Data & metadata tests

Luděk Vecsey

Metadata health

https://github.com/PetrColinSky/AdriaArray

- list of AdriaArray stations: Petr's tables + _ADARRAY
- metadata downloaded from EIDA on Mar 09

TESTS

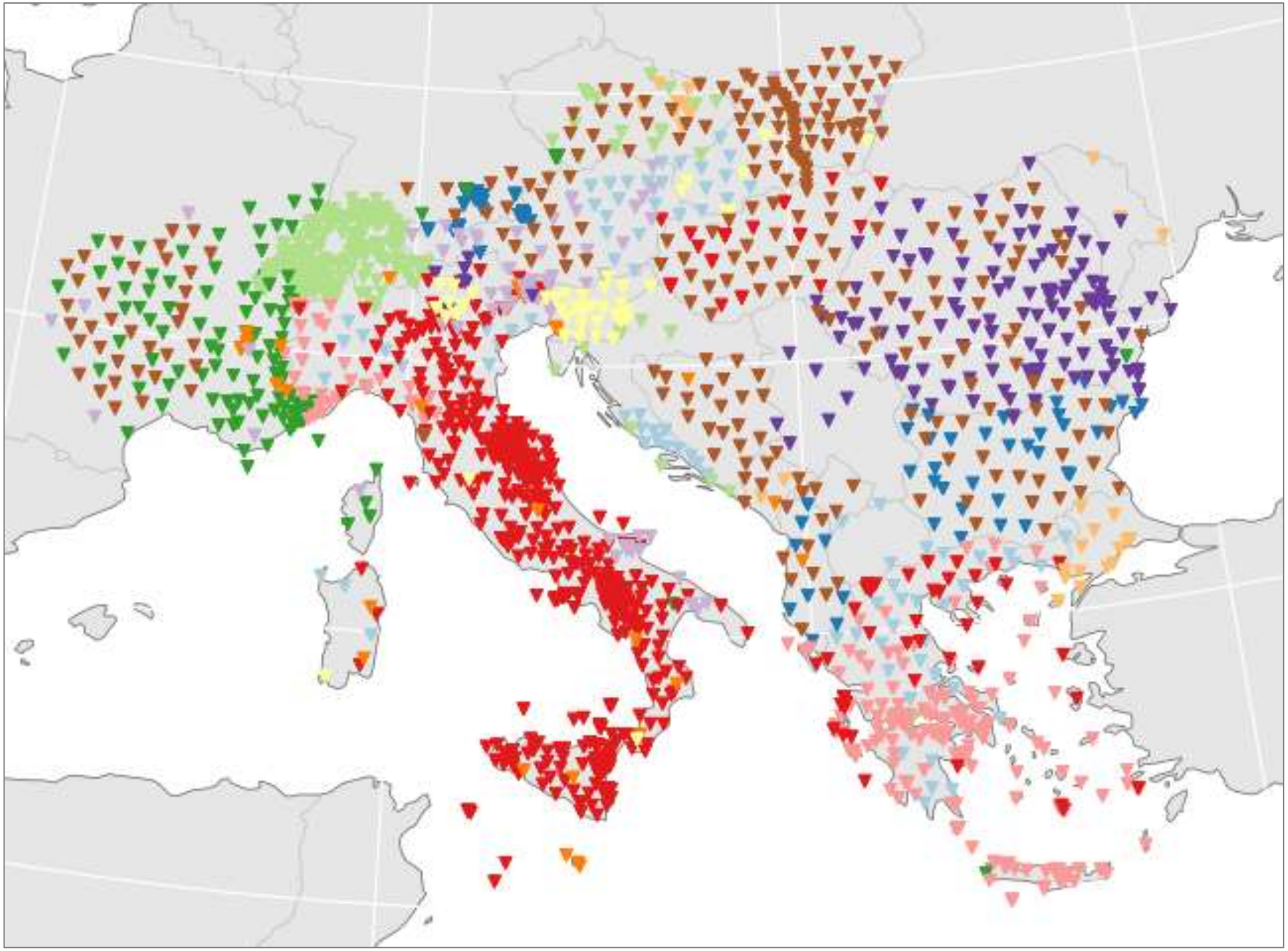
- FDSN StationXML validation
- IRIS StationXML Validator - https://github.com/iris-edu/StationXML-Validator/wiki
  - validation levels: network, station, channel, response
- corner periods computed from the metadata responses
- checks for P&Z consistency
- cross-checks of corner periods from metadata and independent list

Tested stations

Petr's tables
- both broadband and short-period
- only current stations

_ADARRAY
- broadband (corner>=30s)
- stations current or closed after 2022-01-01

Stations: 1861
- downloaded from EIDA
- version 9.3.2024



List of tested issues

IRIS StationXML validator

		id	severity		
FDSN StationXML validation		2	-1	META_VALID	# metadata FDSNStationXML validation
		2.01	3	META_ATTRIB_MISS	# metadata - missing attribute
		2.02	1	META_CLOCK_VALUE	# metadata - wrong value format in ClockDrift
	network	3.111	3	META_IRIS_111	# metadata - Station epochs overlap
		3.112.1	1	META_IRIS_112.1	# metadata - Network and/or Station endDates set to future
		3.112.2	3	META_IRIS_112.2	# metadata - Station startDate before Network startDate
	station	3.212.2	3	META_IRIS_212.2	# metadata - Channel startDate before Station startDate
		3.222	4	META_IRIS_222	# metadata - Station and Channel Latitudes/Longitudes differ too much
		3.223	4	META_IRIS_223	# metadata - Station and Channel Elevations differ too much
	channel	3.304	1	META_IRIS_304	# metadata - no Sensor Description in Channel
		3.310	3	META_IRIS_310	# metadata - Channel.startDate should exist and be before Channel.endDate
		3.320	4	META_IRIS_320	# metadata - missing Azimuth and/or Dip
		3.321	4	META_IRIS_321	# metadata - invalid InputUnits (first Stage) and/or OutputUnits (last Stage)
		3.332	1	META_IRIS_332	# metadata - mis-orientation >= 5, component code should be one of [123]
	response	3.402.1	0	META_IRIS_402.1	# metadata - case inconsistency in Input/OutputUnits
		3.402.2	3	META_IRIS_402.2	# metadata - invalid Input/OutputUnits
		3.402.3	4	META_IRIS_402.3	# metadata - invalid InputUnits in the first Stage
		3.403	4	META_IRIS_403	# metadata - Input/OutputUnits do not follow in Stages
		3.404	5	META_IRIS_404	# metadata - missing Decimation and/or StageGain in Stage(s)
		3.410	4	META_IRIS_410	# metadata - missing Value in InstrumentSensitivity
		3.411	3	META_IRIS_411	# metadata - InstrumentSensitivity.Frequency must be < SampleRate/2
		3.414	4	META_IRIS_414	# metadata - StageGain.Frequency is zero in undue Stage
		3.416	4	META_IRIS_416	# metadata - missing InstrumentSensitivity in Response
		3.420	3	META_IRIS_420	# metadata - missing at least one Decimation in Stages
		3.421	4	META_IRIS_421	# metadata - output samplerate from Stages not equal to Channel.SampleRate
		3.422	4	META_IRIS_422	# metadata - samplerates do not follow in Stages
Response checks, corner periods, band codes		4	5	META_RESP_FAIL	# metadata response failure
		4.01	5	META_NO_POLES	# metadata - no Poles in PolesZerosResponseStage
		4.02	5	META_LAST_STAGE_FAIL	# metadata - the last response stage is faulty
		4.03	3	META_REAL_POLE_POSITIVE	# metadata - real(Pole) >= 0 in the analog PolesZerosResponseStage
		4.04	5	META_POLES_NOT_CONJUG	# metadata - not conjugated Poles in the PolesZerosResponseStage
		4.05	3	META_BAND_SAMPL_RATE	# metadata - incorrect channel band due to sampling rate
		4.06	4	META_BAND_BB_SP	# metadata - channel band confusion in broadband and short-period
		4.07	2	META_CORNER_TABLE	# metadata - corner confusion between real and tabled values

Issue severity levels

(very subjective)

NOTIFICATONS

WARNINGS

ERRORS

0 1 2 3 4 5

not important at all

station operators could check their metadata

station operators **should** check their metadata

- case inconsistency in Input/OutputUnits

Lvl 0 is not shown in our tests!

- wrong value format in ClockDrift
- Network and/or Station endDates set to future
- no Sensor Description in Channel
- **mis-orientation >= 5, component code should be one of [123]** ?

- confusion between real and tabled values of corner periods

- Station epochs overlap
- Station startDate before Network startDate
- invalid Input/OutputUnits
- missing at least one Decimation in Stages
- incorrect channel band due to sampling rate

- Station and Channel Latitudes/Longitudes differ too much
- missing Azimuth and/or Dip
- missing Value in InstrumentSensitivity
- missing InstrumentSensitivity in Response
- channel band confusion in broadband and short-period

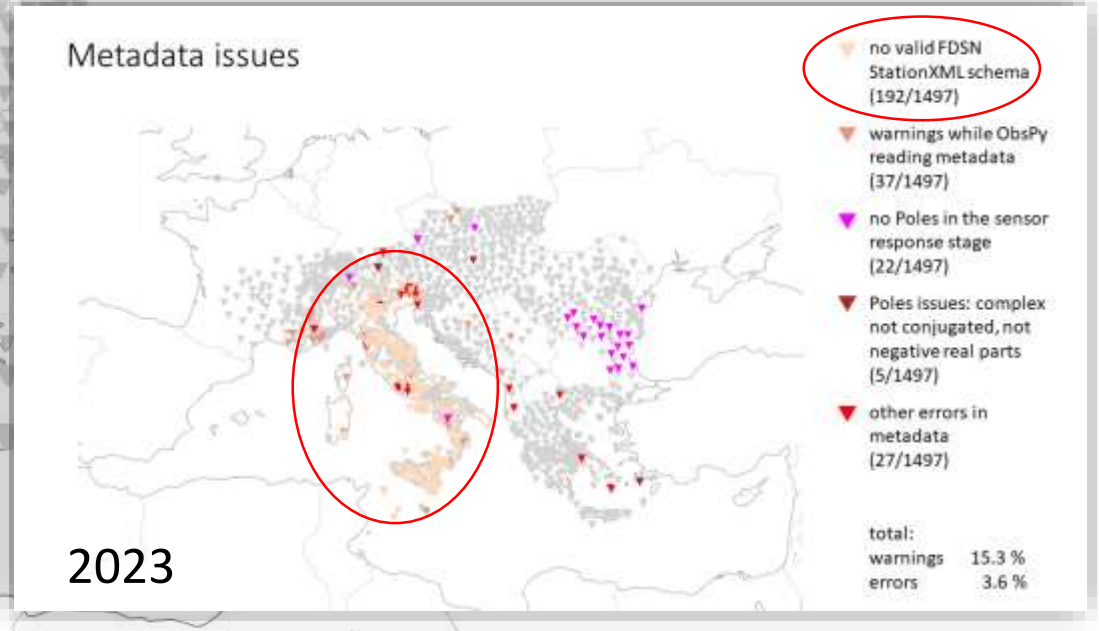
- missing Decimation and/or StageGain in Stage(s)
- metadata response failure
- no Poles in PolesZerosResponseStage
- the last response stage is faulty
- not conjugated Poles in the PolesZerosResponseStage

FDSN StationXML validation

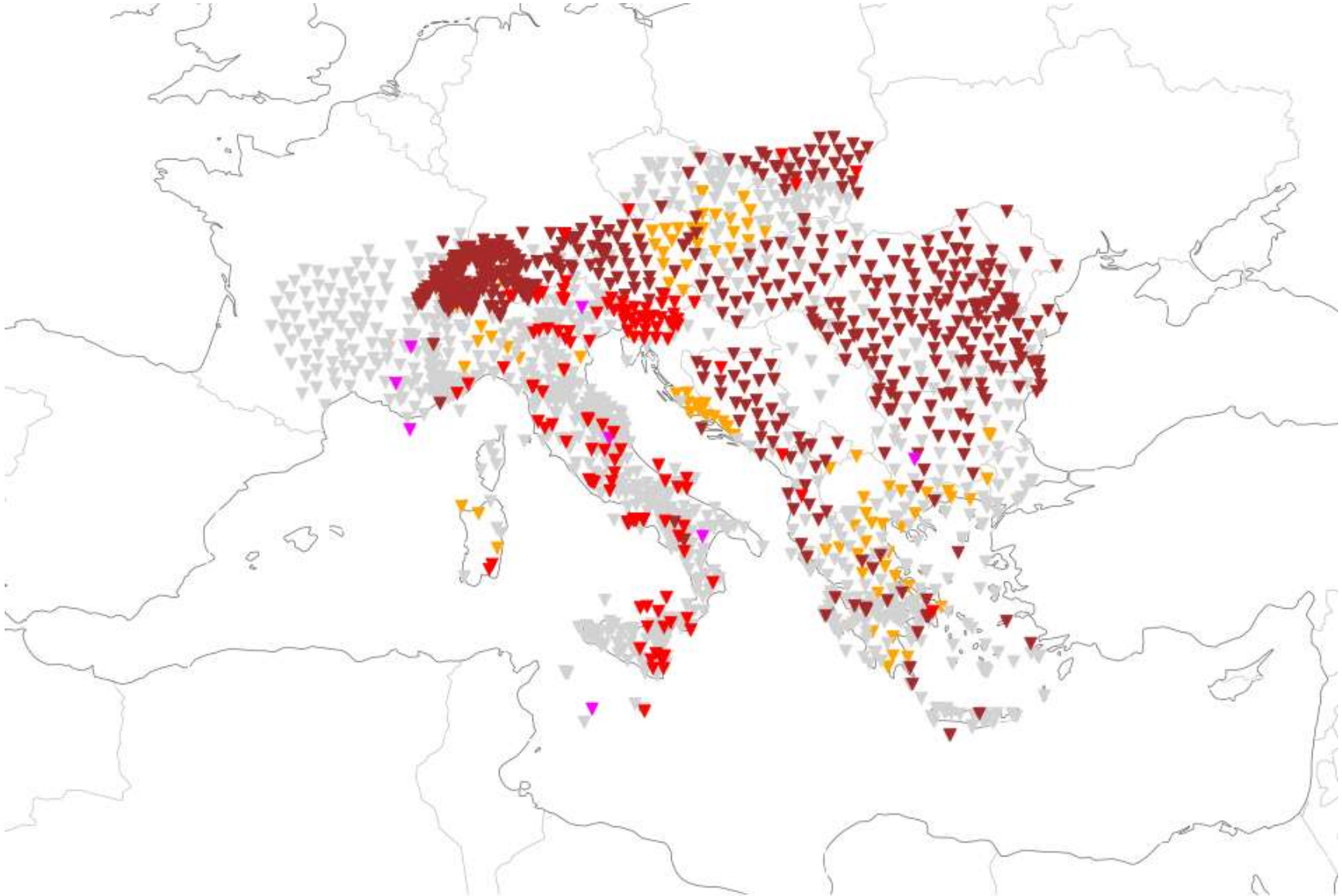
- missing attribute, not readable by ObsPy (3 stations)
- missing attribute (2 stations)
- wrong value format in ClockDrift (1 station)

2024

Big improvement from 30.3.2023!



IRIS StationXML validator



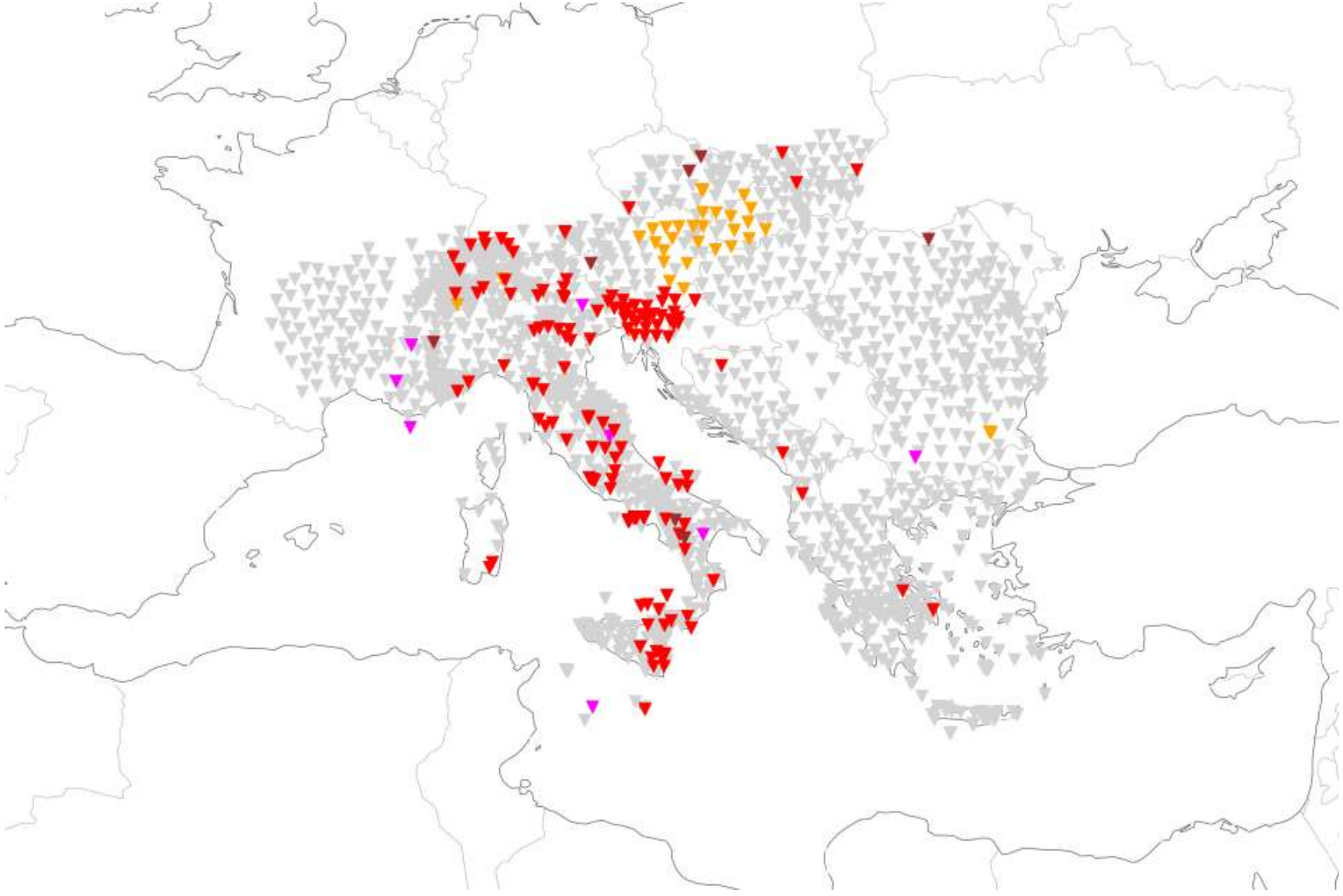
- network level (4.7 %)
- station level (0.5 %)
- channel level (31.6 %)
- response level (8.0 %)

total: 45 %

lots of notifications!

channel level:
no Sensor Description in Channel

IRIS StationXML validator without notifications



network level (1.6 %)

station level (0.5 %)

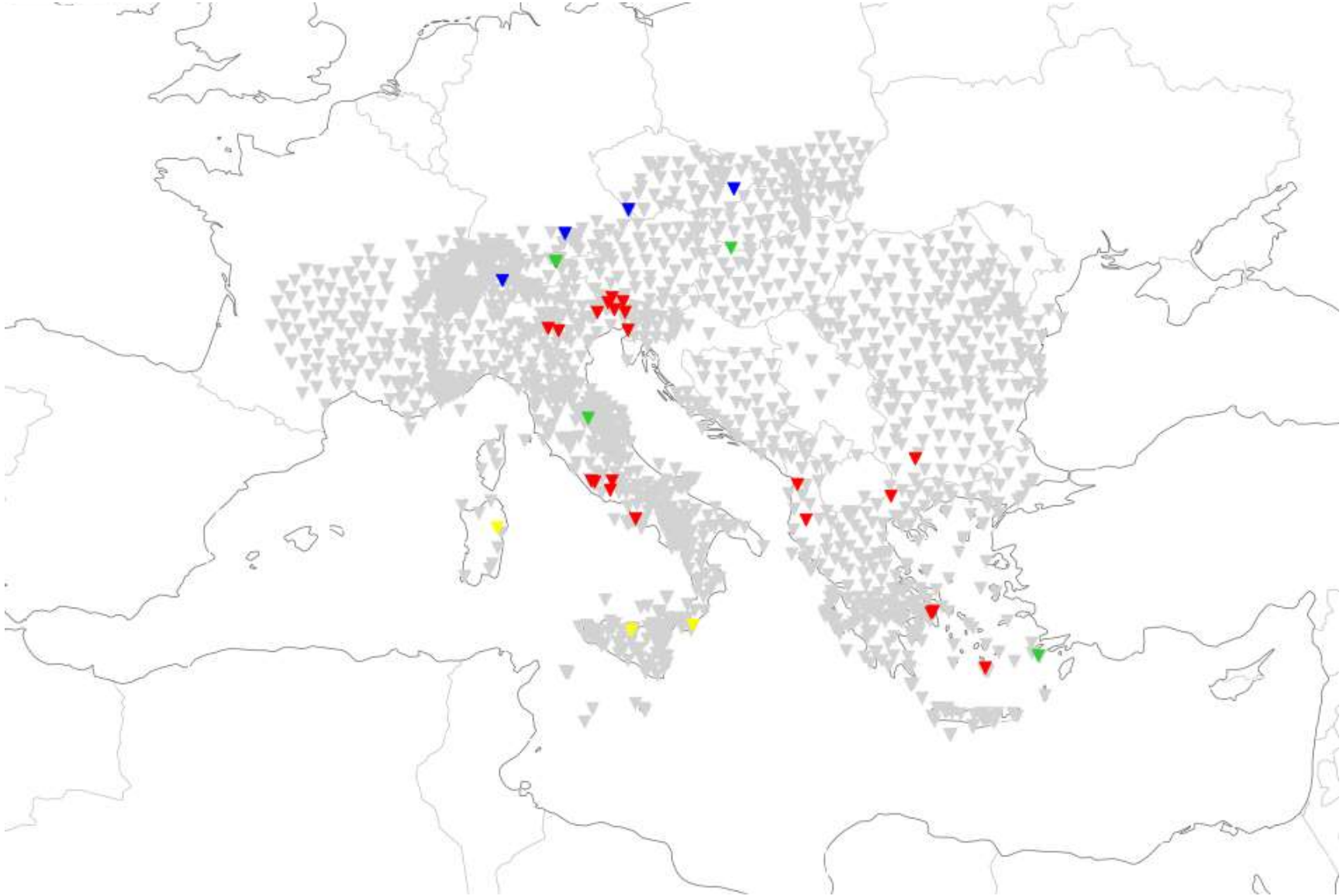
channel level (0.4 %)

response level (8.0 %)

total: 10 %

errors & warnings

Another response checks



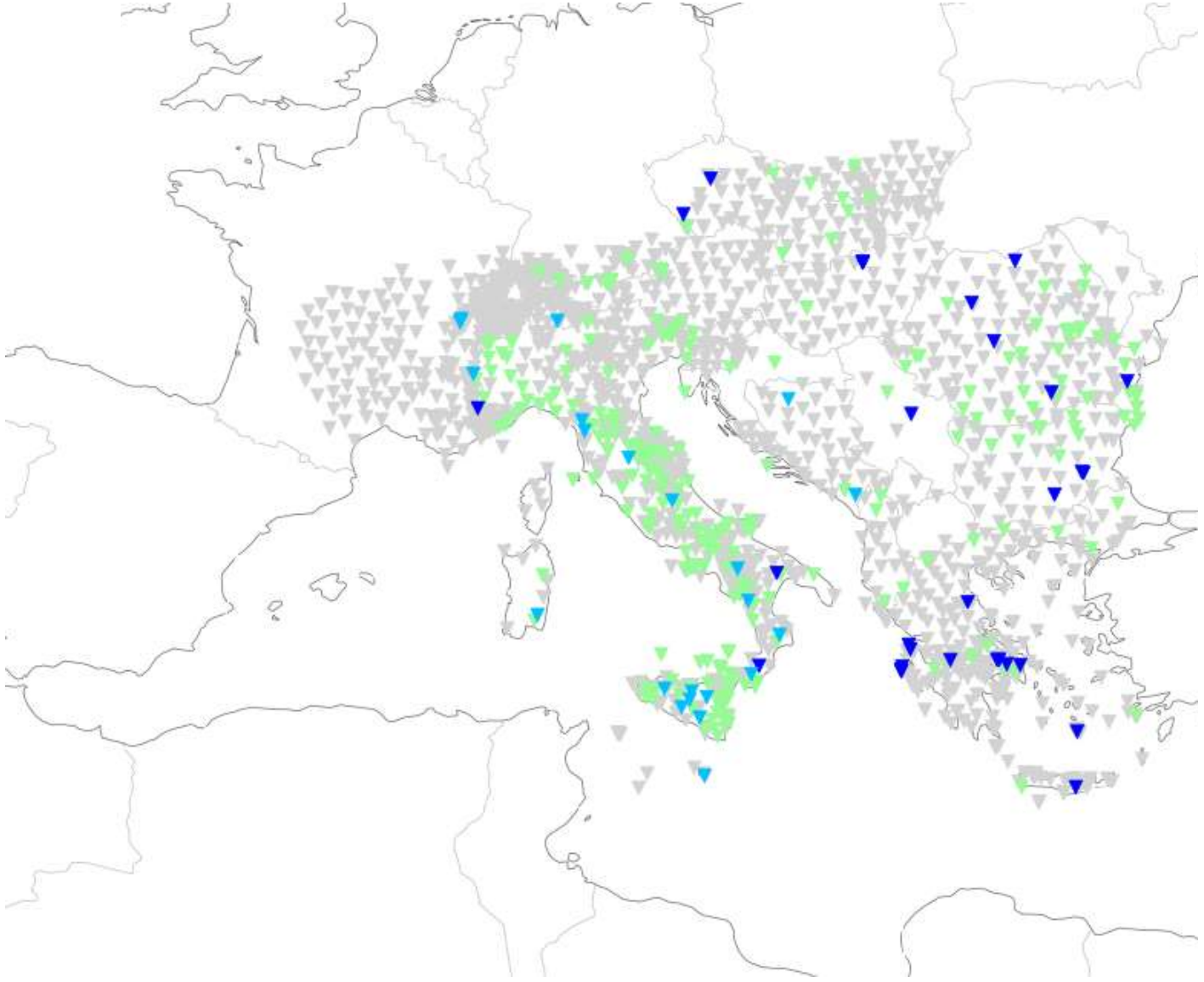
- response failure (1.1 %)
- no Poles in the sensor response stage (0.2 %)
- complex Poles not conjugated (0.2 %)
- not negative real Pole parts in the sensor stages (0.2 %)

total: 1.7 %

errors

Corner period, band codes

sampling frequency	+	corner period	=	band code
20 Hz		60 s		BHZ



Band Code

The first letter specifies the general sampling rate and the response band of the instrument. (The "A" code is reserved for administrative functions such as miscellaneous state of health.)

Band code	Band type	Sample rate (Hz)	Corner period (sec)
F	...	≥ 1000 to < 5000	≥ 10 sec
G	...	≥ 1000 to < 5000	< 10 sec
D	...	≥ 250 to < 1000	< 10 sec
C	...	≥ 250 to < 1000	≥ 10 sec
E	Extremely Short Period	≥ 80 to < 250	< 10 sec
S	Short Period	≥ 10 to < 80	< 10 sec
H	High Broad Band	≥ 80 to < 250	≥ 10 sec
B	Broad Band	≥ 10 to < 80	≥ 10 sec
M	Mid Period	> 1 to < 10	
L	Long Period	= 1	
V	Very Long Period	= 0.1	
U	Ultra Long Period	= 0.01	
R	Extremely Long Period	≥ 0.0001 to < 0.001	
P	On the order of 0.1 to 1 day[1]	≥ 0.00001 to < 0.0001	
T	On the order of 1 to 10 days[1]	≥ 0.000001 to < 0.00001	
Q	Greater than 10 days[1]	< 0.000001	
A	Administrative Instrument Channel	variable	NA
O	Opaque Instrument Channel	variable	NA

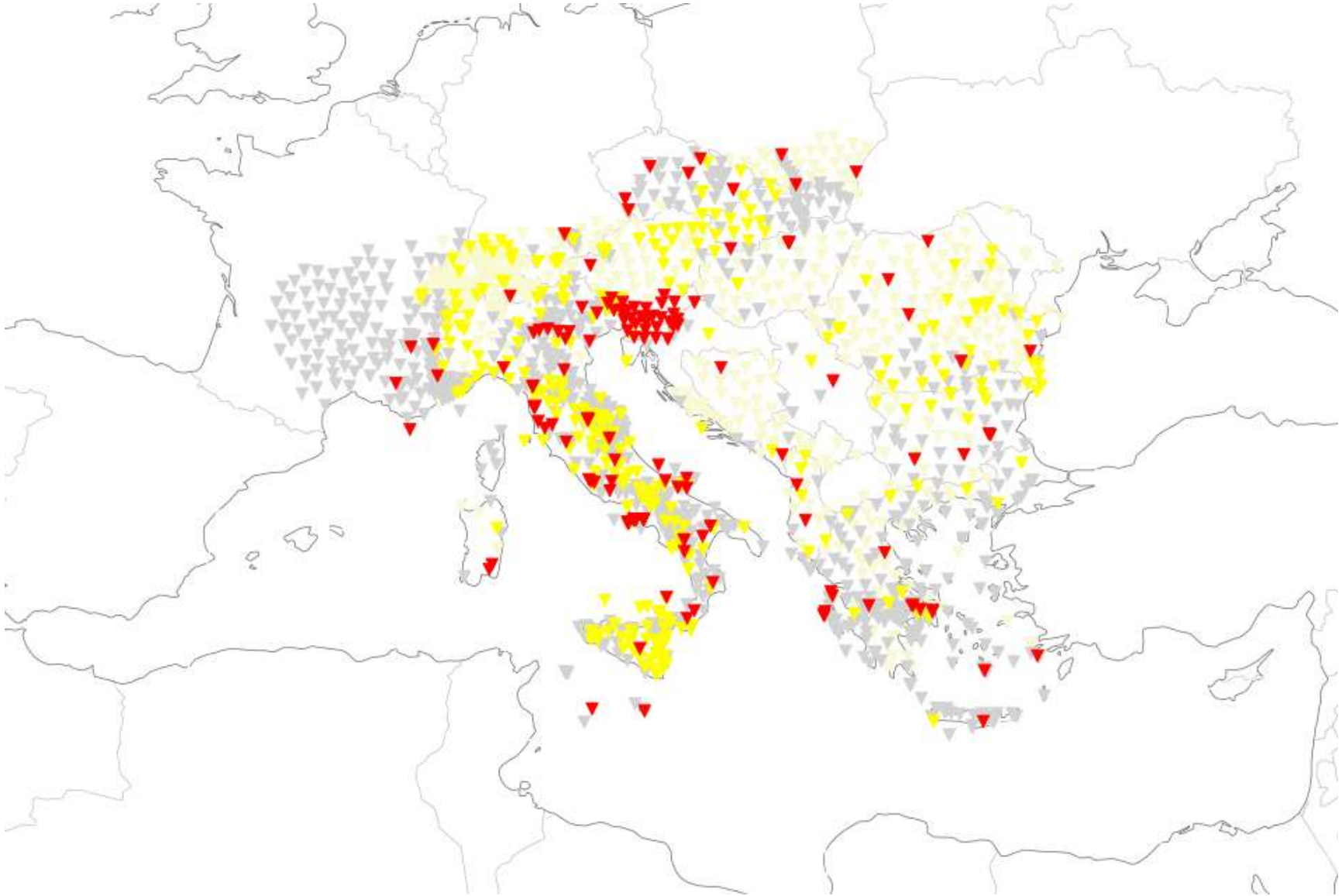
1. These are approximate values. The sample rate should be used for the correct Band Code.

SEED manual

- different corner periods in metadata and the list (14.5 %)
- wrong band code for a given sample rate (1.2 %)
- wrong band code for broad-band/short-period (1.5 %)

total: 17 %

Issue severity levels



notifications
(29.7 %)

warnings
(16.1 %)

errors
(7,6 %)

Report tables prepared …

STATION	VALID_FROM	VALID_T	ISSUE_ID	ISSUE_	ISSUE DESCRIPTION	TESTED	TESTER	INPUT1	INPUT2	STATUS	COMMENTS
7B.A339A.00.LHE	2015-09-02		4.04	5	metadata - not conjugated Poles in the PolesZerosResponseStage	240310	LV:lv_codes	META::240309		found:240310	inventoryP&Z: imag(prod(Poles))!=0 stage1: Poles=
7B.A339A.00.HHE	2015-09-02		4.04	5	metadata - not conjugated Poles in the PolesZerosResponseStage	240310	LV:lv_codes	META::240309		found:240310	inventoryP&Z: imag(prod(Poles))!=0 stage1: Poles=
AC.BCI.00.HHE	2009-12-22		4.07	2	metadata - corner confusion between real and tabled values	240310	LV:lv_codes	META::240309		found:240310	corner period: real: 379.58 s, table: 40.00 s
AC.BCI.00.HHN	2009-12-22		4.07	2	metadata - corner confusion between real and tabled values	240310	LV:lv_codes	META::240309		found:240310	corner period: real: 379.58 s, table: 40.00 s
AC.BCI.00.HHZ	2009-12-22		4.07	2	metadata - corner confusion between real and tabled values	240310	LV:lv_codes	META::240309		found:240310	corner period: real: 379.58 s, table: 40.00 s
AC.BERA.00.HHE	2022-07-19		4	5	metadata response failure	240310	LV:lv_codes	META::240309		found:240310	check_channel: Illegal RESP format
AC.BERA.00.HHN	2022-07-19		4	5	metadata response failure	240310	LV:lv_codes	META::240309		found:240310	check_channel: Illegal RESP format
AC.BERA.00.HHZ	2022-07-19		4	5	metadata response failure	240310	LV:lv_codes	META::240309		found:240310	check_channel: Illegal RESP format
AC.LACI.00.HHE	2022-07-19		4	5	metadata response failure	240310	LV:lv_codes	META::240309		found:240310	check_channel: Illegal RESP format
AC.LACI.00.HHN	2022-07-19		4	5	metadata response failure	240310	LV:lv_codes	META::240309		found:240310	check_channel: Illegal RESP format
AC.LACI.00.HHZ	2022-07-19		4	5	metadata response failure	240310	LV:lv_codes	META::240309		found:240310	check_channel: Illegal RESP format
AC.LSK.00.HHE	2010-01-01		4.07	2	metadata - corner confusion between real and tabled values	240310	LV:lv_codes	META::240309		found:240310	corner period: real: 379.58 s, table: 120.00 s
AC.LSK.00.HHN	2010-01-01		4.07	2	metadata - corner confusion between real and tabled values	240310	LV:lv_codes	META::240309		found:240310	corner period: real: 379.58 s, table: 120.00 s
AC.LSK.00.HHZ	2010-01-01		4.07	2	metadata - corner confusion between real and tabled values	240310	LV:lv_codes	META::240309		found:240310	corner period: real: 379.58 s, table: 120.00 s
AC.PUK.00.HHE	2009-05-29		4.07	2	metadata - corner confusion between real and tabled values	240310	LV:lv_codes	META::240309		found:240310	corner period: real: 40.52 s, table: 30.00 s
AC.PUK.00.HHN	2009-05-29		4.07	2	metadata - corner confusion between real and tabled values	240310	LV:lv_codes	META::240309		found:240310	corner period: real: 40.52 s, table: 30.00 s
AC.PUK.00.HHZ	2009-05-29		4.07	2	metadata - corner confusion between real and tabled values	240310	LV:lv_codes	META::240309		found:240310	corner period: real: 40.52 s, table: 30.00 s
AC.SRN.00.HHE	2024-01-09		4.07	2	metadata - corner confusion between real and tabled values	240310	LV:lv_codes	META::240309		found:240310	corner period: real: 40.52 s, table: 120.00 s
AC.SRN.00.HHN	2024-01-09		4.07	2	metadata - corner confusion between real and tabled values	240310	LV:lv_codes	META::240309		found:240310	corner period: real: 40.52 s, table: 120.00 s
AC.SRN.00.HHZ	2024-01-09		4.07	2	metadata - corner confusion between real and tabled values	240310	LV:lv_codes	META::240309		found:240310	corner period: real: 40.52 s, table: 120.00 s
BS.BLKB..BHE	2012-11-20		4.07	2	metadata - corner confusion between real and tabled values	240310	LV:lv_codes	META::240309		found:240310	corner period: real: 101.94 s, table: 120.00 s
BS.BLKB..BHN	2012-11-20		4.07	2	metadata - corner confusion between real and tabled values	240310	LV:lv_codes	META::240309		found:240310	corner period: real: 101.94 s, table: 120.00 s
BS.BLKB..BHZ	2012-11-20		4.07	2	metadata - corner confusion between real and tabled values	240310	LV:lv_codes	META::240309		found:240310	corner period: real: 101.94 s, table: 120.00 s
BS.BLKB..HHE	2012-11-20		4.07	2	metadata - corner confusion between real and tabled values	240310	LV:lv_codes	META::240309		found:240310	corner period: real: 101.94 s, table: 120.00 s
BS.BLKB..HHN	2012-11-20		4.07	2	metadata - corner confusion between real and tabled values	240310	LV:lv_codes	META::240309		found:240310	corner period: real: 101.94 s, table: 120.00 s
BS.BLKB..HHZ	2012-11-20		4.07	2	metadata - corner confusion between real and tabled values	240310	LV:lv_codes	META::240309		found:240310	corner period: real: 101.94 s, table: 120.00 s
BS.DIM..HHZ	2005-12-02		4.06	4	metadata - channel band confusion in broadband and short-period	240310	LV:lv_codes	META::240309		found:240310	band H -> E (sr: 100.0 Hz, corner period: 1.01 s)
BS.DIM..HHN	2005-12-02		4.06	4	metadata - channel band confusion in broadband and short-period	240310	LV:lv_codes	META::240309		found:240310	band H -> E (sr: 100.0 Hz, corner period: 1.01 s)
BS.DIM..HHE	2005-12-02		4.06	4	metadata - channel band confusion in broadband and short-period	240310	LV:lv_codes	META::240309		found:240310	band H -> E (sr: 100.0 Hz, corner period: 1.01 s)
BS.ELND..BHE	2012-11-20		4.07	2	metadata - corner confusion between real and tabled values	240310	LV:lv_codes	META::240309		found:240310	corner period: real: 101.94 s, table: 120.00 s
BS.ELND..BHN	2012-11-20		4.07	2	metadata - corner confusion between real and tabled values	240310	LV:lv_codes	META::240309		found:240310	corner period: real: 101.94 s, table: 120.00 s
BS.ELND..BHZ	2012-11-20		4.07	2	metadata - corner confusion between real and tabled values	240310	LV:lv_codes	META::240309		found:240310	corner period: real: 101.94 s, table: 120.00 s
BS.ELND..HHE	2012-11-20		4.07	2	metadata - corner confusion between real and tabled values	240310	LV:lv_codes	META::240309		found:240310	corner period: real: 101.94 s, table: 120.00 s

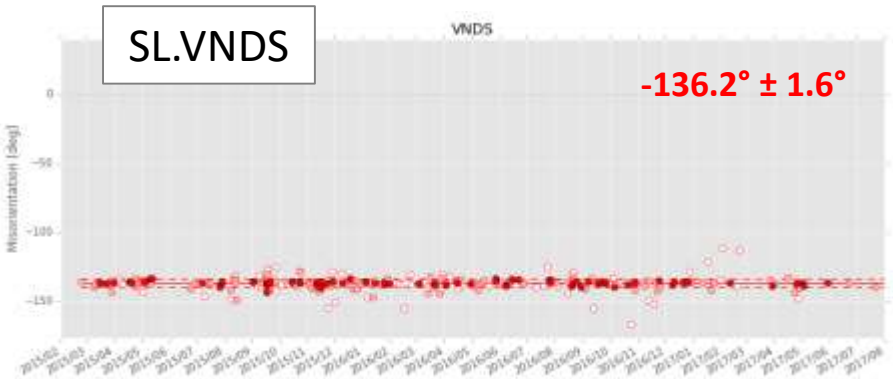
Conclusions

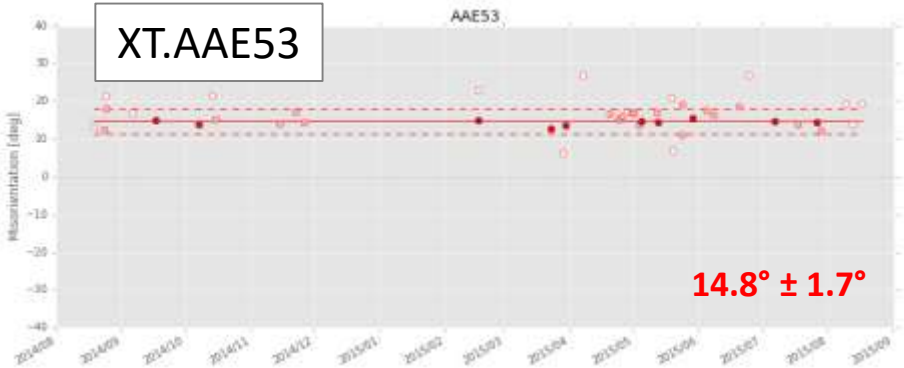
- A non-negligible part of the AdriaArray metadata in EIDA is defective or incorrect
  – errors: 8 %, warnings 16 %

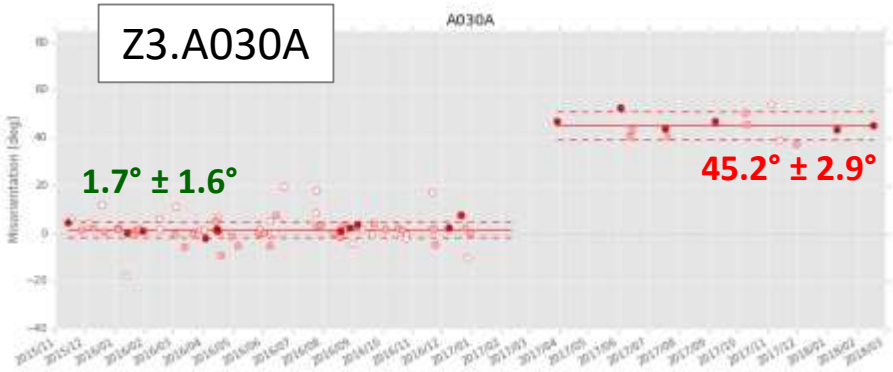
- We identified a wide group of problems in metadata. Next step should be a cooperation with **data providers** (and EIDA) to fix the metadata.
  -> action for AdriaArray - Working Group 3 ‘Data QC’ ?

- What next:
  - check a fit of metadata and data (in channel codes, location codes)
  - sensor orientations
  - amplitude gains

Sensor mis-orientations







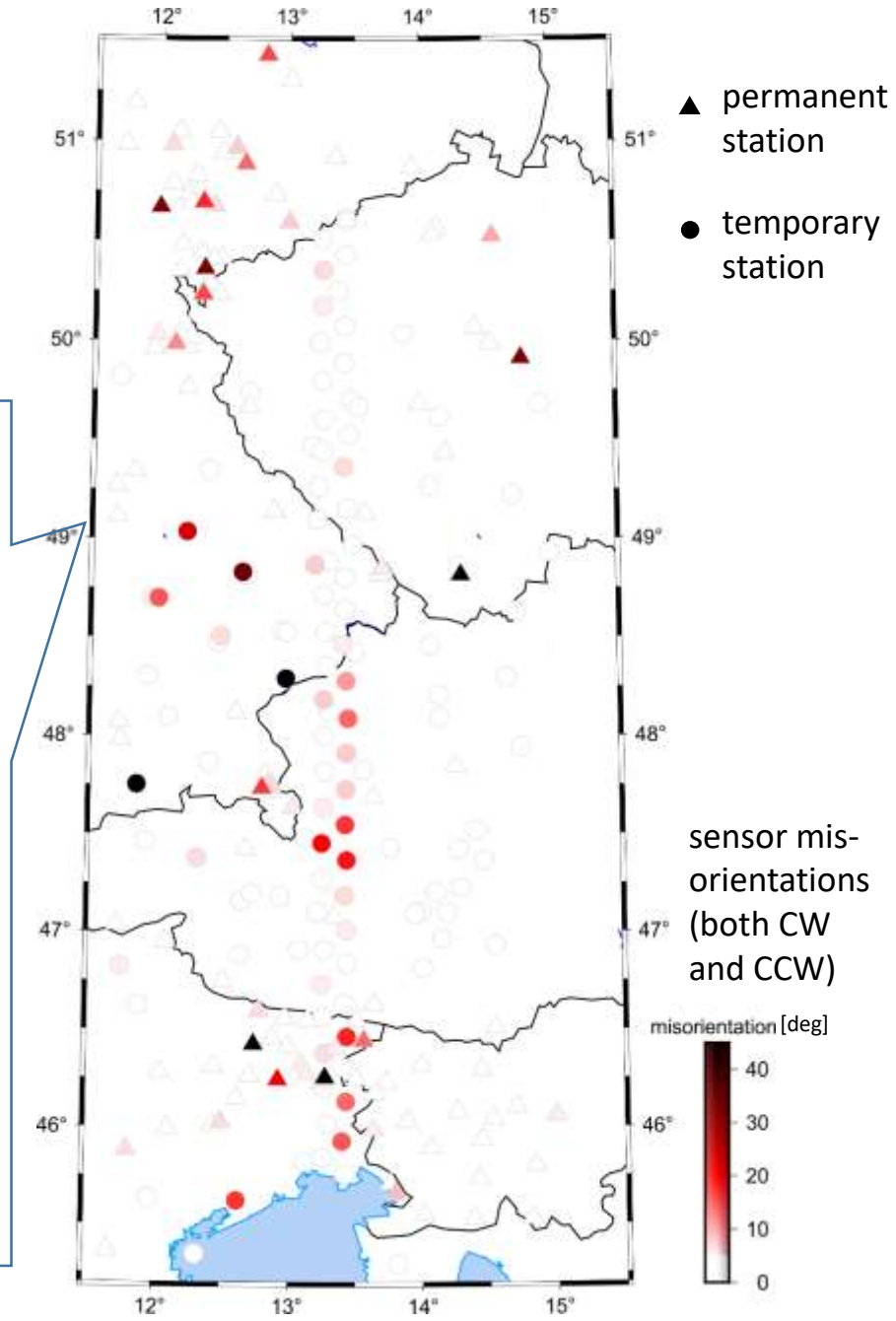
2014-2018: Rayleigh wave method (multi station)

2014-2023: SKS polarization method (single station)

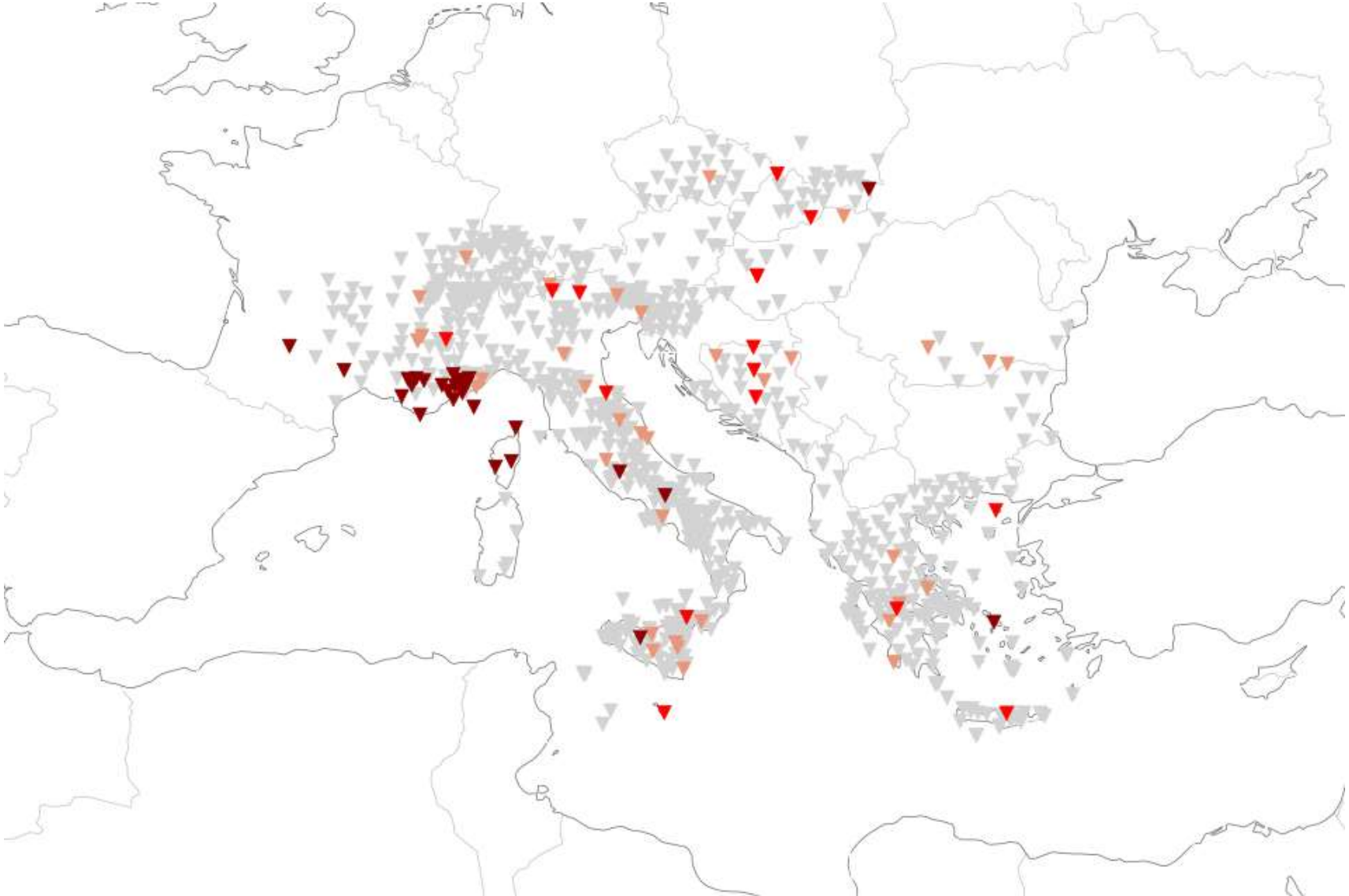
Deviations of sensor orientations from the official orientations given in metadata (from EIDA, 2023-06-05).

Number of stations with sensor mis-orientations:

≥ 5°	77/250	31 %
≥ 10°	29/250	12 %
≥ 20°	12/250	5 %
≥ 30°	9/250	4 %
≥ 45°	5/250	2 %
≥ 90°	3/250	1 %



Saturated medians - map

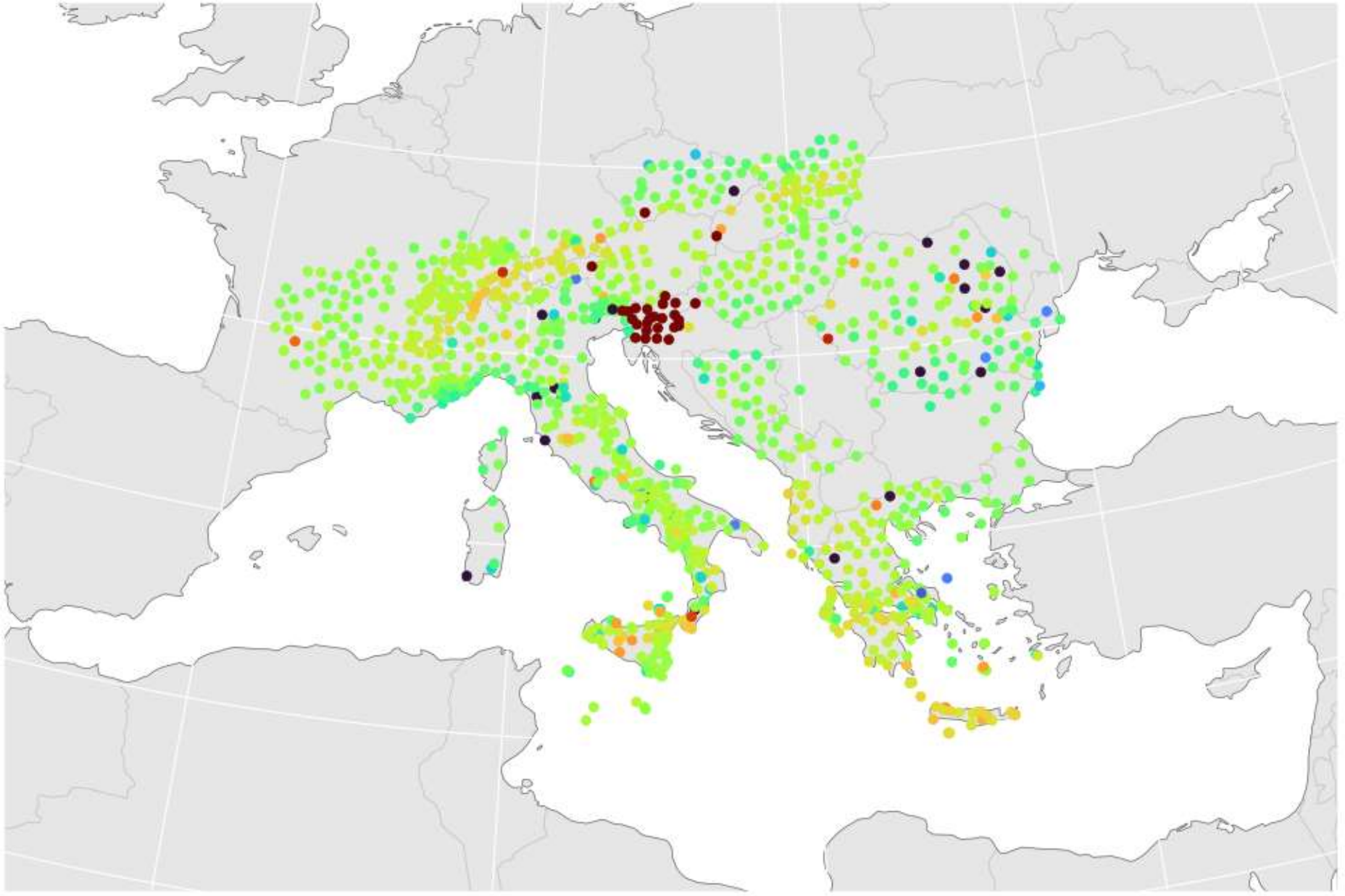


Tested period:
2022-01-01 – 2023-03-21
445 days

Saturation condition:
abs(median) > 1e6

- all stations (861)
- 10+ days: 79/861 9 %
- 100+ days: 43/861 5 %
- 300+ days: 28/861 3 %

Maximal amplitudes from a seismic event



Morocco 6.8

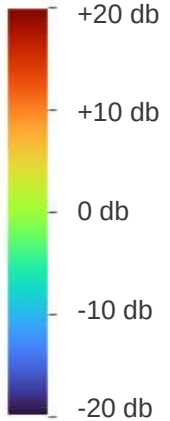
2023-09-08T22:11

N31.055°, W8.389°

depth 19.0 km

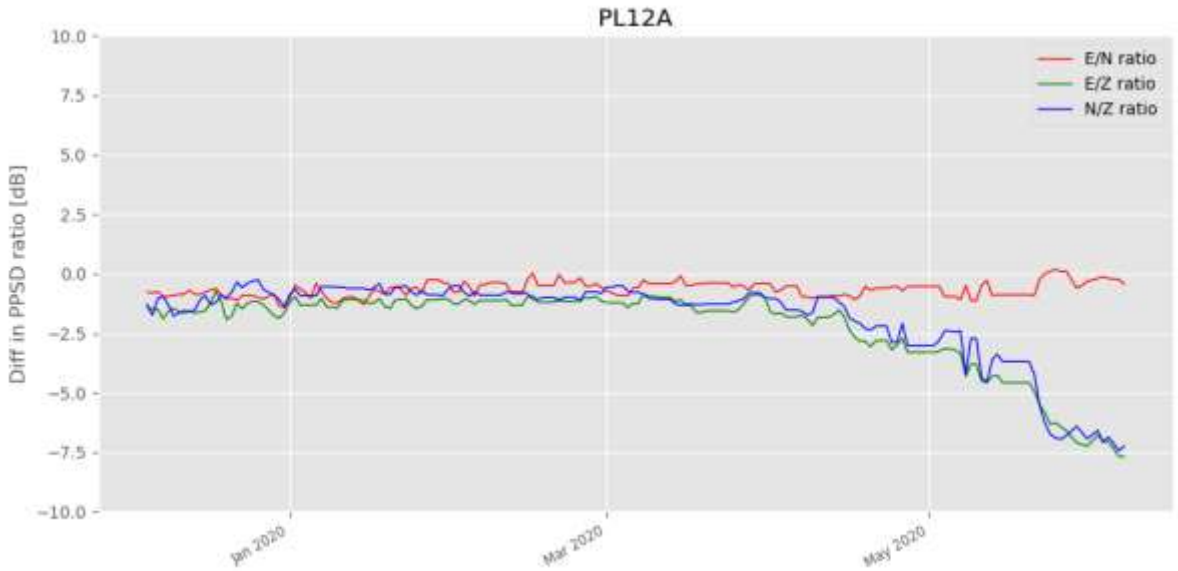
Z component

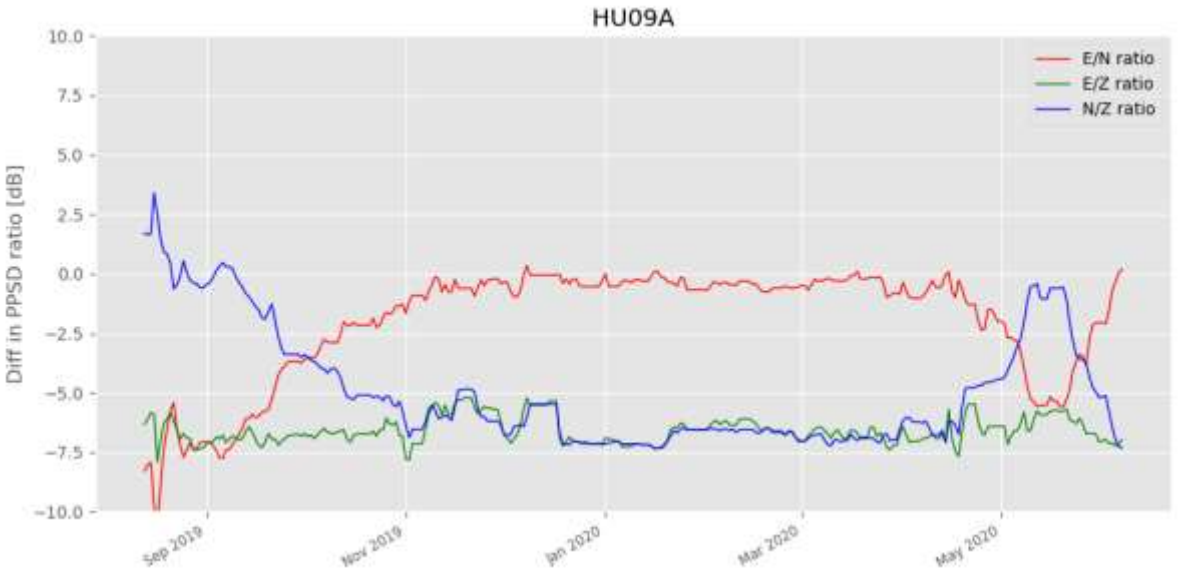
filtered 20-60 s

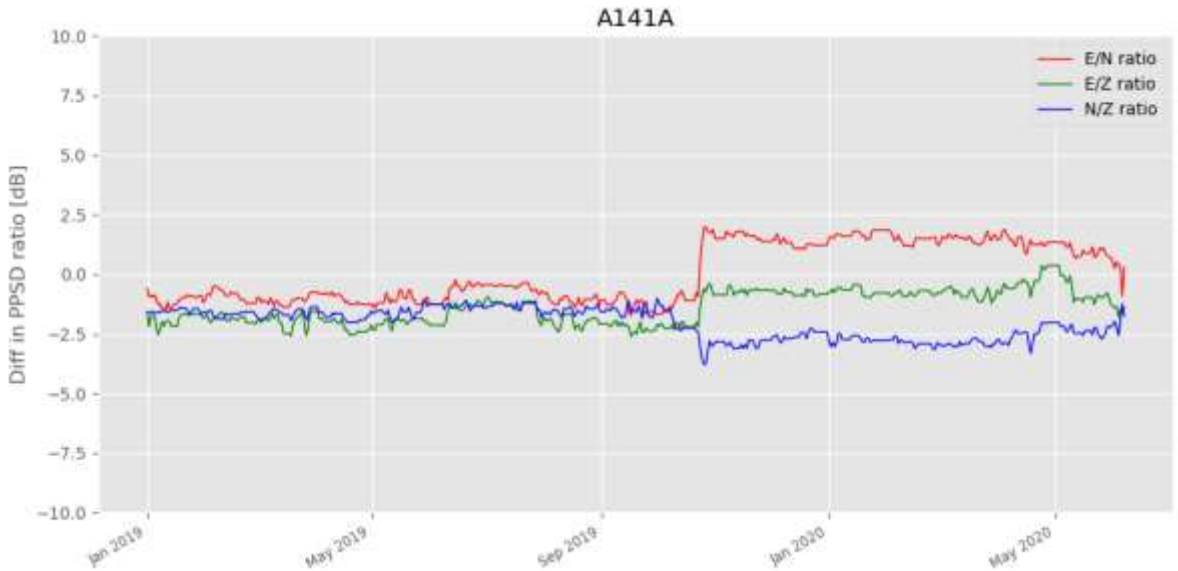


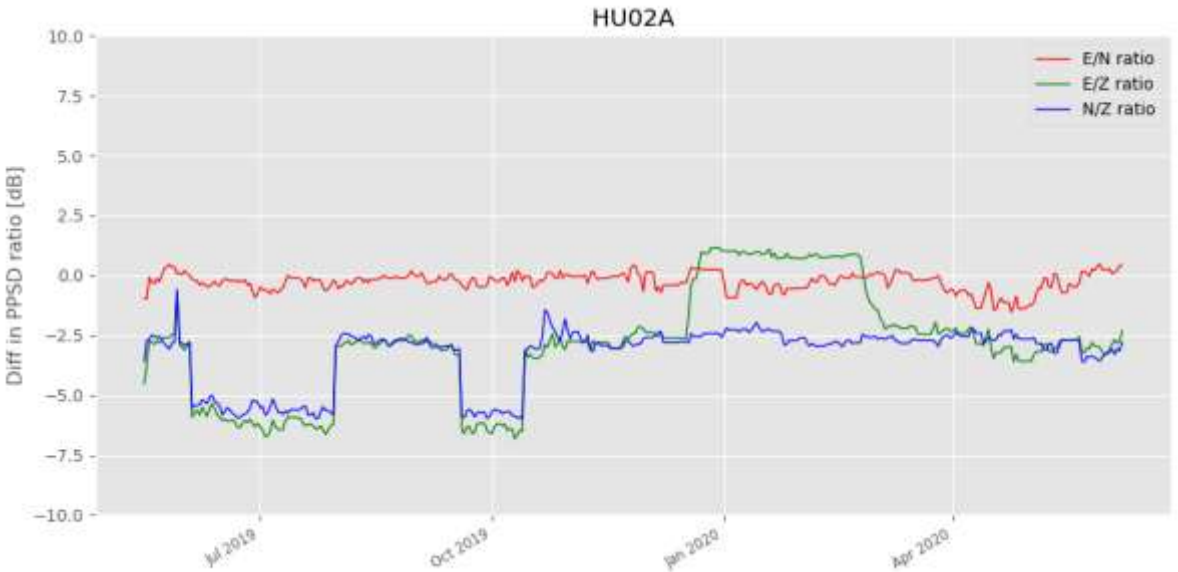
PACASE - amplitude gain issues

- temporal variations of PPSD ratios (for T = 5s)
- global trend subtracted



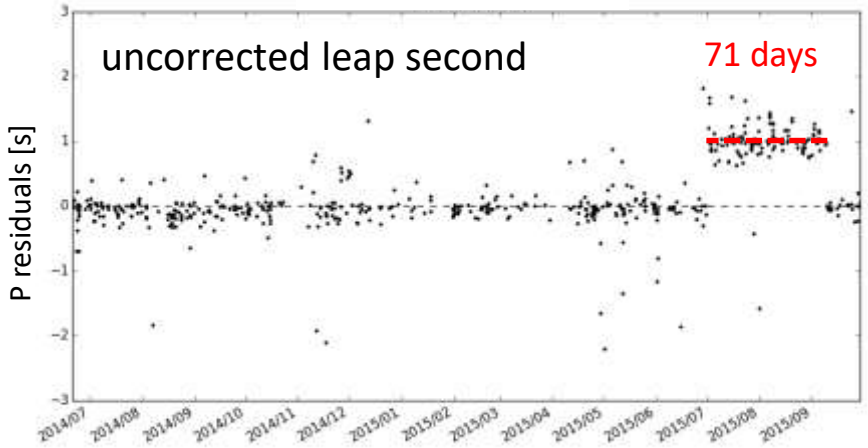




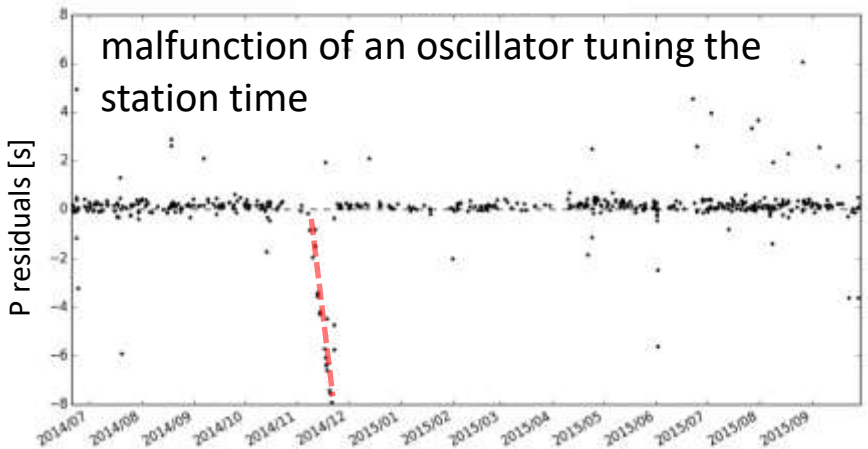


Timing issues - examples

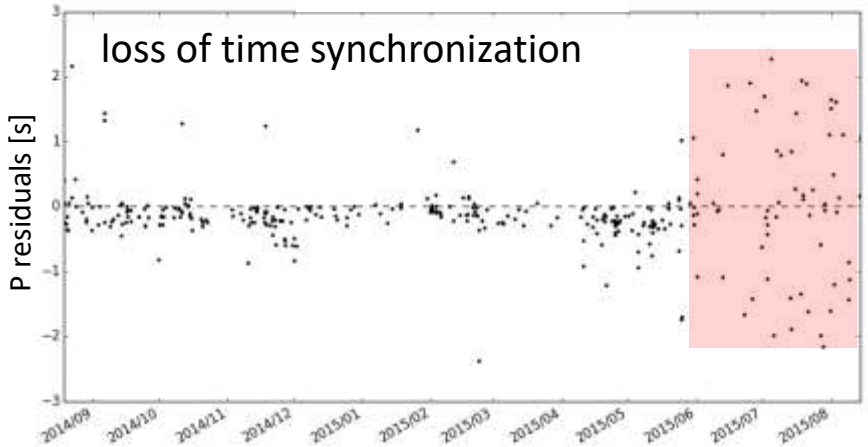
TH.HKWD..HHZ



BW.MGBB..EHZ



XT.AAE21..HHZ



Z3.A147A.00.HHZ