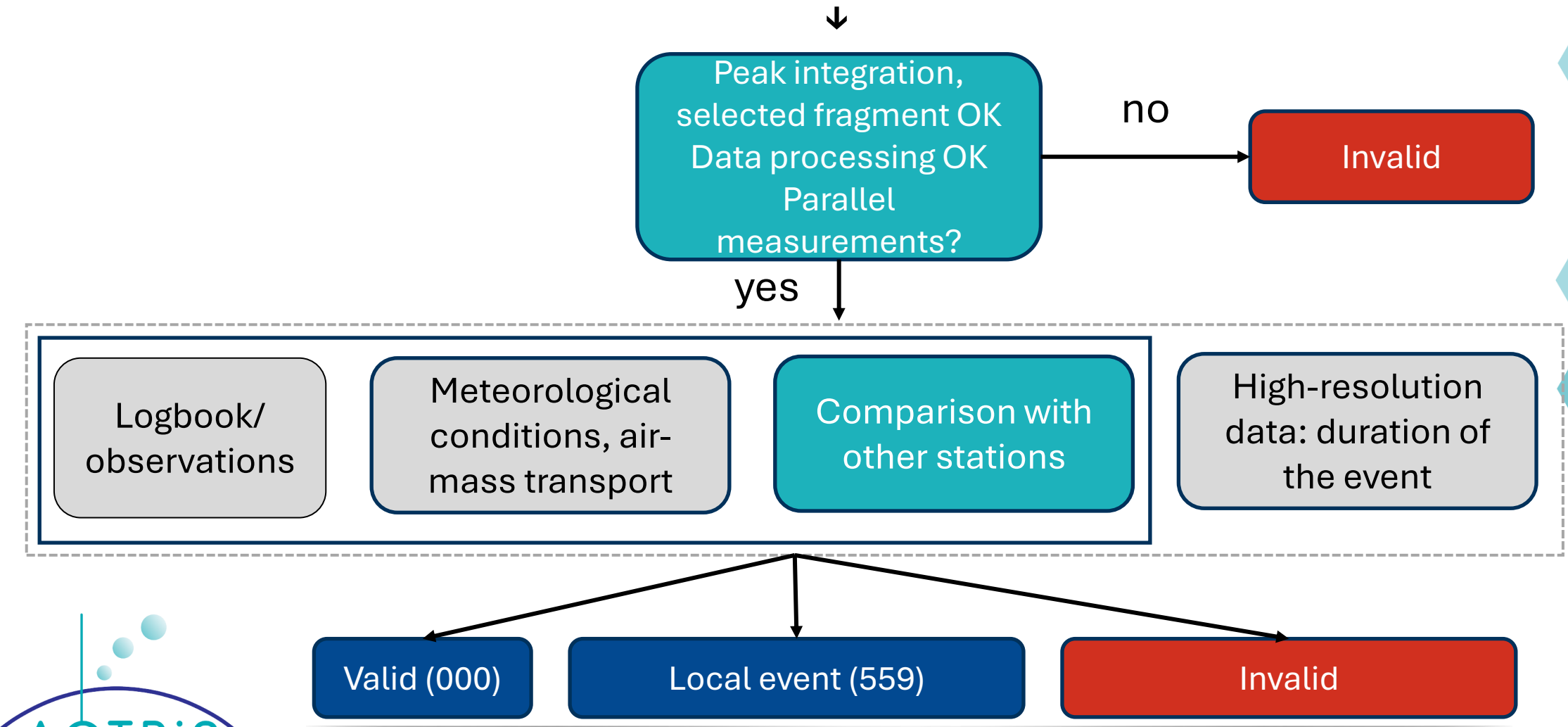



- upper panel, magenta, grey and dashed line: increasing $NO_x$ and CO
- lower panel, temperature decreases (red line), webcam shows rising fog (not shown)

peak is caused by meteorological characteristics of the station → **no flag 559**

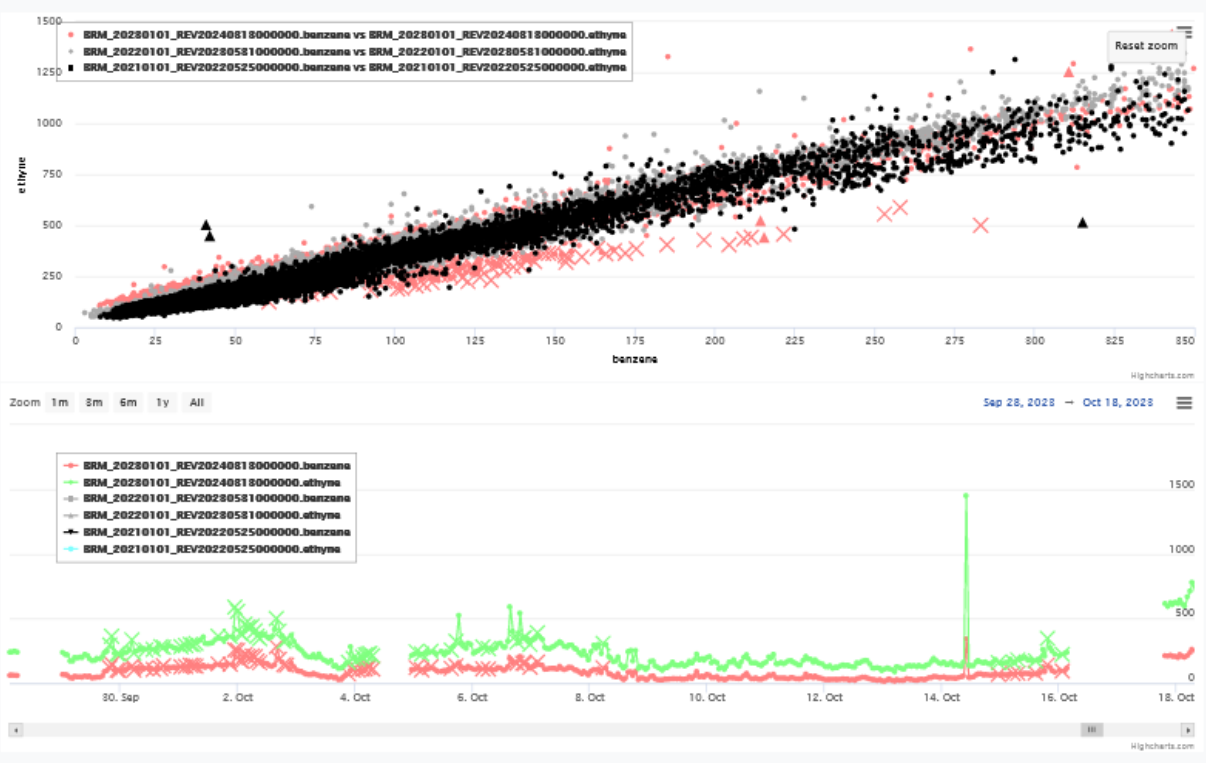
# Draft for guidelines on local event flagging

Unexpected **high concentration** for a VOC over a short period of time



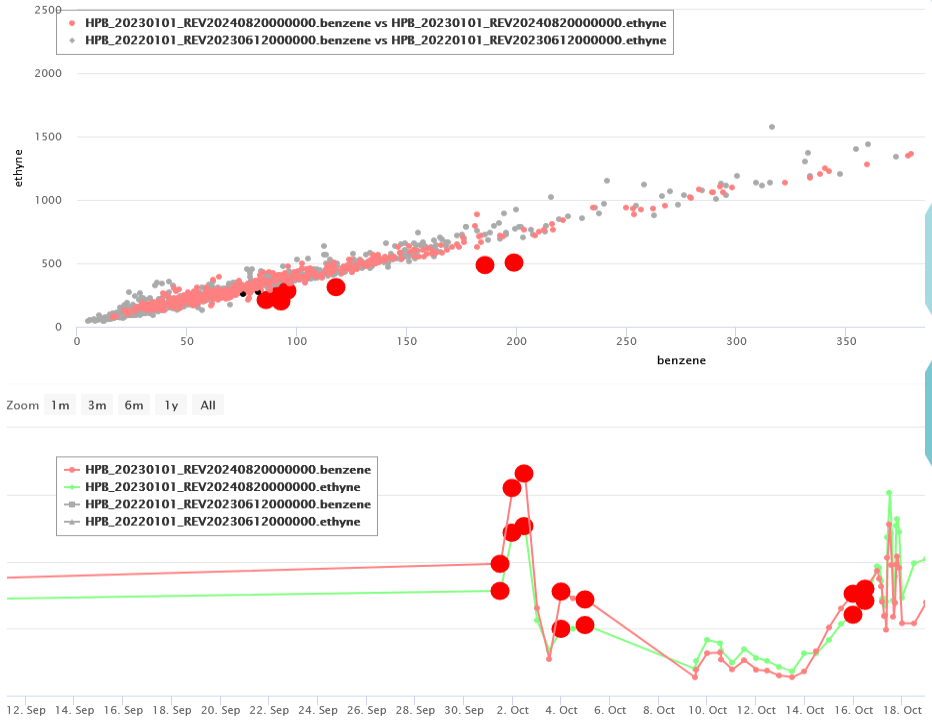
# Untypical correlation: ethyne / benzene

## Beromünster



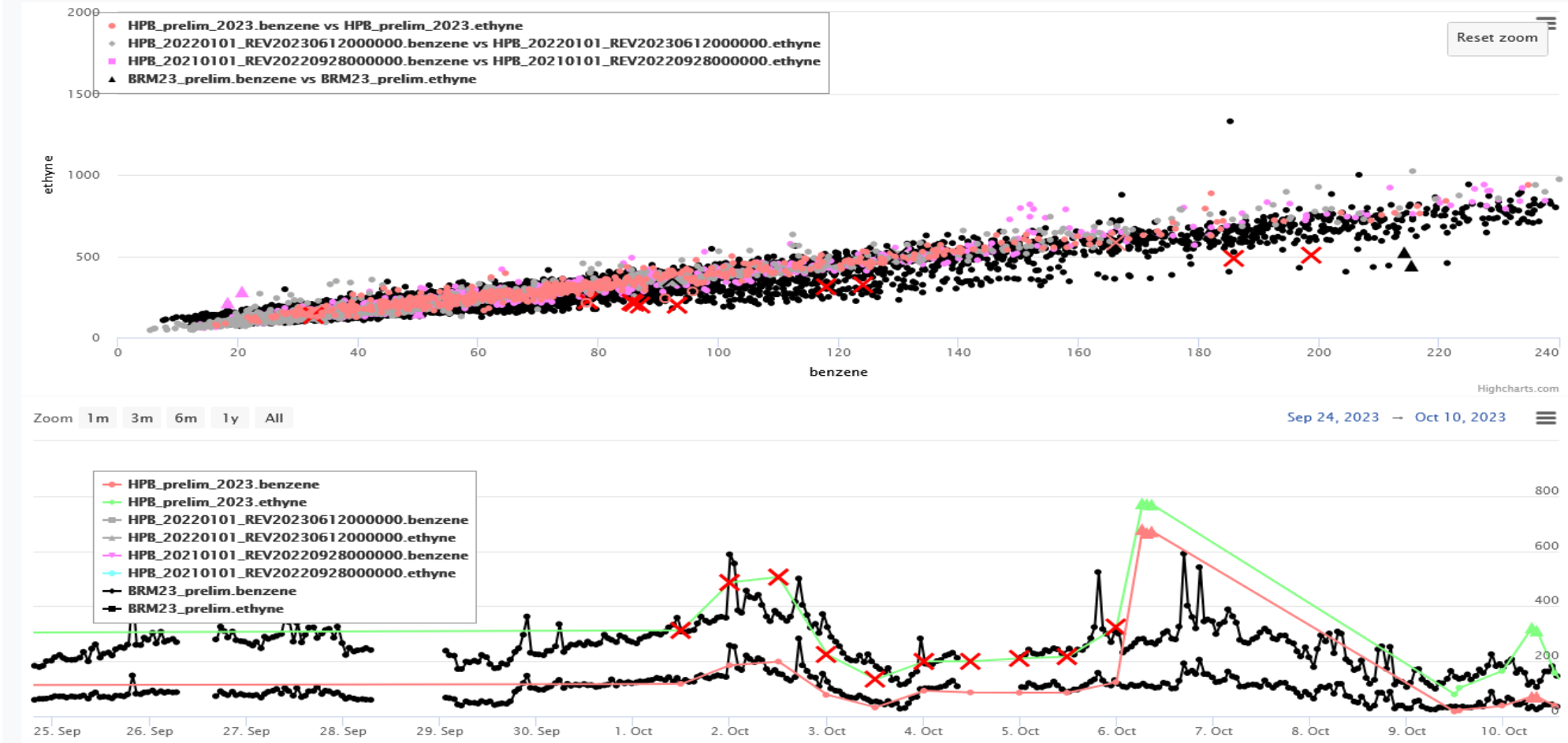

Decision at Beromünster → **invalid**

## Hohepeissenberg



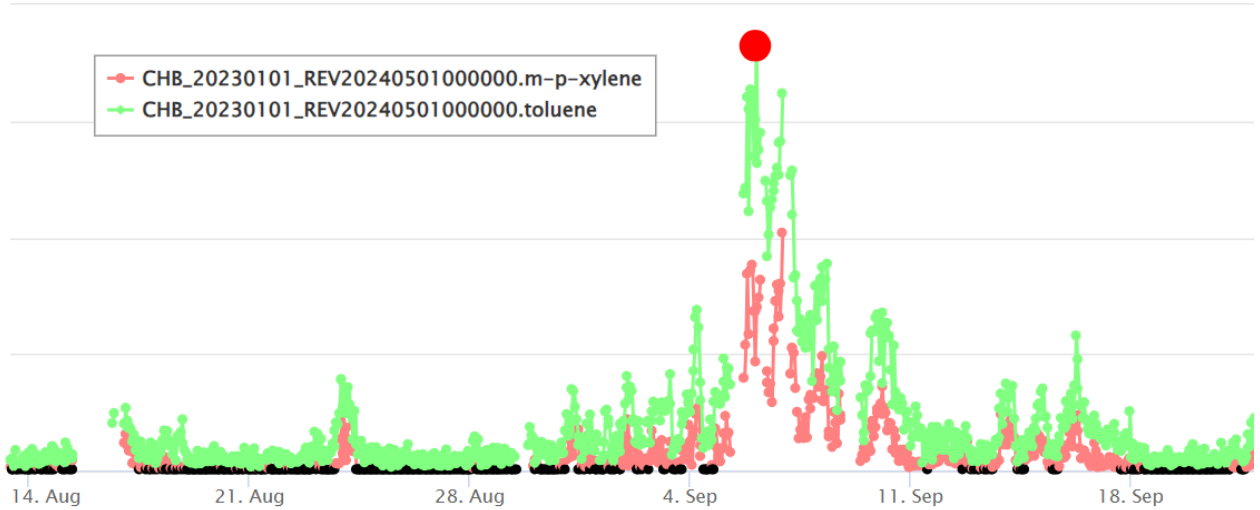

Decision at Hohepeissenberg → **invalid**

# Untypical correlation: ethyne / benzene



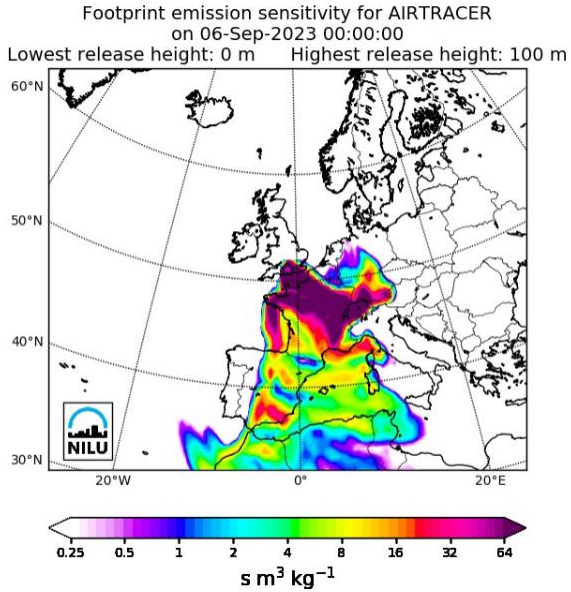

Happens at the same time in two different locations with two differnt GC systems → **valid!**

# Comparison with different stations

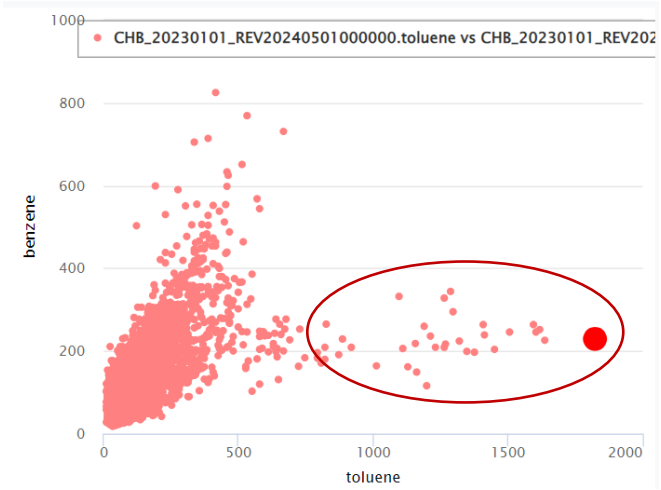



*High concentrations observed for toluene, m-p-xylene in September*

→back-trajectories do not indicate a local event but rather a regional event



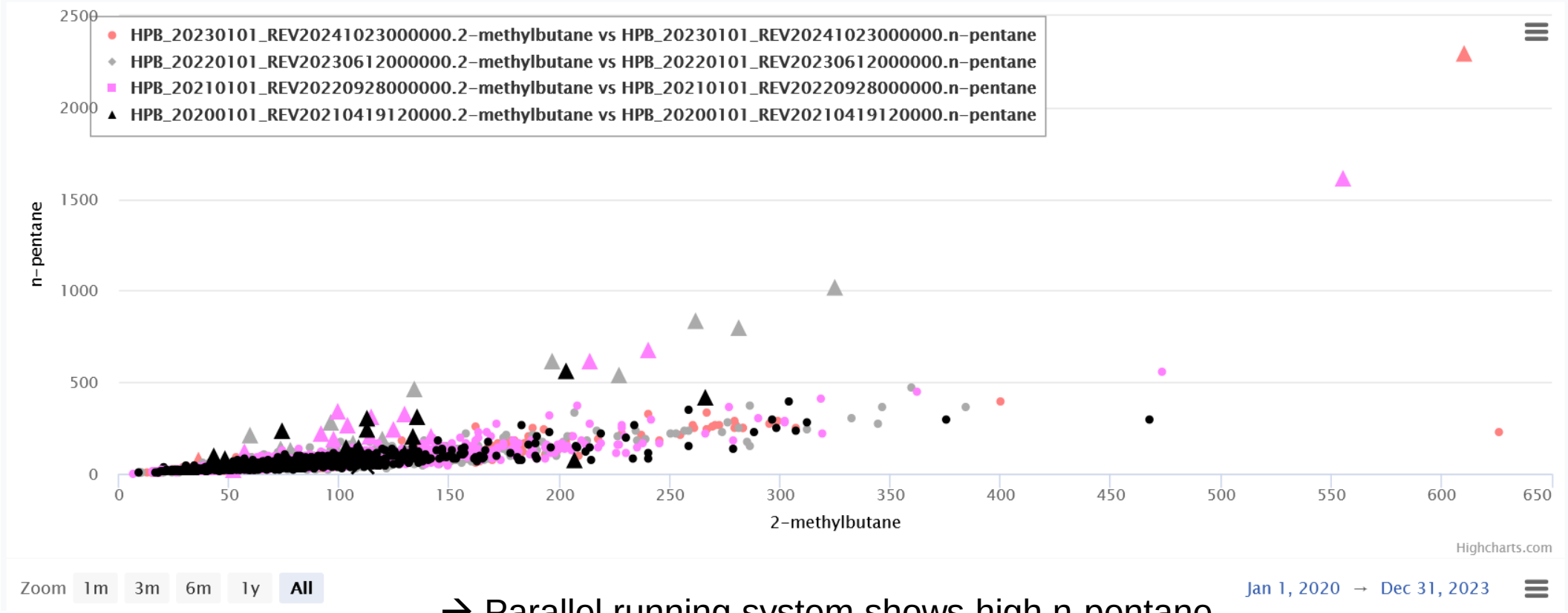

→ Observed at other UK stations




→ confirmation of ozone peak pollution at this period in UK (local ≠ regional)

Outcome: flagged as **valid** as not considered as local event but regional event

# Comparison with parallel measurements: n-pentane singularities at HPB

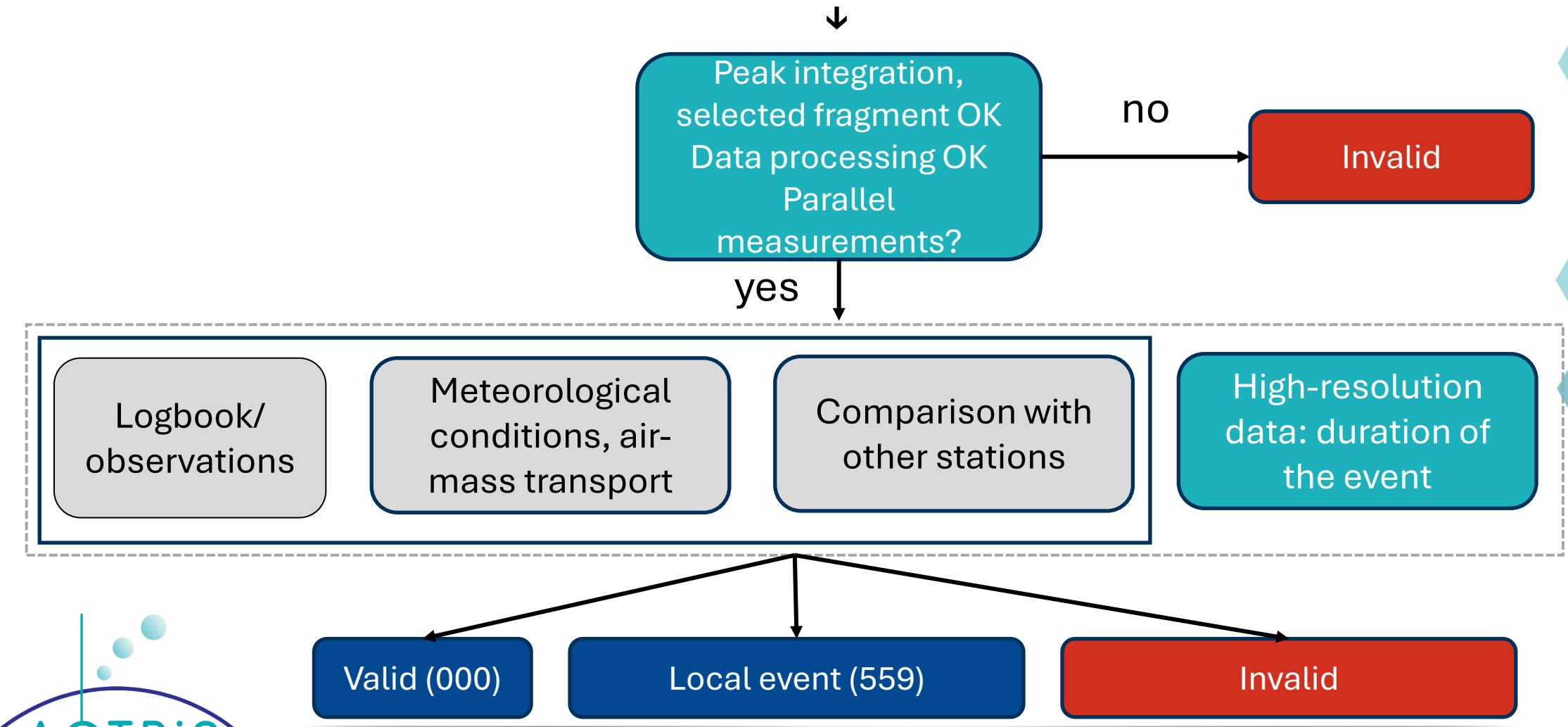



→ Parallel running system shows high n-pentane
→ No correaltion with meteo (e.g. wind direction)
→ No reportings of activity at HPB
→ Observed over many years

Outcome: **As valid local event (559) flagged**

# Draft for guidelines on local event flagging

Unexpected **high concentration** for a VOC over a short period of time

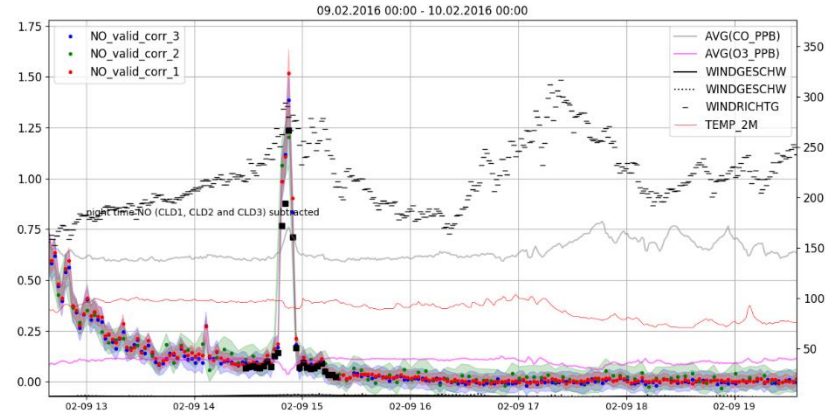


# using a characteristic time scale

- constrain the meaning of „local“
- e.g. literature of topographical climatology: 1-10 km, 1 min – 1 hour
- typical way to notice local influences in time series is to find peaks, which are usually not there
- need to detect peaks
- scale estimate, dimensions: $T = \frac{L}{[v]}$
- **acceptable** peak width for 559, e.g.:

$$\Delta t_{local} = \frac{\Delta x}{v_{wind}} = \frac{1\ km}{v_{wind}}$$

- $\Delta t_{local}$ can be set
  - dynamically as $f(v)$
  - as fixed value from statistics*
  - depending on surrounding sources



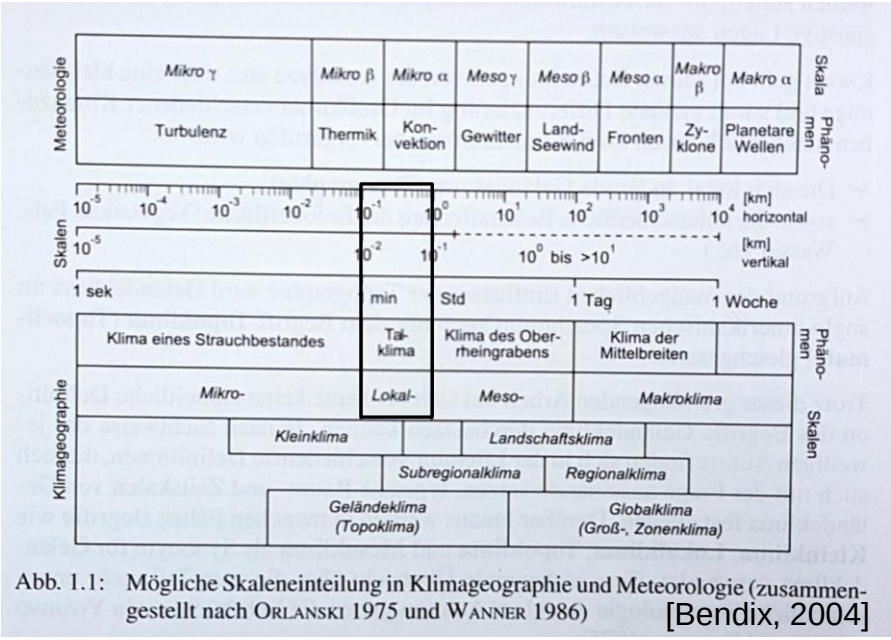



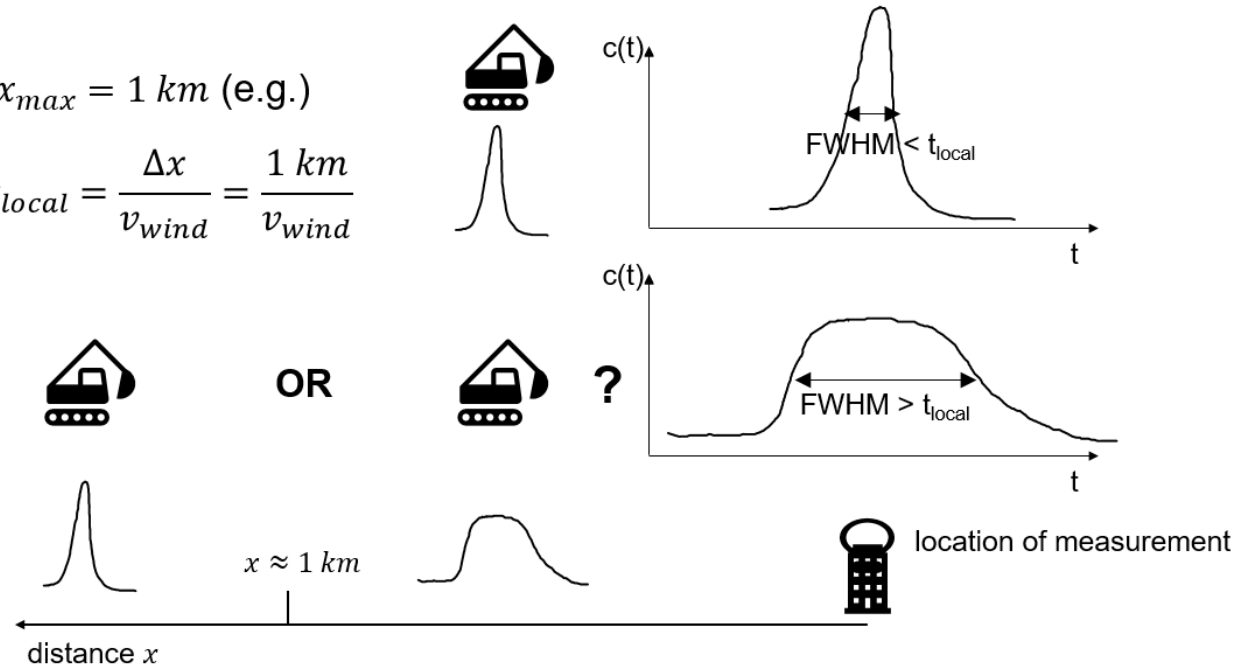

Abb. 1.1: Mögliche Skaleneinteilung in Klimageographie und Meteorologie (zusammengestellt nach ORLANSKI 1975 und WANNER 1986) [Bendix, 2004]

*e.g. for MOHp 2024 98%-percentile of is $\Delta t_{local}$ 21 min

# find peaks, then use additional information

$\Delta x_{max} = 1\ km$ (e.g.)

$$\Delta t_{local} = \frac{\Delta x}{v_{wind}} = \frac{1\ km}{v_{wind}}$$




*„narrow“* peak, 559!

*„broad“* peak, 559?

- ➔ try to find additional information for peaks broader than $\Delta t_{local}$ to identify them as „559“ as well:
  - ➔ log book entries / observations
  - ➔ additional measurements: CO, $O_3$, aerosols, other if available
  - ➔ high $NO/NO_2$ or $NO/O_3$ ratio (age of air mass)
- ➔ flags 559 can be transferred from $NO_x$ to $O_3$ time series or vice versa
- ➔ Outlook: compare flagging results with Anja‘s VOC flagging for 559
- ➔ Outlook: testing automated routines for finding peaks: use it as support or substitute for a part of manual flagging?

# On going

- Harmonization of peak events within one station (HPB, EMEP Campaign)
- Test for peak finder using Python (HPB, NOx)

# Open discussion

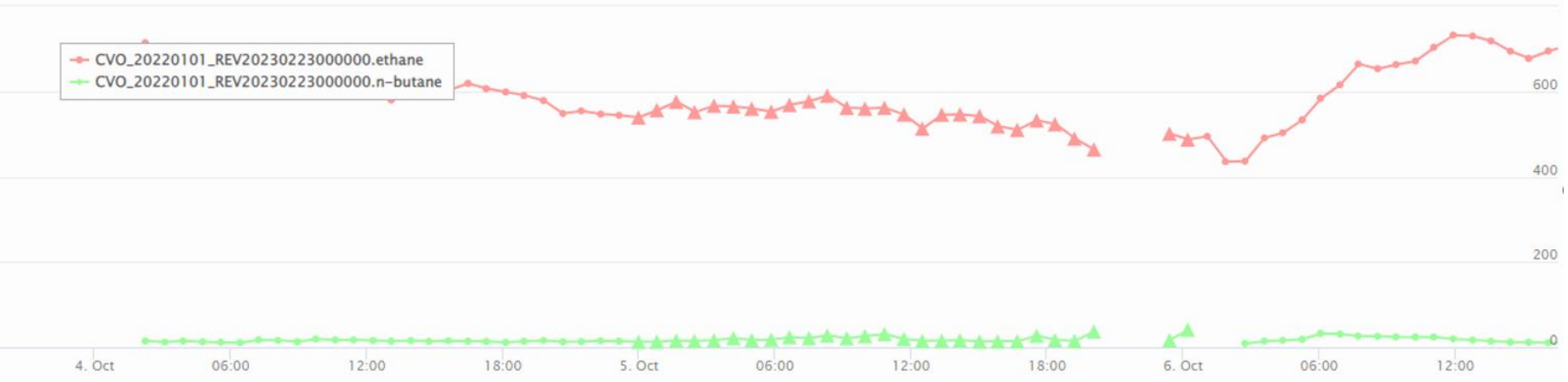
- Logbook entries indication possible contamination (BBQ, people around the station) but no apparent outliers observed – Automatic flagging as locally contaminated or invalid?
- Flag 111 (Irregular data checked and accepted by data originator. Valid measurement)
- How much do we flag, only affected substances or the complete sample? Should the entire sample flagged, especially for VOCs?
- Objective/automatic flagging versus subjective/manually flagging
- Event log book in ACTRIS, is it required? Know-how transfer and sustainability of the station knowledge and supporting of QA/QC process

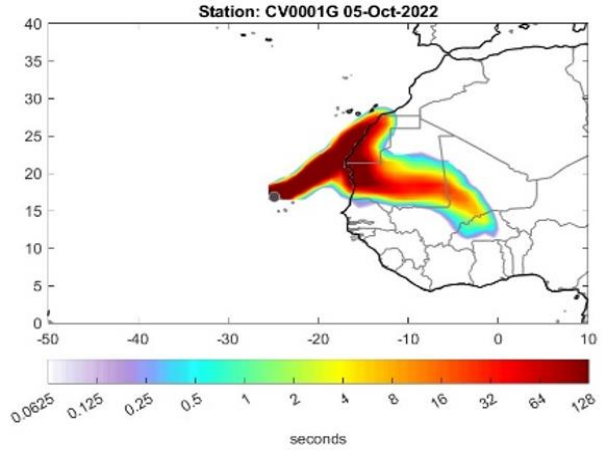
# Next submission

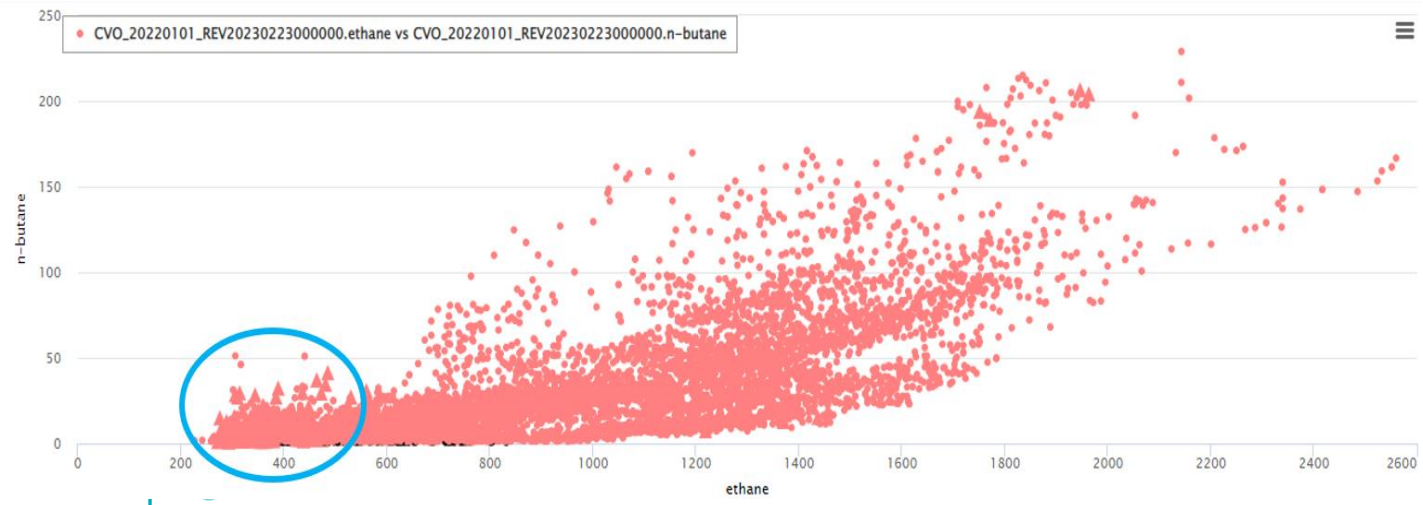
- NOx: as usual, watch for flags 68x when submitting lev0
- VOCs: lev2 as usual, lev0 for areas only for CAMS station, für ACTRIS under discussion, 68x can be used.
- **Data submission 15.03.2025**
- **CiGas QA/QC Workshop: tba, beginning of April.** At least one representative of each station, invitation follows


ACTRiS
CiGas

# Automatic flag (559) for low wind speed




## Discussion:

So far subjective flagging with visually outliers

Low wind speed – potentially local event

Automatic flagging: objective but representative?

But footprints show air masses from Africa