Primary author: **Scherrer, Simon C.** (MeteoSwiss - Federal Office of Meteorology and Climatology, Climate Services), simon.scherrer@meteoswiss.ch

Co-authors: Deborah van Geijtenbeek, Claudine Naguel, Mischa Croci-Maspoli and Christof Appenzeller (MeteoSwiss - Federal Office of Meteorology and Climatology, Climate Services)

Abstract ID: 4O4

Towards semi-automation of manual precipitation data quality control at MeteoSwiss using spatial information

At MeteoSwiss several automatic data quality control systems for automated data are operational and incorporated into our data warehouse system (DWH) for several years now. They allow treating automatically retrieved data at high temporal resolution (usually 10 minute) with a high degree of automation using a graphical user interface for interactive correction of data (see companion paper by van Geijtenbeek et al.).

A few surface measurement networks, such as the manual precipitation network with ~330 stations retrieving 1-day precipitation sums are still treated manually and create a substantial workload which should be reduced. We present a semi-automatic approach that incorporates automatic spatial interpolation and empirical thresholds to flag suspicious values which are then evaluated by expert knowledge. It is shown that although there are several problems with the interpolation in the high Alpine terrain due to spatial density of the network, the manual workload can be reduced substantially by the new tool. In the lowlands the method is very reliable. In more complex terrain, such as the Swiss Alps, or convective cases, expert knowledge, using additional local information about the station, is still very helpful to decide whether an automatically identified suspicious value is indeed wrong and needs to be corrected. Therefore a fully automated data quality control is not recommended especially as long as the measurements are not fully automated and difficult to detect shifts, i.e. wrong allocations of date to a measurement, can occur.

There are several applications of the above tool such as the testing of historical data, e.g. to find “potential” values that are saved as 1-day sum but are several day-sums in reality. Finding and removing these wrong record values can be crucial for a proper extreme value analysis.

Further refinement of the interpolation method (especially for sparse measurement networks) and testing of less empirically based flagging thresholds are planned in the future.

Schweizerische Eidgenossenschaft
Confédération suisse
Confederazione Svizzera
Confederaziun svizra

Federal Departement of Home Affairs FDHA
Federal Office of Meteorology and Climatology MeteoSwiss

Towards semi-automation of manual precipitation data quality control at MeteoSwiss using spatial information

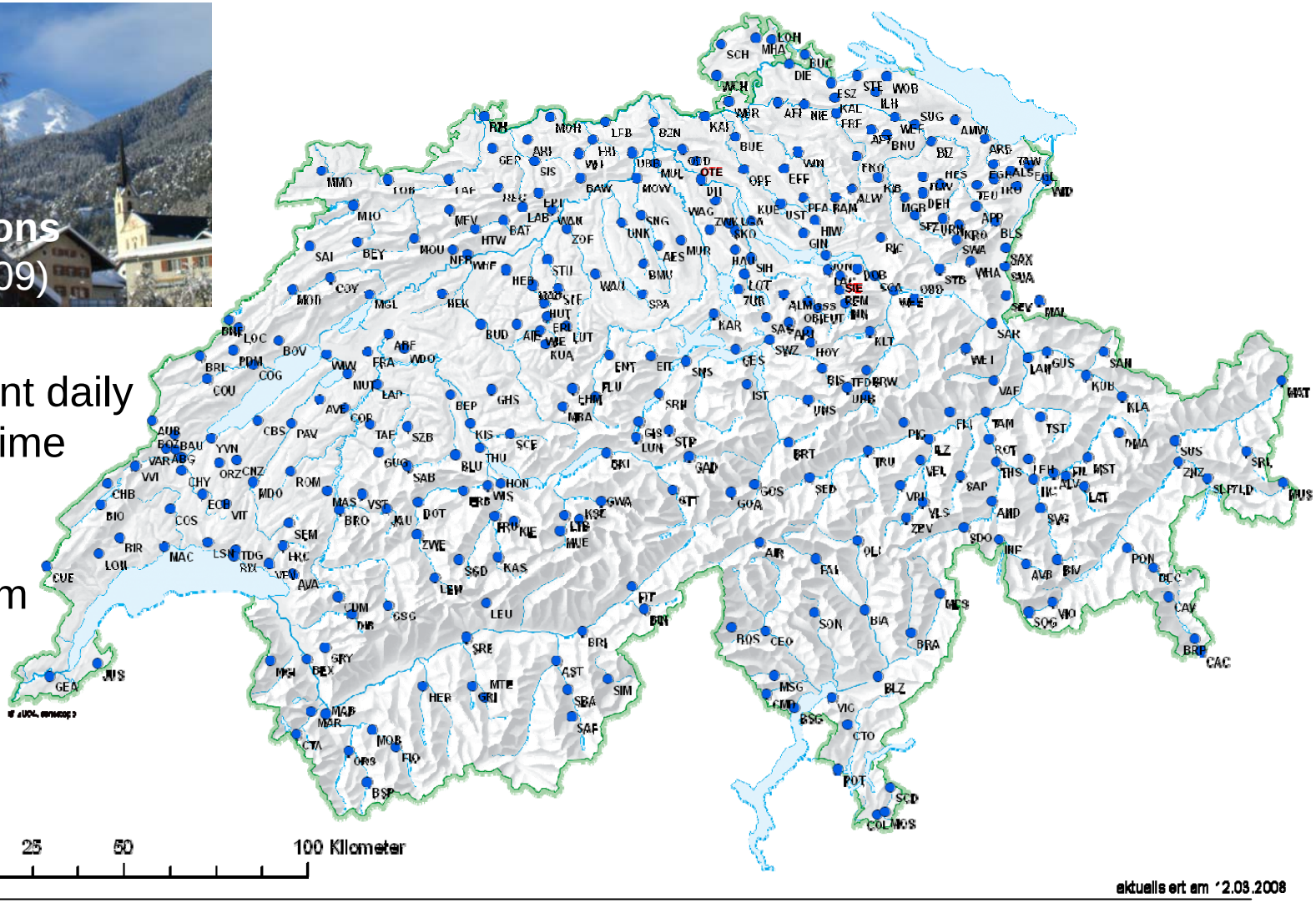
Simon C. Scherrer, D van Geijtenbeek, C Naguel, M Croci-Maspoli and C Appenzeller
Climate Services, Swiss Federal of Office of Meteorology and Climatology MeteoSwiss

5 November 2009
7th ECSN Data Management Workshop 2009, DMI Copenhagen

Manual precipitation network (NIME)

→ measurement daily at 07:30 local time

- precipitation
- new snow sum
- snow height



simon.scherrer@meteoswiss.ch

Manual precipitation network (NIME)

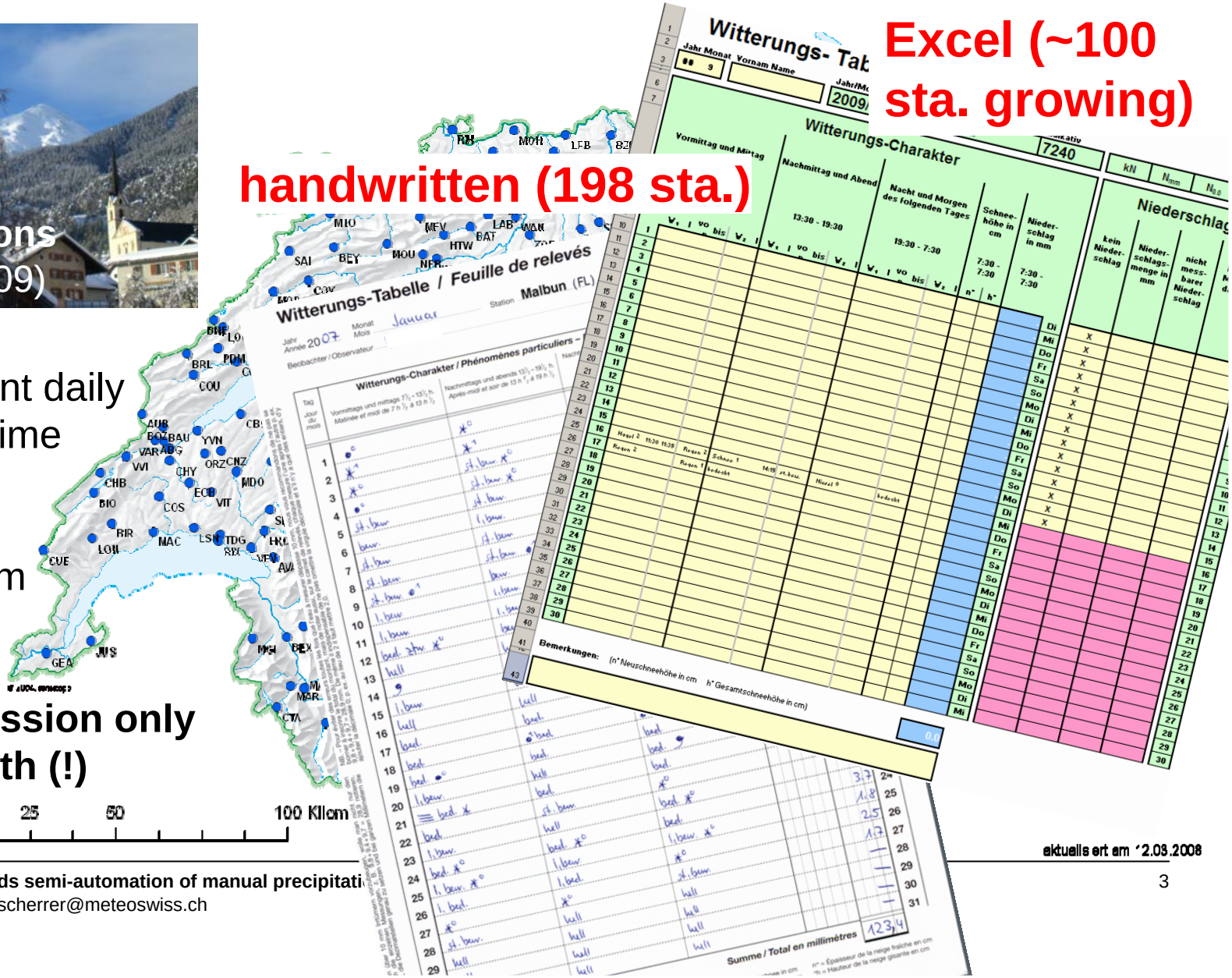
→ measurement daily at 07:30 local time

- precipitation
- new snow sum
- snow height

- **data transmission only once per month (!)**

Excel (~100 sta. growing)

handwritten (198 sta.)



Data quality control NIME

Overview

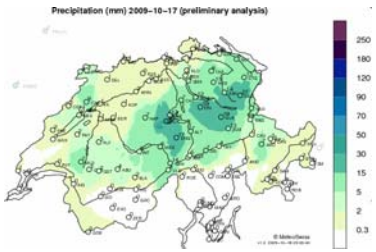
„currently“

manual correction of 331 stations, ~3-4% of values need correction

task: *shift, distribute, interpolate without missing anything*
→ suspicious cases identified optically and subjectively!

Niederschlag – précipitations **help sheet**

Ind.	Station	m ü. M.	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
9580	Simplon-Dorf	1495	.	.	.	0.5	6.5	.	1.2	1.0	.	.
9610	Passo del Bernina	2307	9.0	2.5	.	.	16.4	3.0	38.9	.	.	.
9630	Cavaglia	1706	2.5	2.1	.	.	17.9	6.5	22.3	.	.	.
9670	Poschiavo / Robbia	1078	3.3	0.9	.	.	14.5	9.1	20.9	.	0.2	.
9710	Brusio-Piazzo	830	.	4.9	.	4.7	10.2	9.7	6.9	.	.	.
9730	Campocologno	535	.	3.5	.	.	14.8	8.0	20.4	.	0.2	.
9750	Vicosoprano	1075	1.6	4.0	.	.	8.2	4.3	19.3	.	0.9	.
9780	Soglio	1093	3.4	8.3	.	.	10.7	7.2	37.5	.	0.2	.
9810	Segl-Maria	1798	7.0	1.1	.	.	8.0	2.5	28.4	.	.	.
9820	Piz Corvatsch	3315	7.7	0.2	0.1	0.4	7.8	4.5	15.8	.	0.2	.
9839	Bernina-Curtinatsch	2090	3.6	3.8	3.3	.	19.8	17.3	4.2	.	.	.
9845	Pontresina	1774	3.5	2.9	2.1	.	2.9	5.7	2.7	.	.	.
9849	Samedan	1709	0.1	10.4	0.2	0.1	4.6	5.6	18.7	.	.	.
9870	Buffalora	1970	3.8	5.0	0.5	.	8.7	0.9	25.4	.	.	.

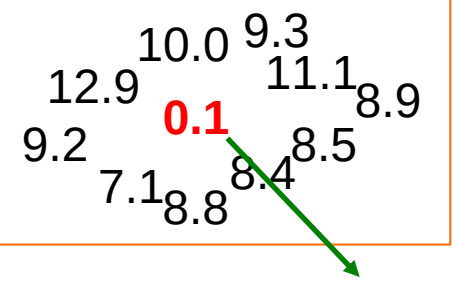


„aim“

semi-automatic correction
using spatio(-temporal) information

task: *find* ***automatically and objectively all*** *possible „suspicious” cases, provide* ***objective interpolation values****, split „several day measurements“*
→ reduce workload

isolated dryness



1) interpol. value **8.6**

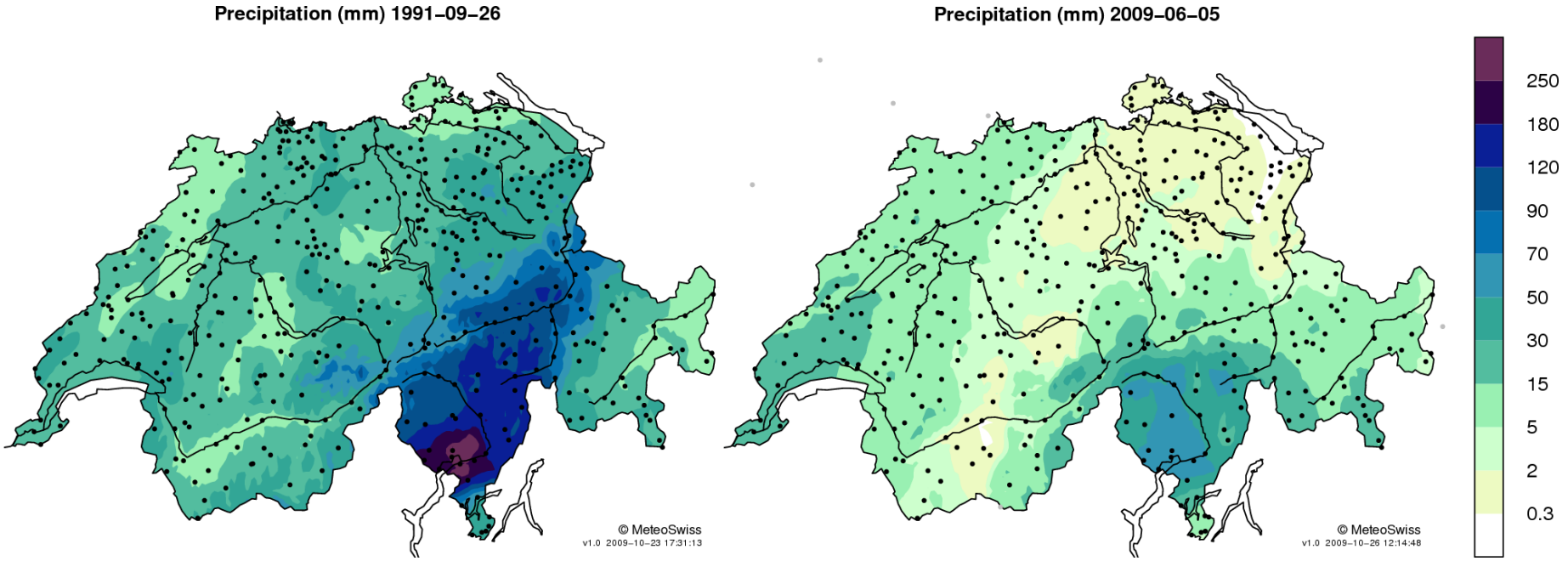
2) uncertainty range **7.3 - 9.4**

simon.scherrer@meteoswiss.ch

Spatial interpolation

objective interpolation values

- ⇨ currently SYMAP gridding products (Frei and Schär 1998[1])
- ⇨ distance-, directional-, and climatological weighting
- ⇨ estimation of values at station but no uncertainty range…



[1]Frei, C. and Schär, C. 1998: A precipitation climatology of the Alps from high-resolution rain-gauge observations. *Int. J. Climatol.*, **18,** 873-900.

Plausibility checks

DWD QUALKO 1992

3 cases of „implausibility“

1. **isolated precipitation**

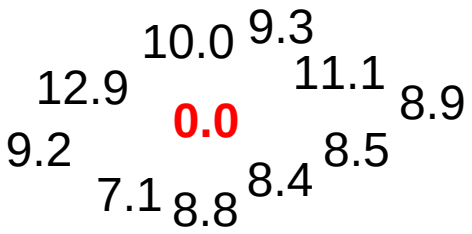
   implausible if

   r_{max}<1 and r_{obs}>10

2. **isolated dryness**

   implausible if

   r_{min}>=0.1 and r_{obs}<0.1 and r_i>=10

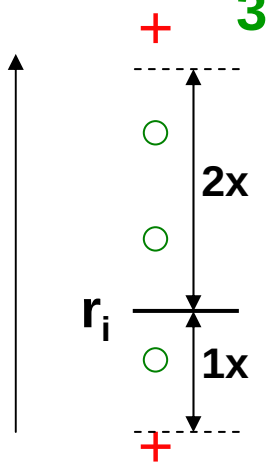


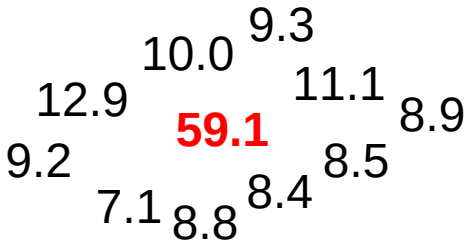
3. **too large deviations (general case) above / below**

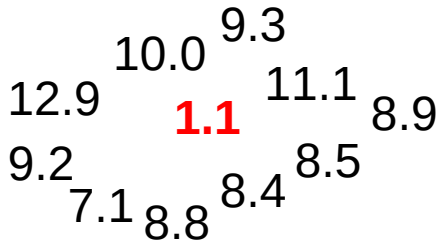
   implausible if

   $r_{obs} > r_i + 2*[(r_{max}-r_{min})/2+30]$

   $r_{obs} < r_i - [(r_{max}-r_{min})/2+30]$

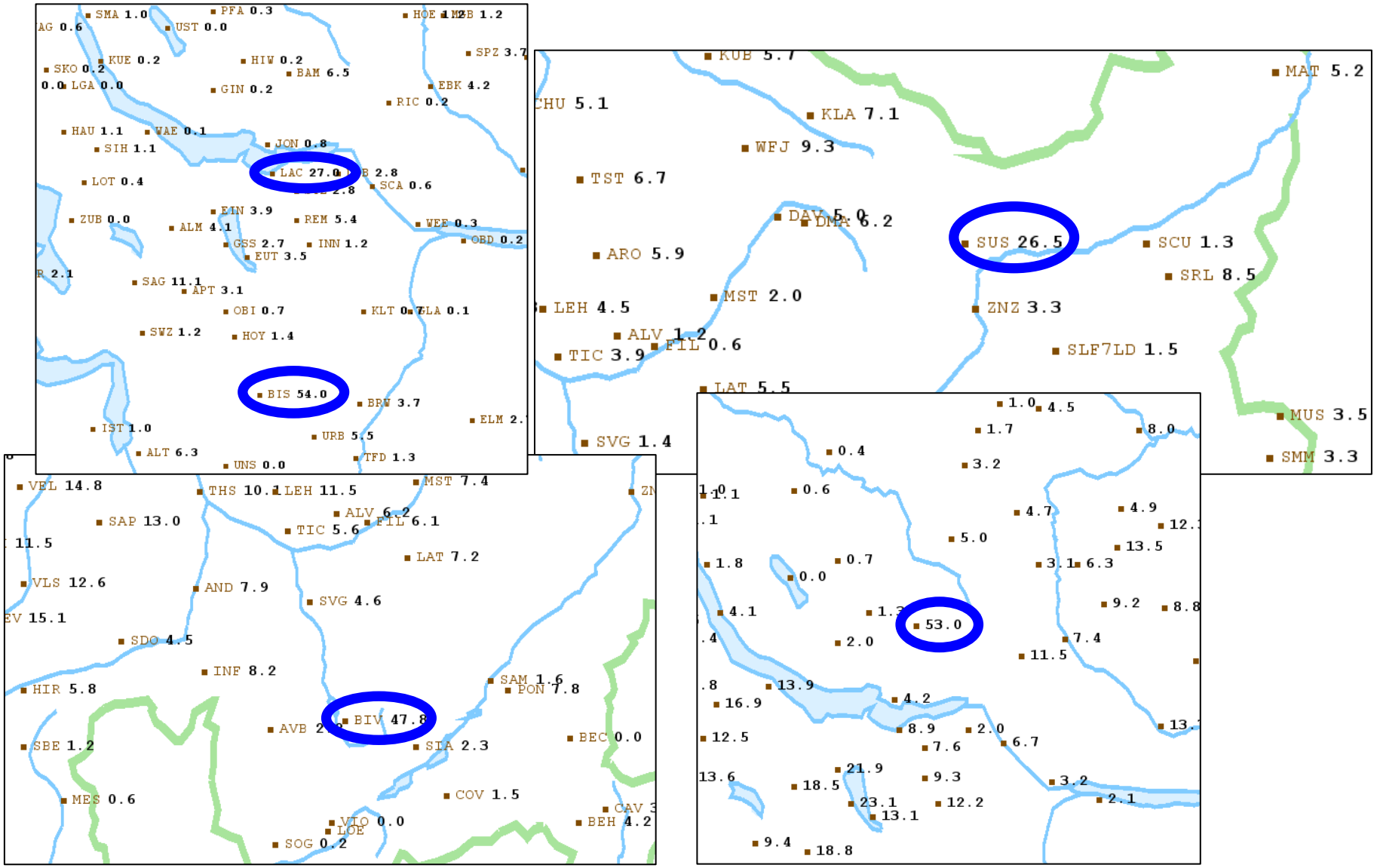






Original QUALKO 1992 checks

some clearly suspicious cases not identified…



Lessons learned from QUALKO limits

Adaptations

- ⇨ DWD QUALKO 1992 limits misses some unplausible precipitation cases (esp on low end)
  - ⇨ isolated precipitation:
    does not find all isolated precipitations
    → adapt limits
  - ⇨ isolated dryness:
    does not find all isolated dynesses
    → adapt limits
  - ⇨ general case (too large devitions):
    does not find all cases we want to identify
    → relative method

simon.scherrer@meteoswiss.ch

Plausibility checks

DWD QUALKO 1992 → **spatial precipitation QC**

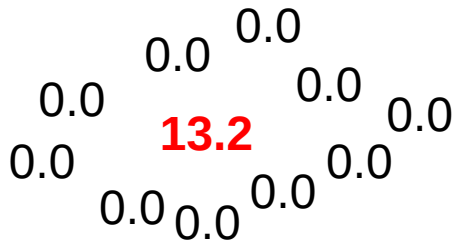
3 cases of „implausibility“

1. **isolated precipitation**

   implausible if

   r_{max}<~~1~~ und r_{obs}>~~10~~ **3**

   0.3

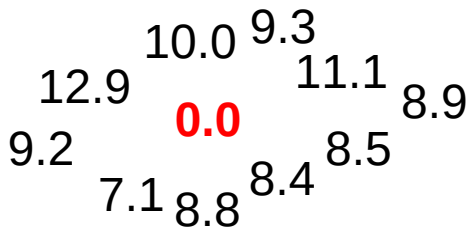


2. **isolated dryness**

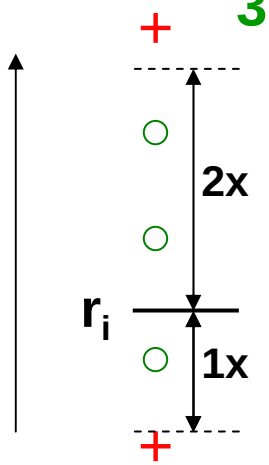
   implausible if **0.3** **3**

   r_{min}>=~~0.1~~ und r_{obs}<~~0.1~~ und r_i>=~~10~~

   0.3



3. **too large deviations (general case) above / below**

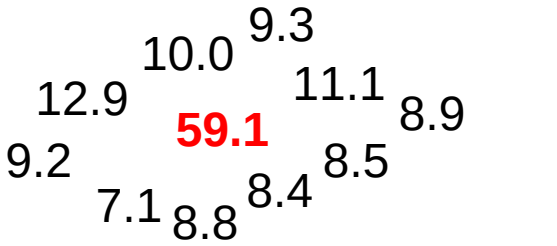


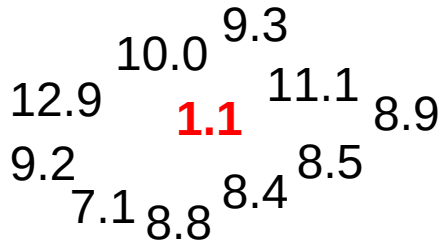
implausible if

$r_{obs} > r_i + 2*[(r_{max}-r_{min})/2+30]$

$r_{obs} < r_i - [(r_{max}-r_{min})/2+30]$

$¾(r_{max}-r_{min})$

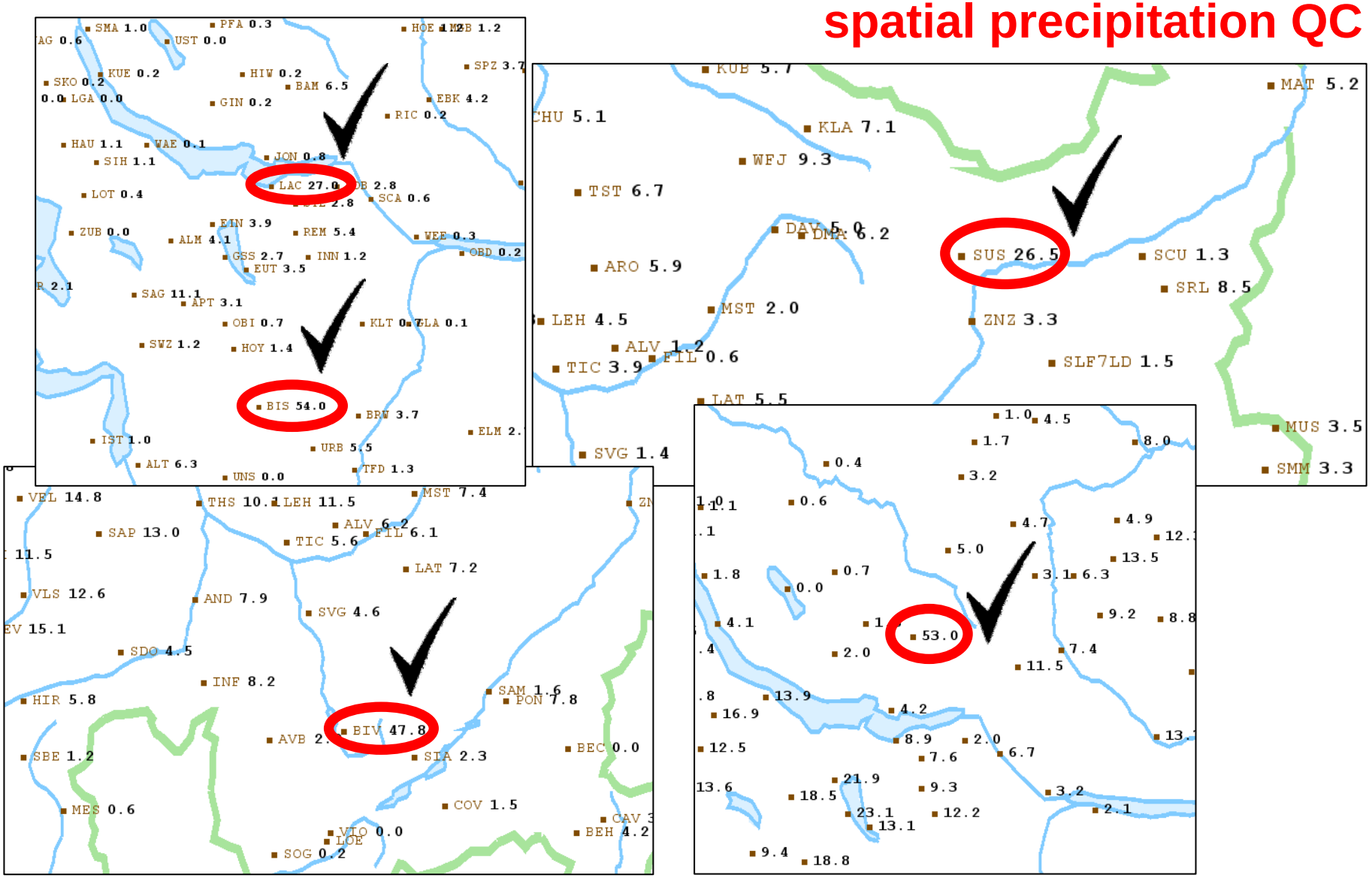




Adapted checks: too large deviation

Outliers now identified as suspicious!

spatial precipitation QC

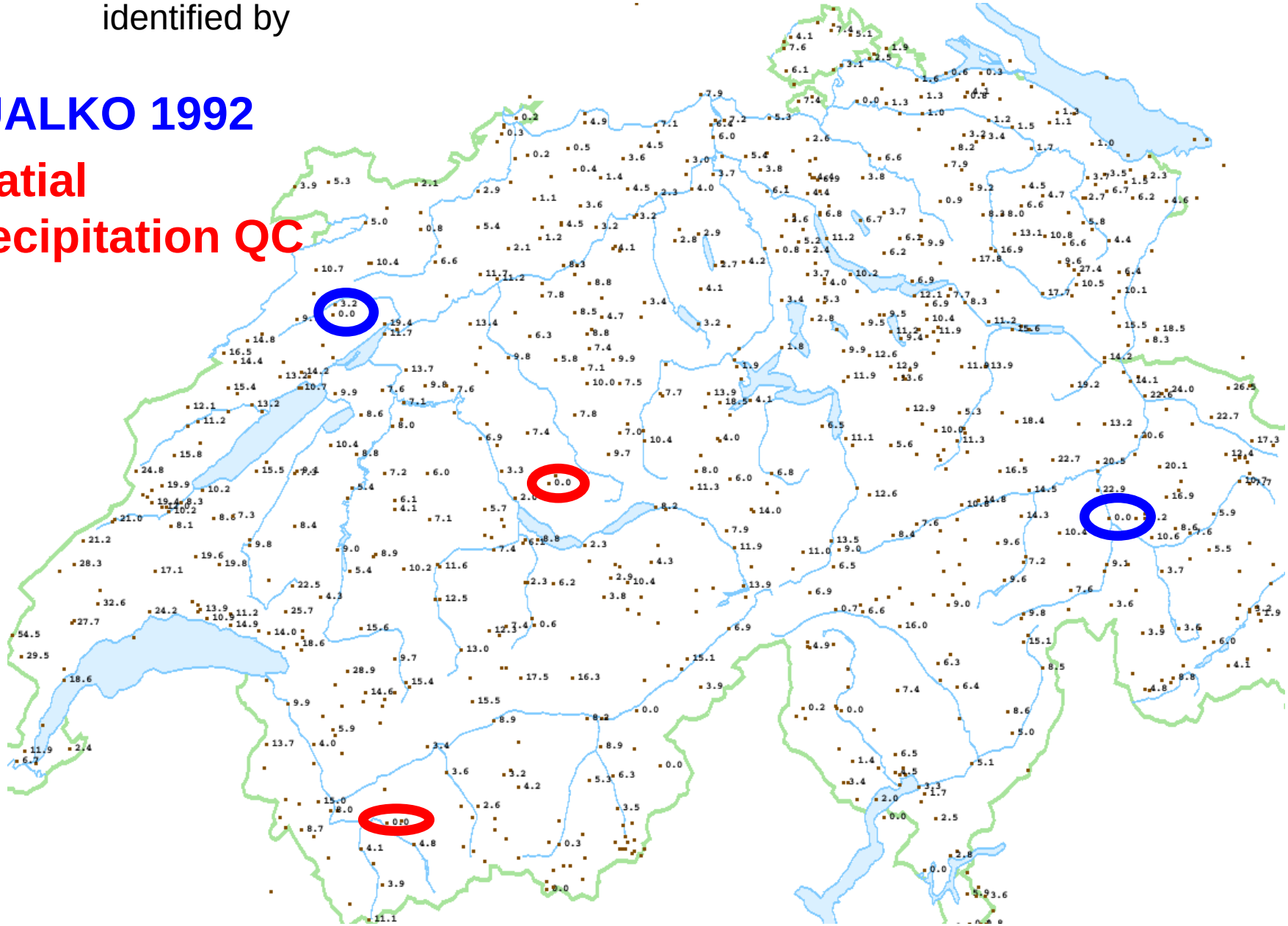


Isolated dryness

identified by

QUALKO 1992

spatial precipitation QC

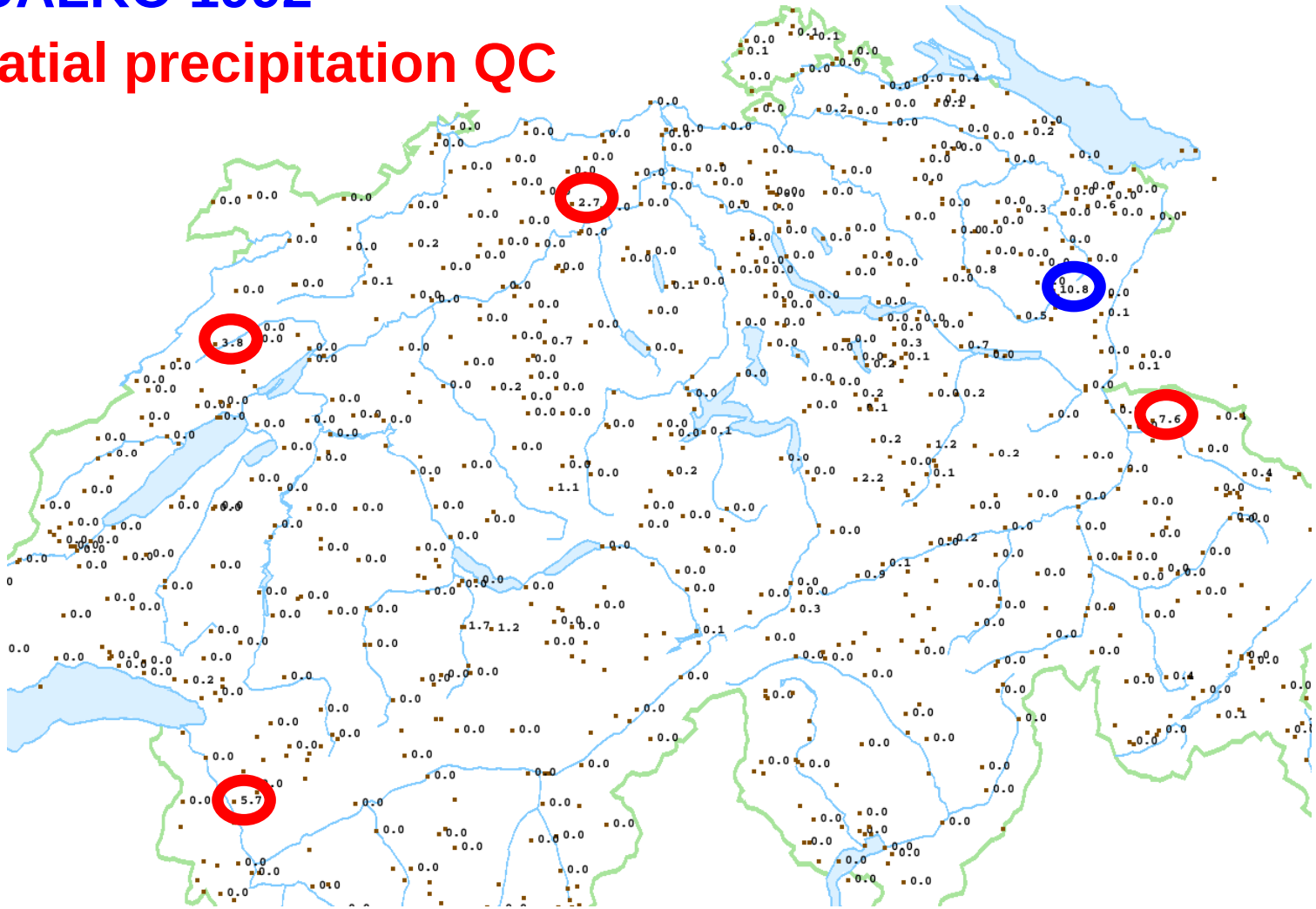


Isolated precipitation

identified by

QUALKO 1992

spatial precipitation QC



NEW help sheet for data editors

example July 2009

stations ↓

→ **day of month (1., 2., 3., … , 30., 31.)**

measured

interpolated

example July 2009

Stand: 16 Aug 2009 08:09

sta/dat	FR 3.	SA 4.	SU 5.	MO 6.
9610/BEH	./1.2	./0.1	16.4/15.1	3.0/10.7
9630/CAV	1.0/0.8	0.1/0.1	16.8/14.1	6.5/8.9
9670/ROE	./0.6	./0.1	14.5/10.1	9.1/5.7
9710/BR	./0.4	4.7/0.1	10.2/12.2	9.7/4.5
9730/CA	./0.5	0.2/0.1	24.6/11.2	./7.4
9750/	./0.2	./0.1	8.2/9.6	4./6.2

sta/dat	SU 12.	MO 13.	TU 14.	WE 15.	TH 16.	FR 17.	SA 18.	SU 19.	MO 20.	TU 21.	WE 22.	TH 23.	FR 24.	SA 25.	SU 26.	MO 27.	TU 28.	WE 29.	TH 30.	FR 31.	sum [mm]	norm %
9610/BEH	./.	./.	./0.1	./1.9	./1.0	31.8/42.5	./5.1	./.	./.	./.	./0.1	./.	8.5/5.7	./0.2	./.	./6.8	./.	./0.6	./.	9.7/9.9	119.8/129.8	82/89
9630/CAV	./.	./.	./.	./.	./.	./.	3.9/5.3	./.	./.	./.	0.5/.	./.	./5.7	./0.1	./.	11.7/4.1	./.	./0.4	./.	./13.3	95.1/120.9	70/89
9670/ROE	./.	./.	./.	./.	./.	7	0.6/5.5	./.	./.	./.	0.1/0.1	./.	4.6/4.6	./0.1	./.	1.2/5.3	./.	./0.3	./.	13.7/9.8	96.2/94.1	92/90
9710/BR	./.	./.	./.	./0.7	./1.3	./0.7	24.9/30.5	6.6/4.2	./.	./.	./.	./0.1	./.	4.9/5.7	./0.2	./.	7.4/4.1	./.	./0.2	./.	22.2/10.8	102.4/104.4	92/94
9730/CA	./.	./.	./.	1.5/.	./0.9	./0.4	./26.9	19.2/4.4	./.	./.	./.	./0.1	./.	1.7/4.8	./0.1	./.	10.7/4.6	./.	./0.2	./.	16.4/10.1	106.4/92.3	102/88
9750/	./.	./.	./.	1.5/1.9	./3.8	47.8/45.1	2.3/4.0	./.	./.	./.	./0.1	./.	12.8/14.6	./0.2	./.	4.8/6.8	./.	./.	./.	12.2/10.0	115.7/139.7	82/99
9780/S	./.	./.	./.	./3.9	4.1/3.4	32.4/63.0	0.2/3.4	./.	./.	./.	0.1/0.1	./.	16.6/16.1	./0.1	./.	5.4/8.7	./.	./.	./.	19.8/8.4	145.9/151.3	96/100
9810/S	./.	./.	./.	./1.1	./2.9	40.4/32.6	10.0/3.8	./.	./.	./.	./.	./.	13.8/7.6	./0.4	./.	8.6/4.8	./.	./0.1	./.	7.4/6.3	127.2/96.1	117/88
9820/CO	./.	./.	./.	./1.1	1.2/1.2	12.2/31.2	29.8/4.3	./.	./.	./.	0.1/.	./.	4.5/6.3	1.2/.	./.	6.0/3.8	./.	./0.2	./.	8.0/4.7	99.7/86.2	110/95
9839/BEC	./.	./.	./.	./.	./.	37.3	3.3/5.9	X/.	X/.	X/.	X/0.1	22.8/.	./6.9	./0.2	./.	3.9/5.2	./.	2.8/.	./.	./8.2	91/114.8	72/91
9845/PON	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	./5.8	./0.2	./.	./4.9	./.	./0.3	./.	7.8/3.6	65.1/90.7	70/98
9849/SAM	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	6.1/5.3	./0.2	./.	5.5/4.6	./.	./0.3	./.	1.4/3.9	99.2/83.9	112/95
9870/BUF	0.1/.	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	./.	13.5/14.1	./.	./.	6.9/8.7	./.	./.	./.	1.3/1.0	95.2/117.6	91/113
9875/PUD	./.	./.	./.	./.	1.6/0.8	./.	./.	./.	./.	./.	./.	./.	1.7	./.	./.	9.5/6.4	./.	./0.1	./.	0.4/1.0	89.1/111.4	88/110
9890/ZNZ	./.	./.	./.	./.	1.2/1.0	3.7/3.7	29.0/29.4	./.	./.	./.	./.	./.	./.	./.	./.	15.8/5.9	X/.	X/.	X/.	X/0.3	99.2/98.7	105/104
9900/SUS	./.	/0.1	./.	./.	./3.1	4.7/1.9	38.8/31.6	8.2/7.0	./.	./.	./.	./.	./.	./.	./.	/10.9	./.	./.	./.	./0.3	115.7/105.7	115/105
9926/SRL	./.	/0.1	./.	./.	./1.3	./1.4	30.7/22.5	5.2/5.2	./.	./.	./1.4	./.	./.	./.	./.	11.8/7.7	./.	./.	./.	./0.8	110.4/94.5	X/X
9930/SCU	1.2/1.9	5.3/2.3	1.3/0.4	./0.1	4./0.2	./.	./.	0.1/.	./.	0.1/0.1	./.	./.	0.8/1.0	./1.5	24.5/23.2	9.6/4.8	./.	./.	./1.4	./.	./.	11.3/8.4	./.	./.	./.	./0.4	87.5/92.1	101/106
9960/MAT	3.4/1.7	3.1/2.8	0.2/0.5	0.4/.	3.8/12.5	0.5/4.3	24.5/24.4	./.	./.	0.2/.	./.	0.3/.	./.	./.	0.2/1.5	./1.0	18.5/24.6	8.5/5.1	./.	./.	0.2/1.6	./.	./.	11.7/10.1	./.	./.	20.5/7.8	./.	./.	./.	./0.5	96/98.4	102/105
9980/SMM	1.2/2.1	3.5/2.8	./0.4	./.	33.8/14.7	1.3/7.8	36.0/25.6	./.	./.	./.	./.	./.	./.	./.	1.7/1.4	./1.3	27.2/24.0	5.4/5.2	./.	./.	6.6/0.7	0.4/.	./.	14.5/12.0	./.	./.	5.4/8.2	./.	./.	./.	3.2/0.9	140.2/107.1	142/108
9990/MUS	0.9/1.7	2.5/2.3	./0.3	./.	26.0/12.7	16.4/1.9	21.1/21.1	./.	./.	./.	./.	./.	./.	./.	1.4/1.2	./0.9	18.5/20.7	6.6/3.8	./.	./.	/2.1	/0.1	./.	13.1/9.0	./.	./.	5.6/7.0	./.	./.	./.	1.6/0.9	113.7/85.7	140/105

→ **red: automatically detected „unplausible value“ candidate**

→ **green: automatically splitted „several day measurement“ (relative)**

→ **blue: several day measurement to split manually**

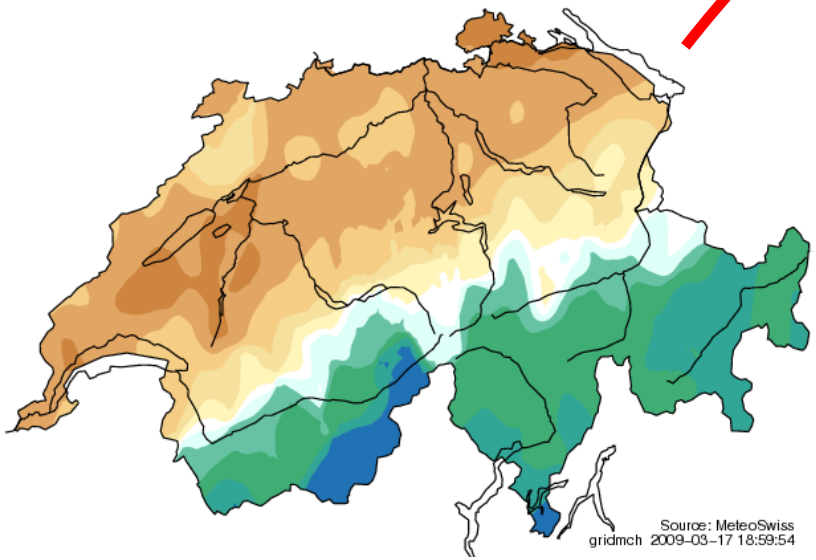
simon.scherrer@meteoswiss.ch

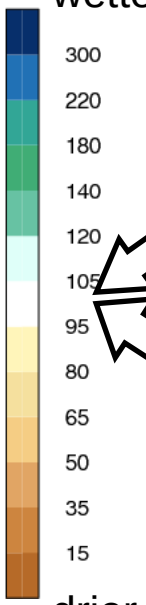
monthly precip. anomaly

example NOV 2008 (%)

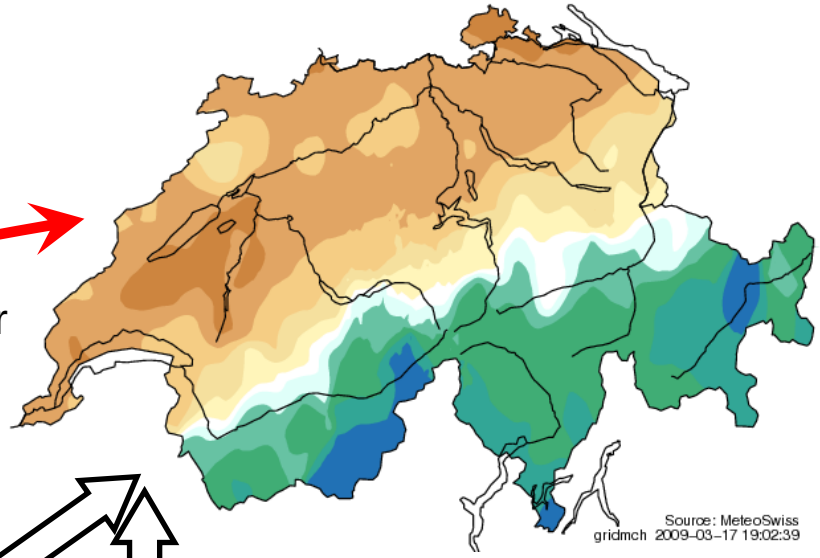
heavy precipitation south of and in Alps, dry in north

raw data

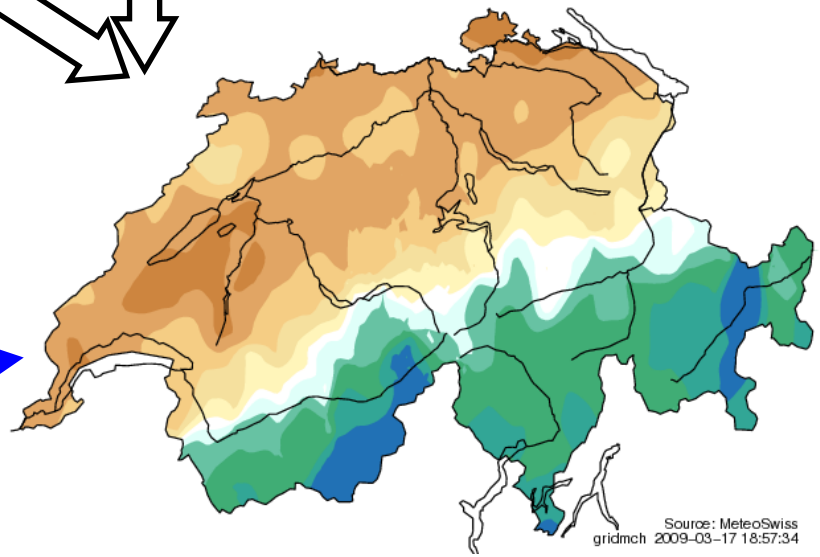




i) auto-DQC interpol. data

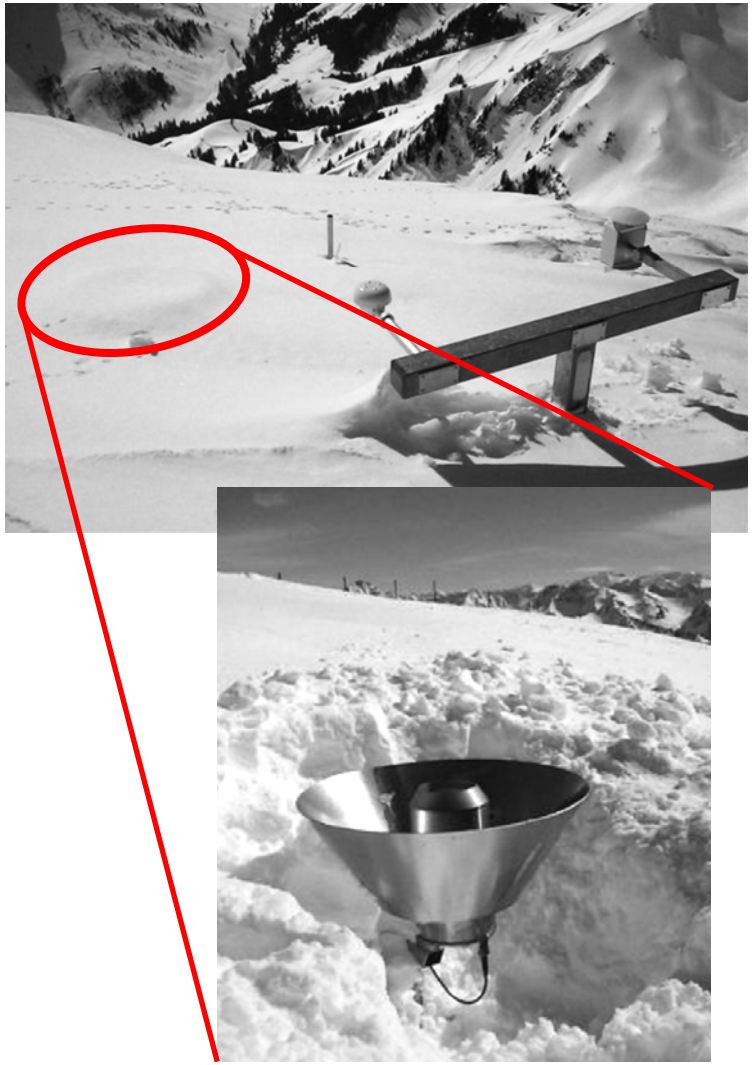


ii) manually treated data

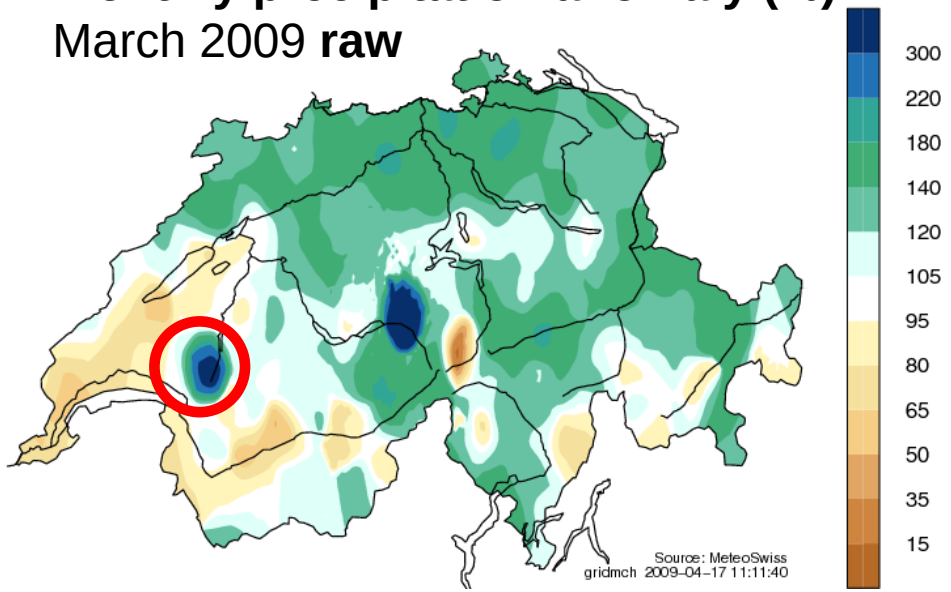


with possibility to suppress
certain stations (e.g. at high altitudes)

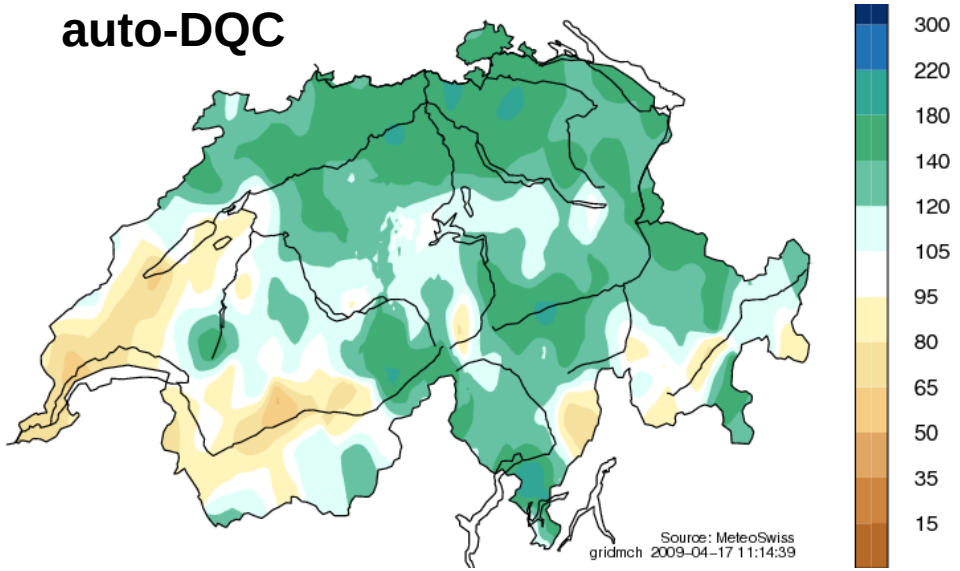
Example with obvious problems…



monthly precipitation anomaly (%)
March 2009 **raw**



auto-DQC



Summary

⇨ implemented method to **objectively detect suspicious** daily precipitation sums using spatial interpolation plausibility in complex terrain (incl. splitting of several day measurements)

⇨ method helps to **objectify** the **editing** of precipitation data and to **reduce the workload** for the data editors
→ concentration on editing outliers instead of finding them

⇨ major challenges:

1. **interpolation** over mountainous terrain
   → sparse network, esp. for convection
2. **plausibility limits**
   → What is an outlier? Site and amount specific?

simon.scherrer@meteoswiss.ch

Challenge: interpolation (I)

SYMAP mean absolute error (MAE) and mean relative error (MRE)

year 2007
(2005/2006 similar)

precip [mm]	**MAE** [mm]	**MRE** [%]
<0.3	0.54	360
0.3-0.5	0.76	190
0.5-1	**0.97**	**131**
1-3	1.4	70.9
3-5	1.8	45.6
5-10	2.4	33.1
10-20	3.6	25.1
20-30	**5.1**	**21.0**
30-50	8.2	22.2
>50	13.3	20.4

for values < ~2 mm:

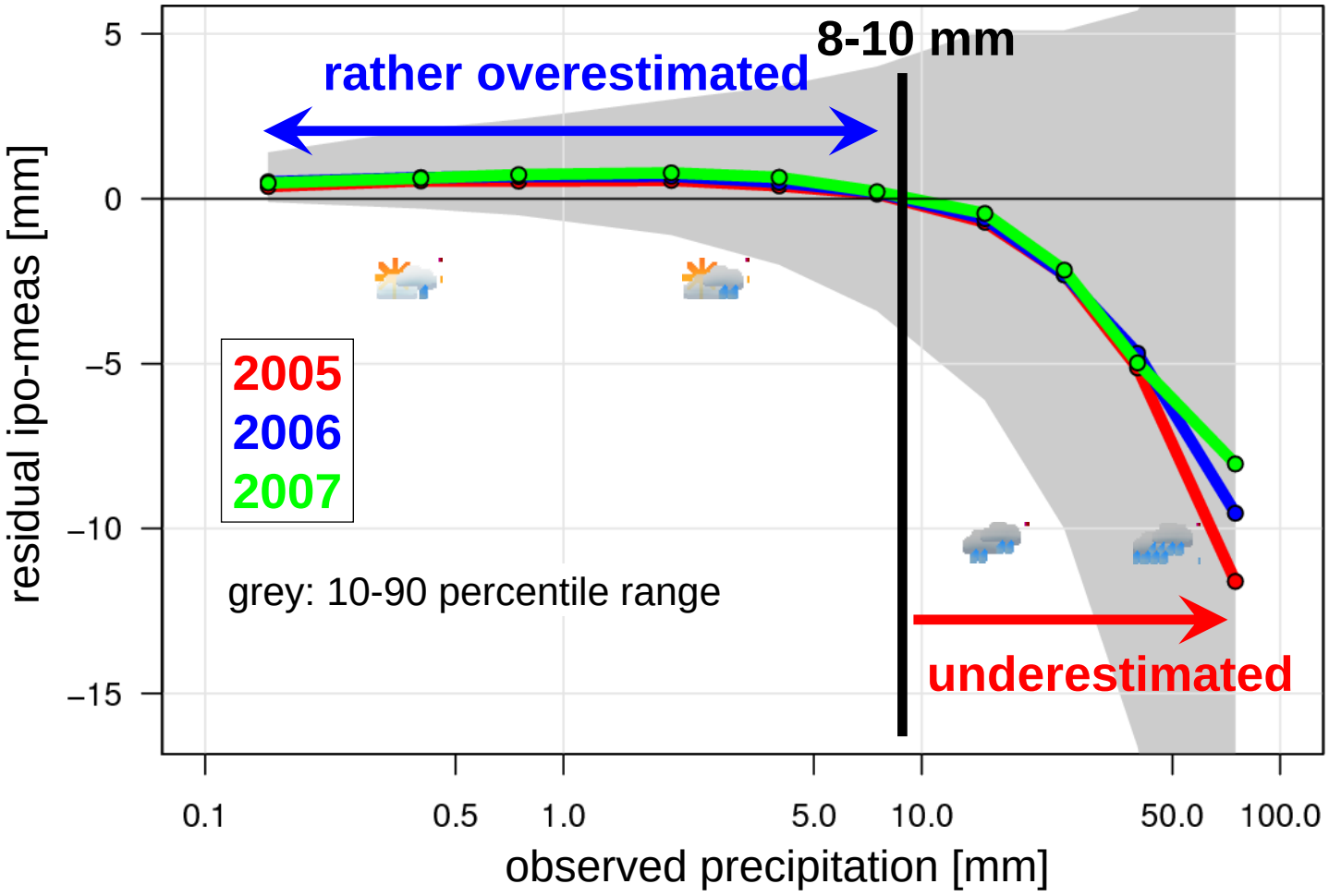
mean error larger than measurement!

~20%

simon.scherrer@meteoswiss.ch

Challenge: Interpolation (II)

SYMAP interpolation performance 2005-2007



simon.scherrer@meteoswiss.ch

Challenge: plausibility limits

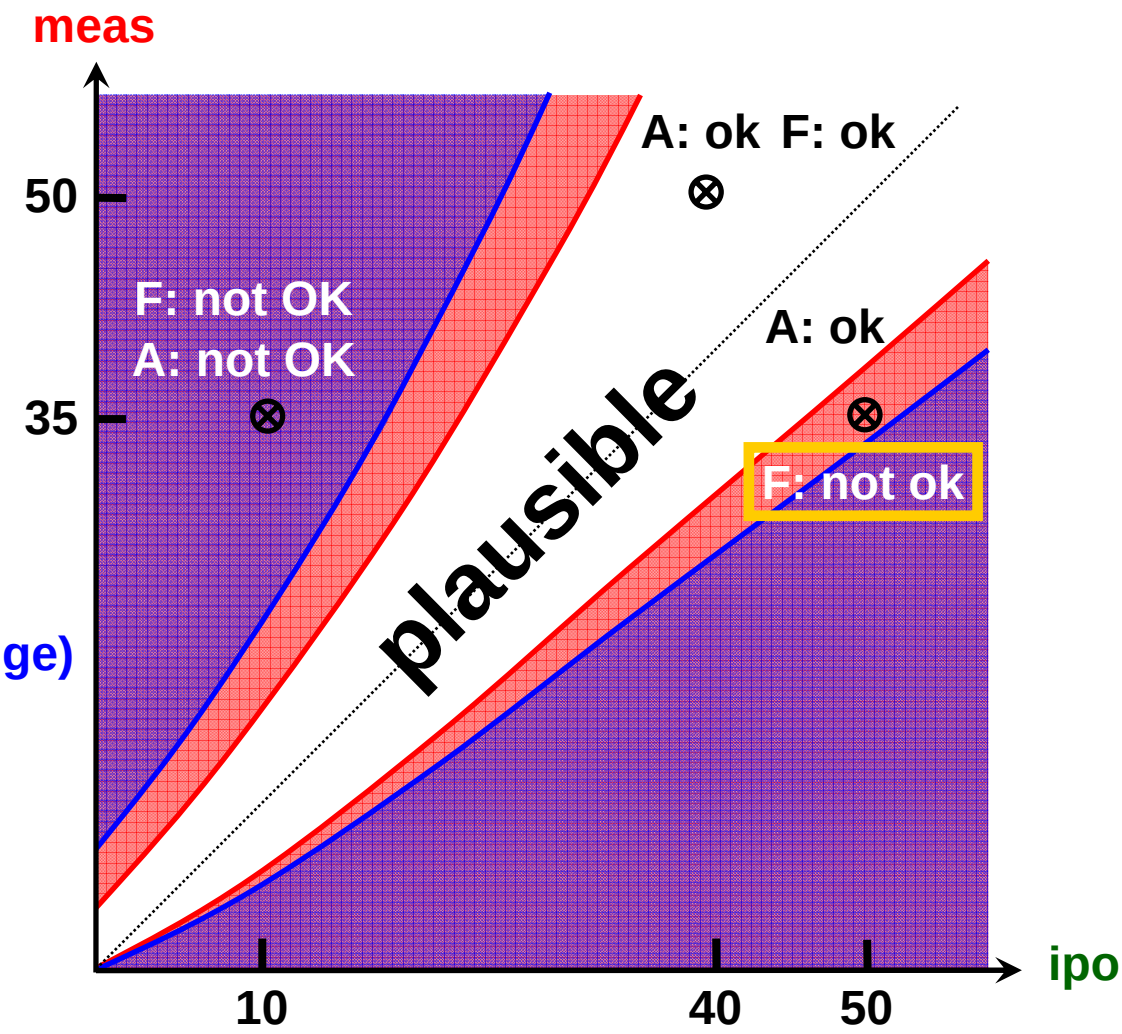
site&amount-specific based on climatological interpolation errors

Are the three measurements plausible?

ANSWER: depends on „interpolation error climatology“ of the site!

Alpine site A (blue range)

Flat site F (red range)



simon.scherrer@meteoswiss.ch

Outlook

- **measurements / data transmission**
  - manual precipitation measurements will go on
  - daily data transmission via SMS/phone (~200 stations)
  - automation of >100 stations or certification of partner sites
- **enhanced interpolation**
  - reduced space optimal interpolation (RSOI) ie use of spacio-temporal information (Schiemann et al. 2009)
  - combination *"ground obs + radar"* (PhD thesis R. Erdin)
- **plausibility limits:** site & amount specific statistical modelling of the interpolation error to identify possible outliers
- **operational implementation** of interpolated values / plausibility info into MeteoSwiss data warehouse (DWH)

more on DQC at MeteoSwiss…

Poster **7-P2**:

„The data quality control chain for automatic surface observation data at MeteoSwiss“

by Deborah van Geijtenbeek et al.