Extensive survey of NM databases with source recommendation list

Pauli Väisänen[1] Ilya Usoskin[1,2] Kalevi Mursula[1]

pauli.vaisanen@oulu.fi

[1] Space Physics and Astronomy Research Unit, University of Oulu, Finland

[2] Sodankylä Geophysical Observatory, (University of Oulu), Finland

Introduction

Neutron monitor (NM) measurements are used to study the variations of galactic cosmic ray (GCR) fluxes.

NM can especially be used to study SEP/flare-induced ground level enhancement (GLE) events and CME or CIR –induced Forbush decreases.

Sources for NM datasets include:

1. Station or institute homepages or other services
2. The Pushkov Institute of Terrestrial Magnetism, Ionosphere and Radiowave Propagation (IZMIRAN)
3. World Data Center for Cosmic Rays (WDCCR)
4. the Neutron Monitor Database (NMDB),

All statistics and results correspond to the situation in 2020

Data repository	Available stations	Recommended sources	Secondary sources
NMDB (1h)	53	29	10
NMDB (revori)	51	3	2
WDCCR	138	59	24
IZMIRAN	81	50	18
Polar Geophys. Inst.	1	1	
Bartol Inst.	8	5	3
Jungfraujoch NM	2	0	2
Lomnický Štit NM	1	1	
Mexico NM	1	0	1
Oulu NM	3	3	
South African stations	5	2	2
Yakutsk + Tixie Bay	2	0	0

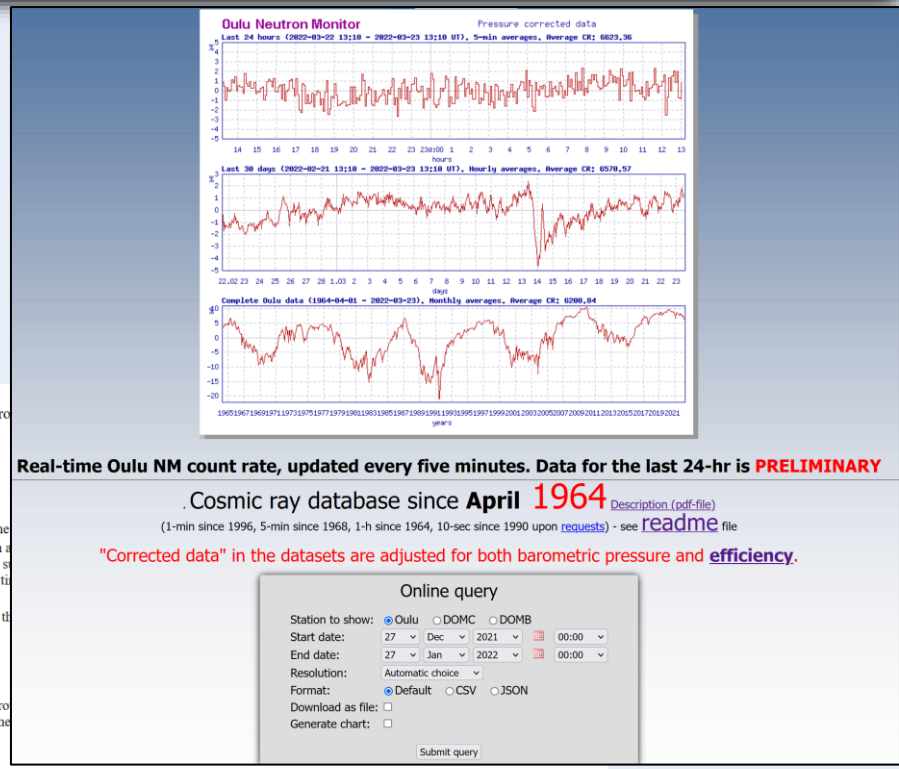
Total unique stations: 147

Overview of sources - Stations

Many NM stations or operating institutes have their own web service or FTP for their data.

Notes:

- Many stations have different systems for obtaining data with individual restrictions.
- In many cases, stations have focused on the NM databases for distributing data, so their own services might be depreciated
- For example, Bartol institution's FTP service is only available up to 2017.



Overview of sources - NMDB

The NMDB was established in 2008 as a part of a European Union funded project (FP7 Program) to create a modern database of NM data, including real-time updates (Mavromichalaki et al., 2011).

Originally, it was built on mostly European NMs, but data from several non-European stations have been added later. In total, NMDB has data folders.

NMDB has three data tables for data:

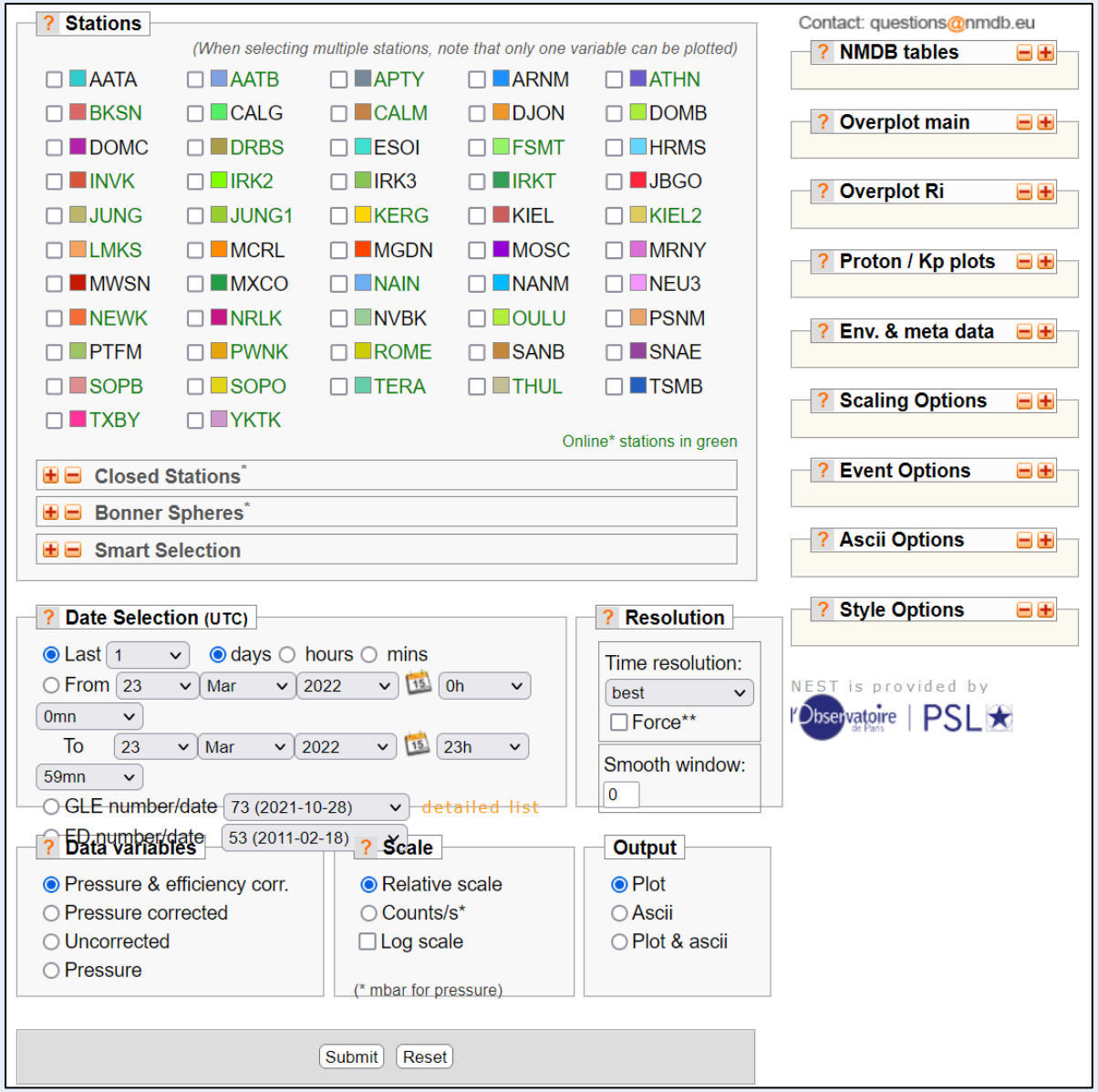
"**ori**" (original) table contains first upload of data to NMDB.

"**revori**" (revised original) contains updated data at the highest resolution, usually 1 min.

"**1h**" (1-hour validated) contains data at 1-hour resolution

Notes:

- Data from these three tables might not be the same!
- When using NEST, the system might automatically decreases the resolution (e.g., from 1 h to 1 month) for too long queries.
- The data table can also change (**even to one with no data!**) when this happens (e.g., from -1 h to –revori).



Overview of sources - WDCCR

The WDCCR started its operation in 1957 (Lincoln & Shea, 1973). It collects pressure-corrected data from NM stations and makes them available online as ASCII files of 1 h time resolution, through an FTP service (now a https service).

WDCCR has three ASCII data formats:

LONGFORMAT
SHORTFORMAT
CARDFORMAT
They all contain the same data.

Notes:

- WDCCR has the most extensive selection of available stations, but in many cases the data might be outdated and/or not corrected for errors.

Index of /WDCCR/files/STATIONS

Name	Last modified	Size	Description
Parent Directory		-	
AHMEDA/	2021-12-22 14:07	-	
ALBUQU/	2014-03-04 11:27	-	
ALERT/	2021-12-22 14:07	-	
ALMA_A/	2021-12-22 14:07	-	
ALMA_B/	2021-12-22 14:07	-	
APATIT/	2021-12-22 14:07	-	
ATHENS/	2021-12-22 14:08	-	
BAGNER/	2014-03-04 11:28	-	
BAKSAN/	2021-12-22 14:08	-	
BARENT/	2021-12-22 14:08	-	
BEIJIN/	2021-12-22 14:08	-	
BEIRUT/	2014-03-04 11:28	-	
BERGEN/	2014-02-06 12:21	-	
BERKEL/	2021-12-22 14:08	-	
BRISBA/	2021-12-22 14:08	-	
BUENOS/	2021-12-22 14:08	-	
BURE/	2021-12-22 14:08	-	
CALGAR/	2021-12-22 14:08	-	
CAPE_H/	2014-02-06 12:22	-	
CAPE_S/	2021-12-22 14:08	-	
CASEY/	2014-03-04 11:29	-	
CHACAL/	2021-12-22 14:08	-	
CHICAG/	2021-12-22 14:08	-	
CHURCH/	2021-12-22 14:08	-	
CLIMAX/	2021-12-22 14:08	-	
COLLEG/	2021-12-22 14:08	-	
CORDOB/	2021-12-22 14:09	-	
DALLAS/	2021-12-22 14:09	-	

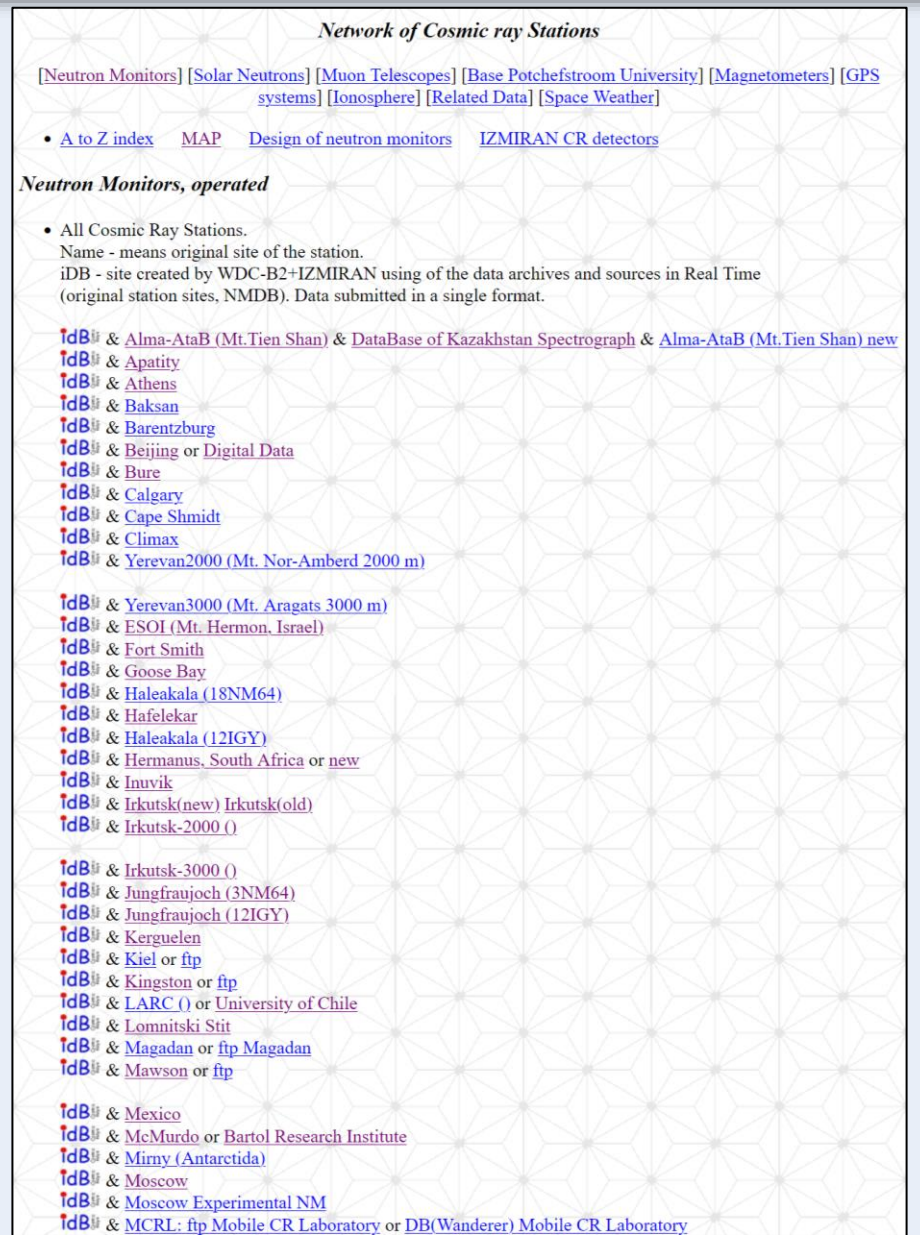
Overview of sources - Izmiran

The Pushkov Institute of Terrestrial Magnetism, Ionosphere, and Radiowave Propagation (IZMIRAN) of the Russian Academy of Sciences was established in 1939 and offers data for most Russian NM stations, but it also offers data from other NM stations. Altogether, IZMIRAN has data for 81 NMs.

You can access data via the idB-service and/or via html address commands. (see Väisänen et al. 2021).

Notes:

- Izmiran often has undocumented changes or corrections to the data. They often seem like improvements to the data, but without documentation it is hard to say if they introduce any bias or other effects.



Network of Cosmic ray Stations

[Neutron Monitors] [Solar Neutrons] [Muon Telescopes] [Base Potchefstroom University] [Magnetometers] [GPS systems] [Ionosphere] [Related Data] [Space Weather]

- A to Z index MAP Design of neutron monitors IZMIRAN CR detectors

Neutron Monitors, operated

- All Cosmic Ray Stations.
  Name - means original site of the station.
  iDB - site created by WDC-B2+IZMIRAN using of the data archives and sources in Real Time (original station sites, NMDB). Data submitted in a single format.

idB & Alma-AtaB (Mt.Tien Shan) & DataBase of Kazakhstan Spectrograph & Alma-AtaB (Mt.Tien Shan) new
idB & Apatity
idB & Athens
idB & Baksan
idB & Barentzburg
idB & Beijing or Digital Data
idB & Bure
idB & Calgary
idB & Cape Shmidt
idB & Climax
idB & Yerevan2000 (Mt. Nor-Amberd 2000 m)

idB & Yerevan3000 (Mt. Aragats 3000 m)
idB & ESOI (Mt. Hermon, Israel)
idB & Fort Smith
idB & Goose Bay
idB & Haleakala (18NM64)
idB & Hafelekar
idB & Haleakala (12IGY)
idB & Hermanus, South Africa or new
idB & Inuvik
idB & Irkutsk(new) Irkutsk(old)
idB & Irkutsk-2000 ()

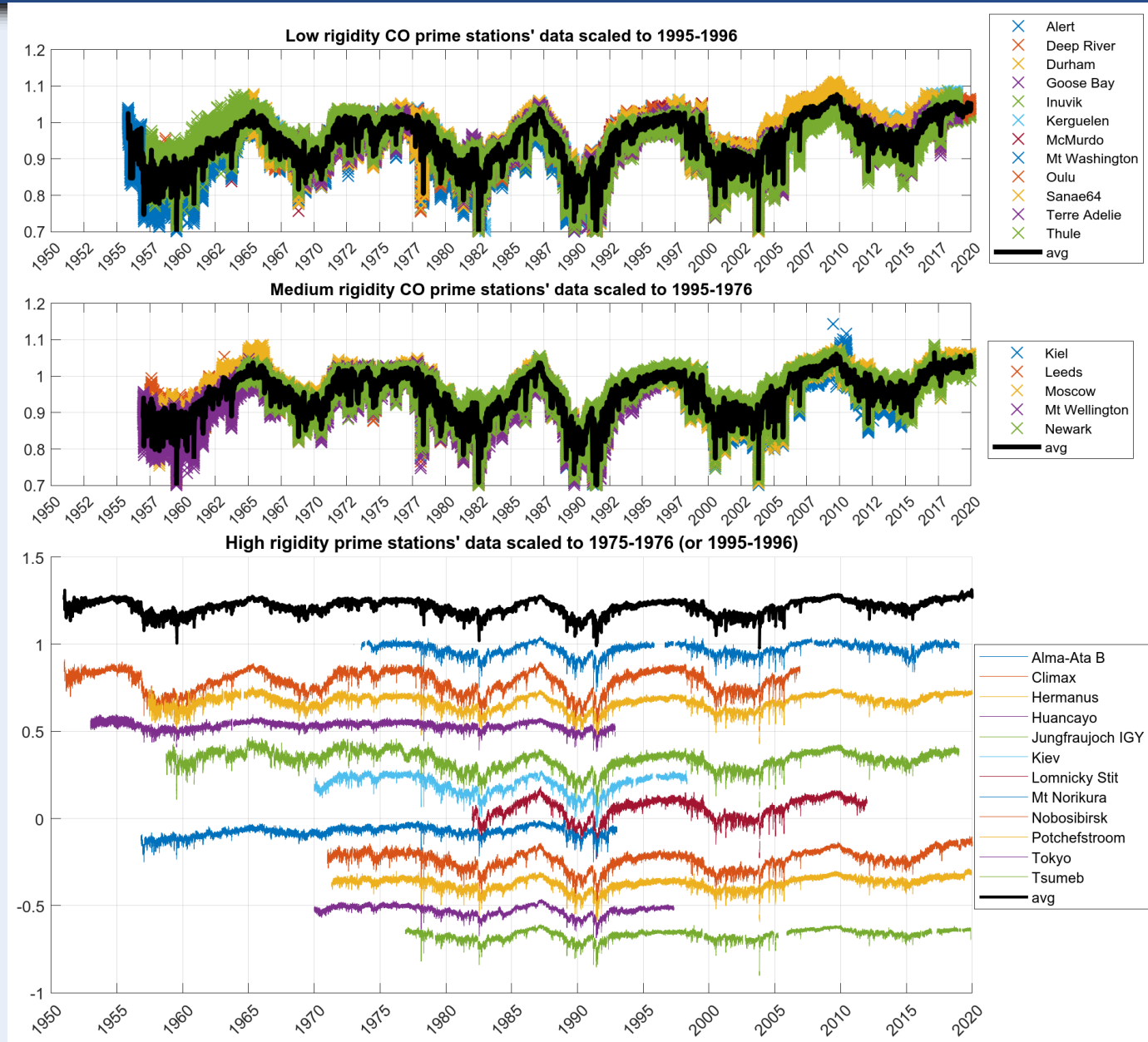
idB & Irkutsk-3000 ()
idB & Jungfraujoch (3NM64)
idB & Jungfraujoch (12IGY)
idB & Kerguelen
idB & Kiel or ftp
idB & Kingston or ftp
idB & LARC () or University of Chile
idB & Lomnitski Stit
idB & Magadan or ftp Magadan
idB & Mawson or ftp

idB & Mexico
idB & McMurdo or Bartol Research Institute
idB & Mirny (Antarctida)
idB & Moscow
idB & Moscow Experimental NM
idB & MCRL: ftp Mobile CR Laboratory or DB(Wanderer) Mobile CR Laboratory

Prime data criteria

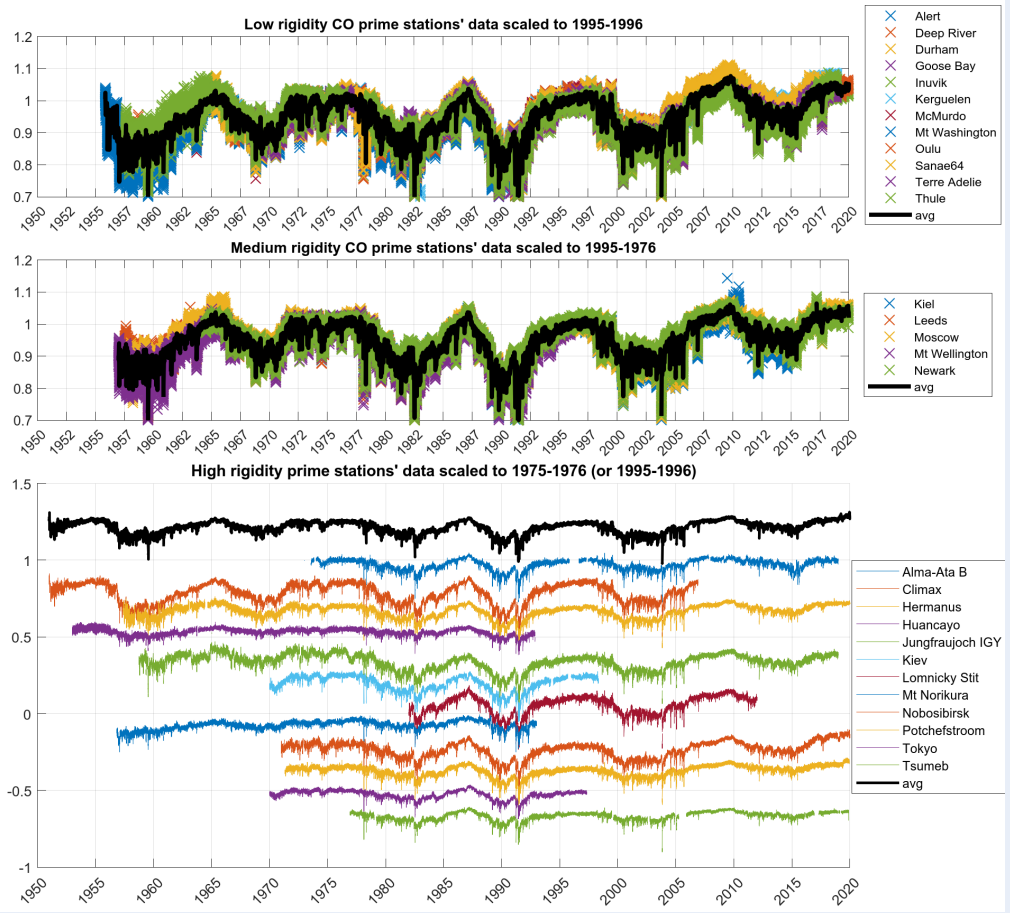
- Instead of just subjectively choosing data from "well-known" NMs, we built specific criteria for the prime data:

1. Data is normalized to years 1995-1996 (or 1975-1976) so that the median for those years is 1. After removing GLE's (GLE list gle.oulu.fi) and outliers (5-point movmedian filter), the **data must have at least 20 years of data**.
2. The ratio of the max and min values of the data must not exceed 2. This means that **variation cannot exceed +-33%.** This step excludes data with notable steps.
3. We manually check for any remaining obvious errors, steps or drift in the data. **Minor and easily fixed problems were corrected**, otherwise dataset is excluded.
4. For prime candidates with multiple possible sources, we use the station with the most available data.

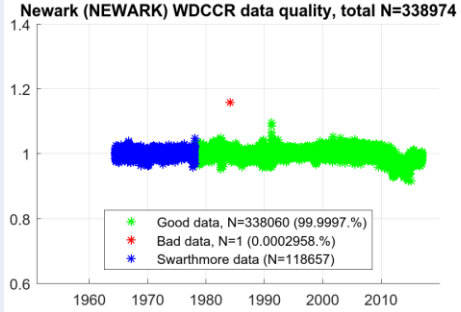


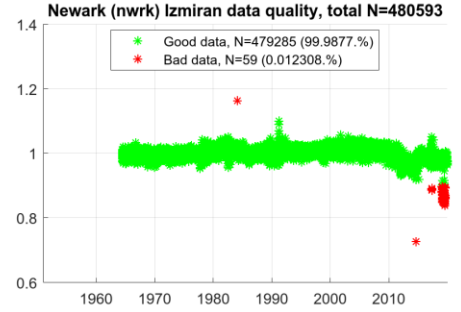
Determining which source to use

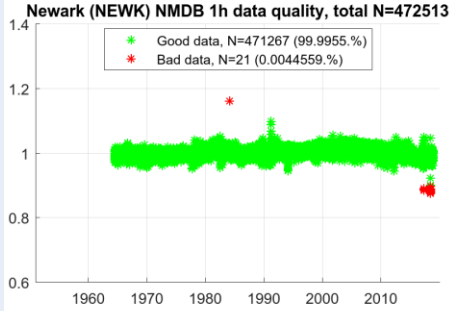
1. Compile a list of long-lived, stable stations to generate a "prime" dataset for different rigidities.

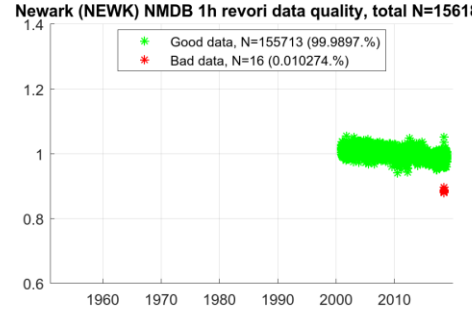


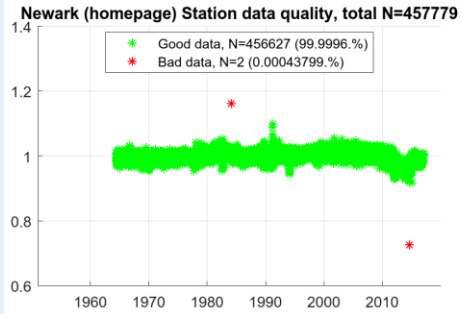
2. Compare each data source of each station to the corresponding prime dataset.











3. Select the source with best "good data" coverage and/or otherwise stable dataset as the **recommended source**.

In Newark's case, we select NMDB1h, even though the coverage is not the best.

Recommendation list

Ahmedabad	4	**Climax**	**4**	Herstmonceux	3	Lincoln	3	**Newark**	**4**	Sverdlovsk	2
Albuquerque	3	College	3	Hobart	3	Lindau_IGY	3	**Nobosibirsk**	**2**	Sydney	3
Alert	**2**	Cordoba	3	**Huancayo**	**4**	Lindau_NM64	3	Nor-Amberd	4	Syowa	3
Alma-Ata A	2	Daejeon	4	**Inuvik**	**2**	**Lomnický Štit**	**1**	Norilsk	2	Tashkent	2
Alma-Ata B	**4**	Dallas	3	Invercargill	3	London	3	Northfield	3	Tbilisi	2
Alma-Ata C	2	Darwin	3	Irkutsk	2	Magadan	2	Ottawa	2	**Terre Adelie**	**4**
Apatity	1	**Deep River**	**2**	Irkutsk 2	2	Makapuu_Pt	3	**Oulu**	**1**	Thailand	4
Aragats	4	Denver	3	Irkutsk 3	2	Mawson	2	Peawanuck	1	**Thule**	**4**
Athens	4	Dome B	1	Jang Bogo	5	**McMurdo**	**1**	Pic du Midi	2	Tibet	4
Bagneres	3	Dome C	1	**Jungfraujoch IGY**	**4**	Mexico	3	**Potchefstroom**	**1**	Tixie Bay	2
Baksan	2	Dourbes	4	Jungfraujoch NM64	4	Mina Aguilar	3	Prague	3	**Tokyo**	**2**
Barentsburg	2	**Durham**	**2**	Kampala	3	Mirny	4	Predigtstuhl	3	**Tsumeb**	**4**
Beijin	2	Ellsworth	3	**Kerguelen**	**4**	Mobile CR Laboratory	2	Resolute Bay	3	Uppsala	3
Beirut	3	ESOISR	2	Khabarovsk	3	Morioka	3	Rio De Janeiro	3	Ushuaia	3
Berkeley	3	Fort Smith	5	**Kiel**	**4**	**Moscow**	**2**	Rome	2	Utrecht	3
Brisbane	3	Freiburg	3	Kiel 2	4	Moscow experimental	2	**Sanae64**	**2**	Weissenau	3
Buenos Aires	3	Fukushima	3	**Kiev**	**3**	**Mt Norikura**	**2**	Sanae80	4	Wellington	3
Bure	2	Goettingen	3	Kingston	2	**Mt Washington**	**2**	Santiago	2	Victoria	3
Calgary	2	**Goose Bay**	**2**	Kiruna	3	**Mt Wellington**	**2**	Seoul	3	Wilkes	3
CALM	5	Hafelekar	2	Kodaikanal	3	Munchen	3	Simferopol	3	Vostok	2
Cape Schmidt	2	Haleakala_IGY	2	Kuhlungsborn	3	Murchison Bay	3	South Pole	1	Yakutsk	2
Casey	3	Haleakala_SM	2	Kula	3	Murmansk	3	South Pole Bare	4	Zugspitze	4
Chacaltaya	3	Halle	3	Lae	3	Nain	1	Sulfur Mt IGY	3		
Chicago	2	Heiss Is	3	Larc	2	Nederhorst	3	Sulfur Mt NM64	2		
Churchill	2	**Hermanus**	**1**	**Leeds**	**2**	Neumayer 3	4	Swarthmore	2		

List of recommended data sources for each station.

1=Station homepage,
2=IZMIRAN,
3=WDCCR,
4=NMDB1h,
5=NMDBrevori.

Prime stations are in **bold**.

Station info list

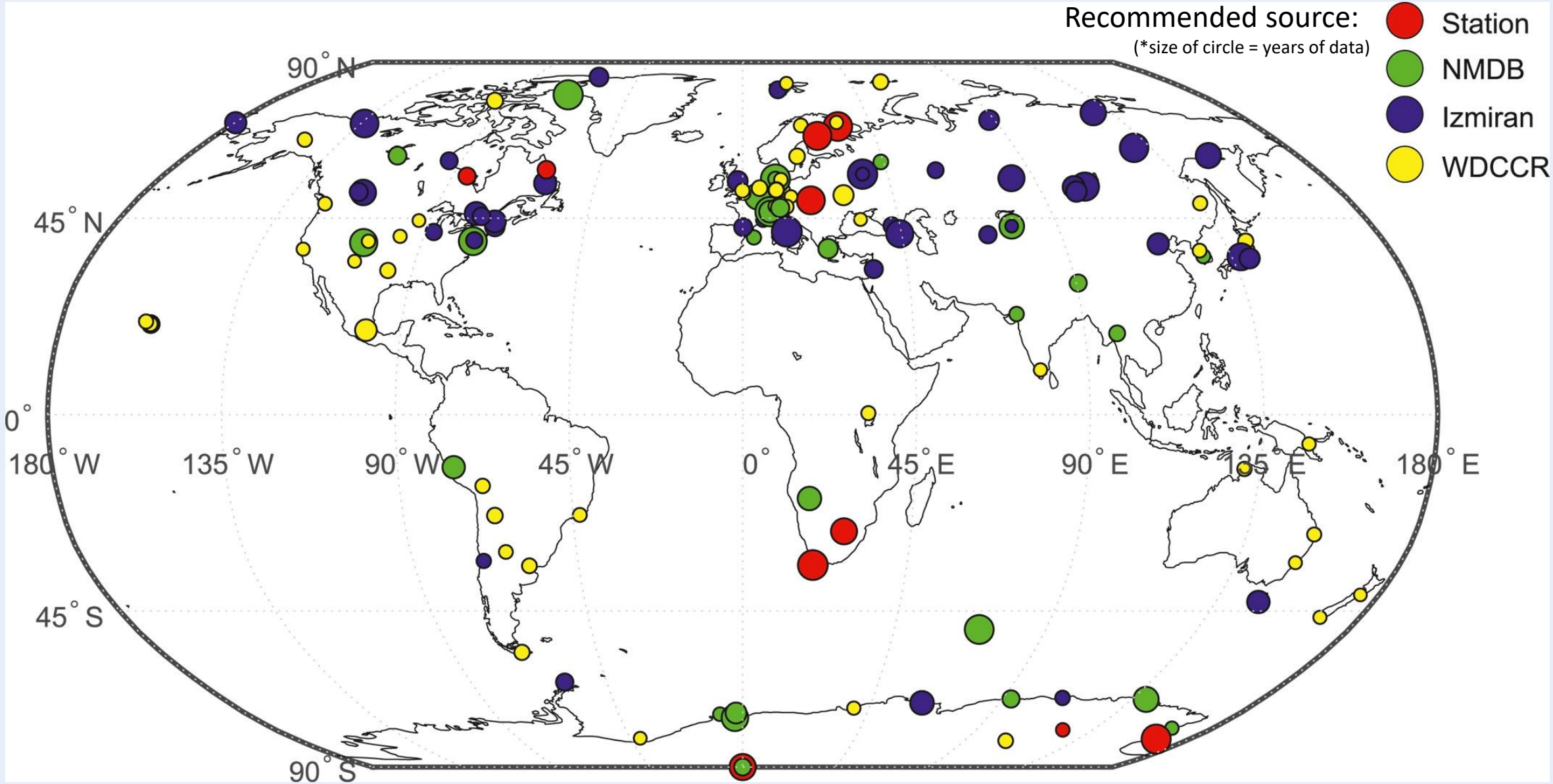
- We also compiled an extensive spreadsheet with station information, acronyms, source links, lat&lon etc.
- Available as a supplement in the article

Table S5. The data table named NMStationInfoList.csv contains metadata, coverage and data quality information. It includes the following 29 columns:

1. Station name	16. Geographical latitude
2. Other name(s)	17. Geographical longitude
3. Detector type	18. Altitude of the NM location
4. Start year - End year	19. Geomagnetic Cut-off rigidity
5. WDCCR acronym	20. Number of "Good" data points in WDCCR
6. NMDB acronym	21. Fraction of "Good" data points in WDCCR
7. IZMIRAN acronym	22. Number of "Good" data points in IZMIRAN
8. URL of Station homepage	23. Fraction of "Good" data points in IZMIRAN
9. Number of available data sources	24. Number of "Good" data points in NMDB-1h
10. Data points (1h) in WDCCR data	25. Fraction of "Good" data points in NMDB-1h
11. Data points (1h) in NMDB-1h data	26. Number of "Good" data points in NMDB-revori
12. Data points (1h) in NMDB-revori data	27. Fraction of "Good" data points in NMDB-revori
13. Data points (1h) in IZMIRAN data	28. Number of "Good" data points in Station
14. Data points (1h) in Station data	29. Fraction of "Good" data points in Station
15. Maximum coverage in years	30. Maximum coverage by "Good" data in years

WDCCR	NMDB	Izmiran	Data homepage	# of data source	WDCCR hours	NMDB1h hours	NMDBori hours	NMDBrevori hours	Izmiran hours	Station hours	Max length [y	Mean length, yea	Latitud	Longitud	Altitude	Cut-off Rigidity	WDCCR good dat
AHMEDA	AHMD			2	61798	46714	0	0	0	0	7.05	2.06	23.01	72.61	0	15.94	39214
ALBUQU				1	12706	0	0	0	0	0	1.45	0.24	35.08	-106.62	1567	4.47	0
ALERT		alrt		2	194491	0	0	0	194491	0	22.20	7.40	82.50	-62.33	57	0.00	194171
ALMA_A	AATA	aata		3	352964	0	0	0	352275	0	40.29	13.42	43.25	76.92	806	6.61	173047
ALMA_B	AATB	aatb		3	379138	368289	311357	310594	379225	0	43.29	33.27	43.14	76.92	3340	6.61	348080
		aatc		1	0	0	0	0	5031	0	0.57	0.10	43.18	76.92	1670	6.61	0
APATIT	APTY	apty	http://pgia.ru/data/	4	504204	161587	161564	161564	506535	438154	57.82	36.79	67.55	33.33	177	0.57	439958
EREVN3	ARNM	yrv3		3	67612	36956	36635	36637	68610	0	7.83	4.69	58.47	44.17	3200	7.10	20784
ATHENS	ATHN	athn	http://cosray.phys	4	159507	153086	151698	151588	217281	0	24.80	15.85	37.98	23.78	260	8.53	81679
BAGNER				1	2663	0	0	0	0	0	0.30	0.05	43.01	0.02	550	5.45	0
BAKSAN	BKSN	bksn		3	133949	80215	80350	80216	134018	0	15.30	9.68	43.28	42.69	1700	5.60	71941
BARENT	BRBG	brbg	http://pgia.ru/data/	4	122147	0	0	0	137234	0	15.67	4.93	78.06	14.22	51	0.00	121938
BEIJIN		bjng		2	277231	0	0	0	277251	0	31.65	10.55	39.08	116.26	48	10.00	230728
BEIRUT				1	2494	0	0	0	0	0	0.28	0.05	33.90	35.47	15	10.42	2367
BERKEL				1	18752	0	0	0	0	0	2.14	0.36	37.87	-122.27	70	4.54	0
BRISBA				1	47922	0	0	0	0	0	5.47	0.91	-27.43	153.08	0	7.21	20193
BUENOS				1	70668	0	0	0	0	0	8.07	1.34	-34.60	-58.48	0	10.63	42653
BURE	BURE	bure		3	35342	0	0	0	53878	0	6.15	1.70	44.63	5.91	2252	5.00	33776
CALGAR	CALG	calg		3	350062	0	250972	250972	406849	0	46.44	23.95	51.08	-114.13	1128	1.08	175719
	CALM			1	0	49238	54312	54312	0	0	6.20	3.00	40.56	3.16	708	6.95	0
CAPE_S		caps		2	244578	0	0	0	275021	0	31.40	9.89	68.55	180.32	0	0.45	221845
CASEY				1	14935	0	0	0	0	0	1.70	0.28	-66.28	110.53	0	0.01	14935
CHACAL				1	62001	0	0	0	0	0	7.08	1.18	-16.32	-68.15	5200	13.10	14649
CHICAG		chgo		2	119902	0	0	0	119903	0	13.69	4.56	41.83	-87.67	200	1.72	51049
CHURCH		chur		2	140075	0	0	0	140076	0	15.99	5.33	58.75	-94.08	39	0.21	113973
CLIMAX	CLMX	clmx		3	444236	459934	0	0	446710	0	52.50	25.70	39.37	-106.18	3400	2.99	419189
COLLEG				1	59890	0	0	0	0	0	6.84	1.14	64.08	-147.83	0	0.54	42601
CORDOB				1	39025	0	0	0	0	0	4.45	0.74	-31.42	-64.19	434	11.45	31816
	DJON			1	0	46565	46597	46368	0	0	5.32	2.65	36.24	127.22	200	11.22	0
DALLAS				1	86324	0	0	0	0	0	9.85	1.64	32.98	-96.73	208	4.35	77387
DARWIN				1	56649	0	0	0	0	0	6.47	1.08	-12.43	130.87	30	14.09	0

Map of all NM stations



Summary

- Data from 147 stations were analysed to create a list of recommended sources.
- NM datasets from different sources often have different versions, leading to discrepancies in the data and affecting reproducibility and reliability of studies.
- Users need to be careful when selecting which data sources to use.
- The recommendation list gives a good basis for data selection, but best selection depends on usage of the data and possible corrections employed.
- Manual inspection and corrections to the selected data are recommended, depending on usage.
- These results are based on a specific version of datasets at time of writing. When databases do updates or changes, the validity of the results decreases.
- An initiative to fix the highlighted problems and preserving, improving and documenting the data is needed. A new EU or other project?

Thank you! Questions?

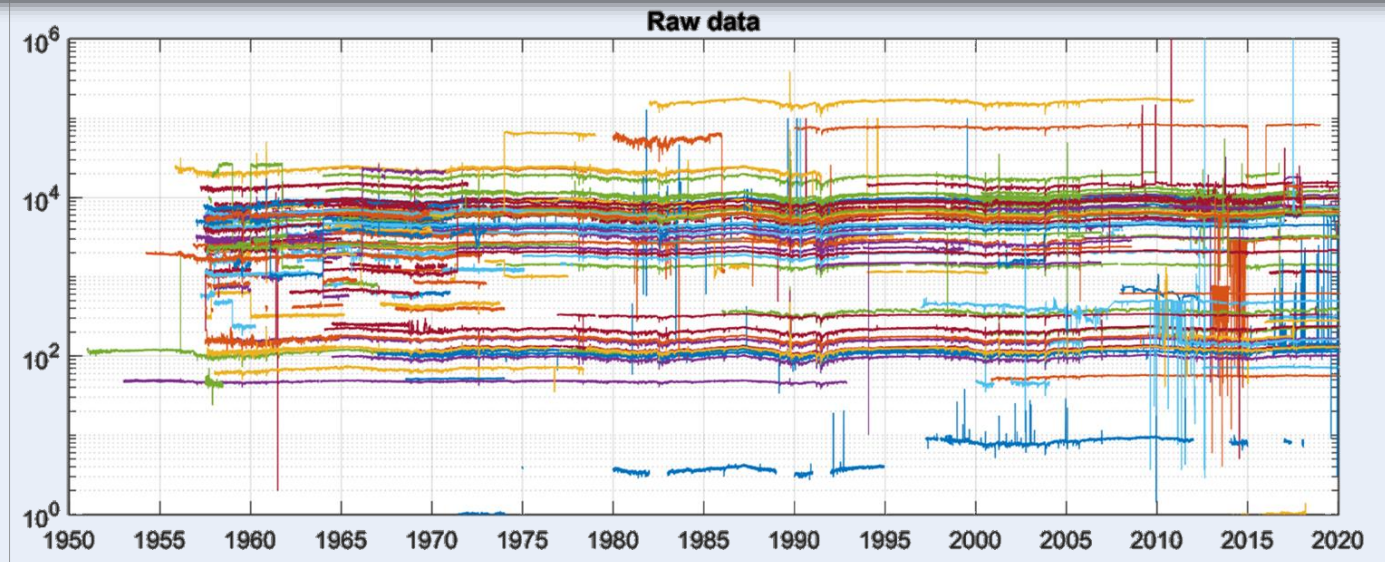
Reference: Väisänen, P., Usoskin, I., & Mursula, K. (2021). Seven decades of neutron monitors (1951–2019): Overview and evaluation of data sources. *Journal of Geophysical Research: Space Physics*, 126, e2020JA028941. https://doi.org/10.1029/2020JA028941

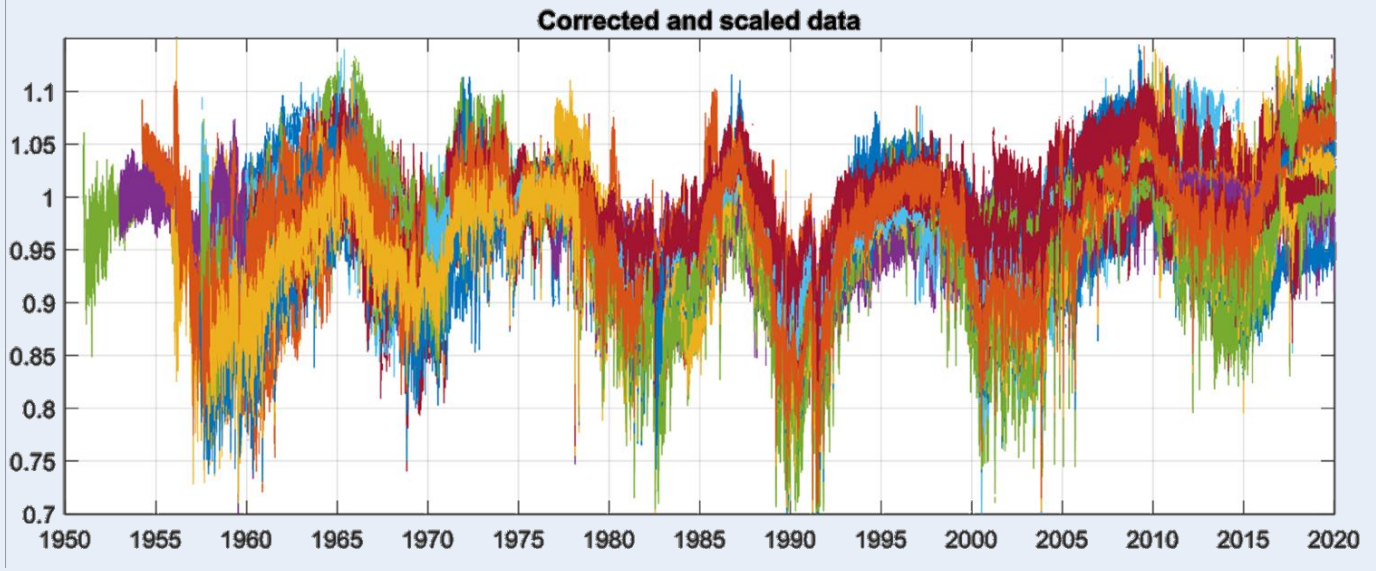
Compilation of the recommended datasets

Come see the poster

"Compiled "multi-NM" recommended dataset of global NM network" #32

to see what the full dataset looks like

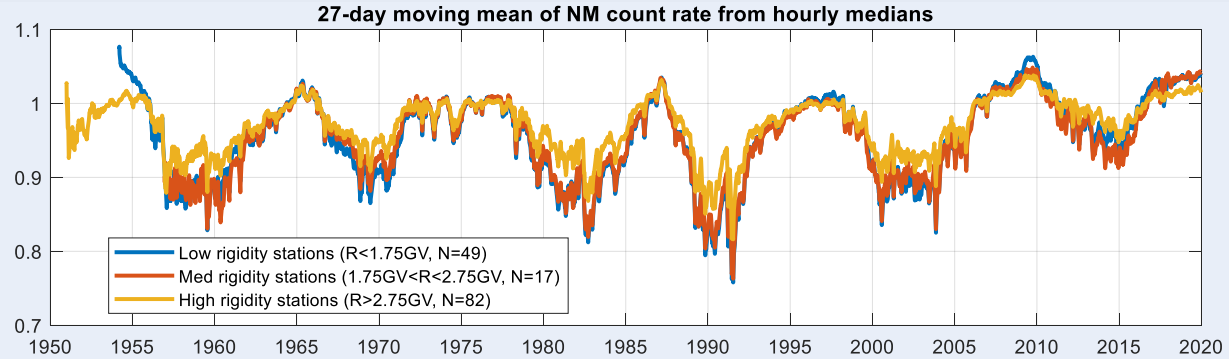


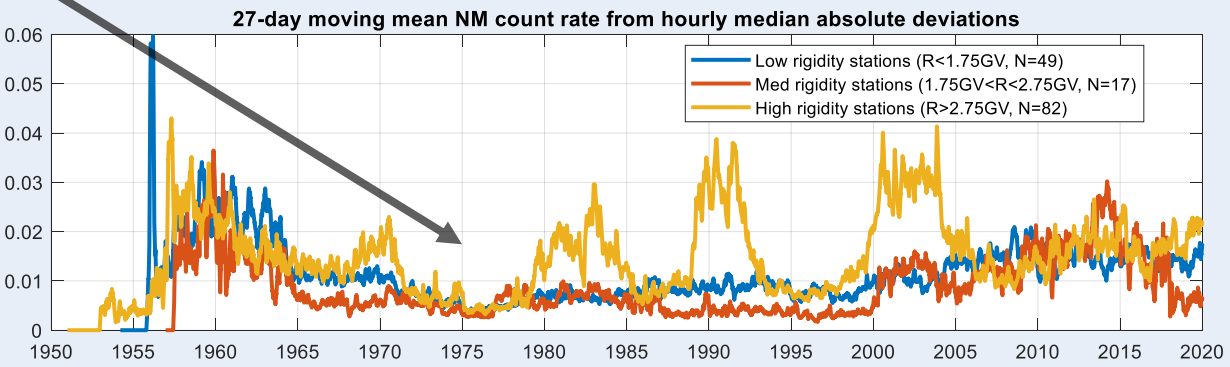


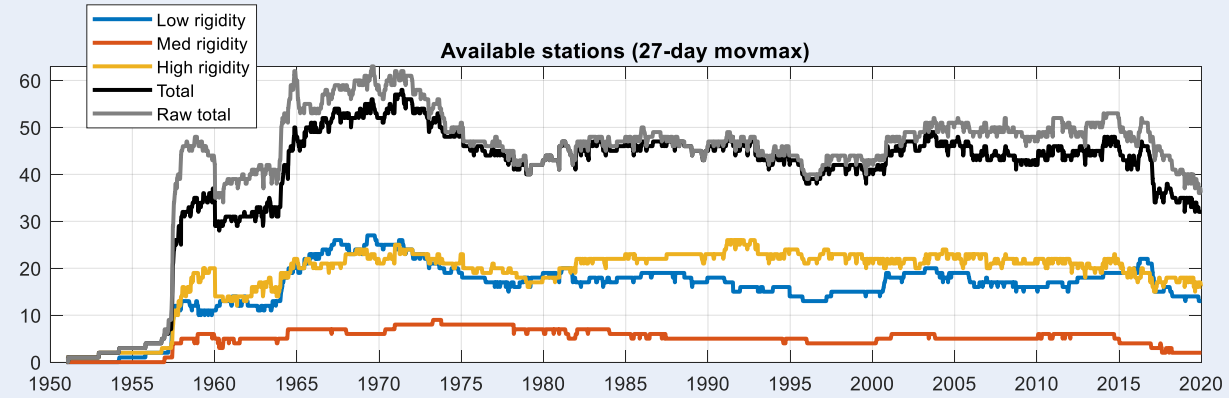
All data combined

Low deviation in 1975-1977 is no coincidence:

Datasets were scaled to those years. If there are any drifts or other similar effects in the datasets, the deviations gets larger the further away from the scaling years we are.





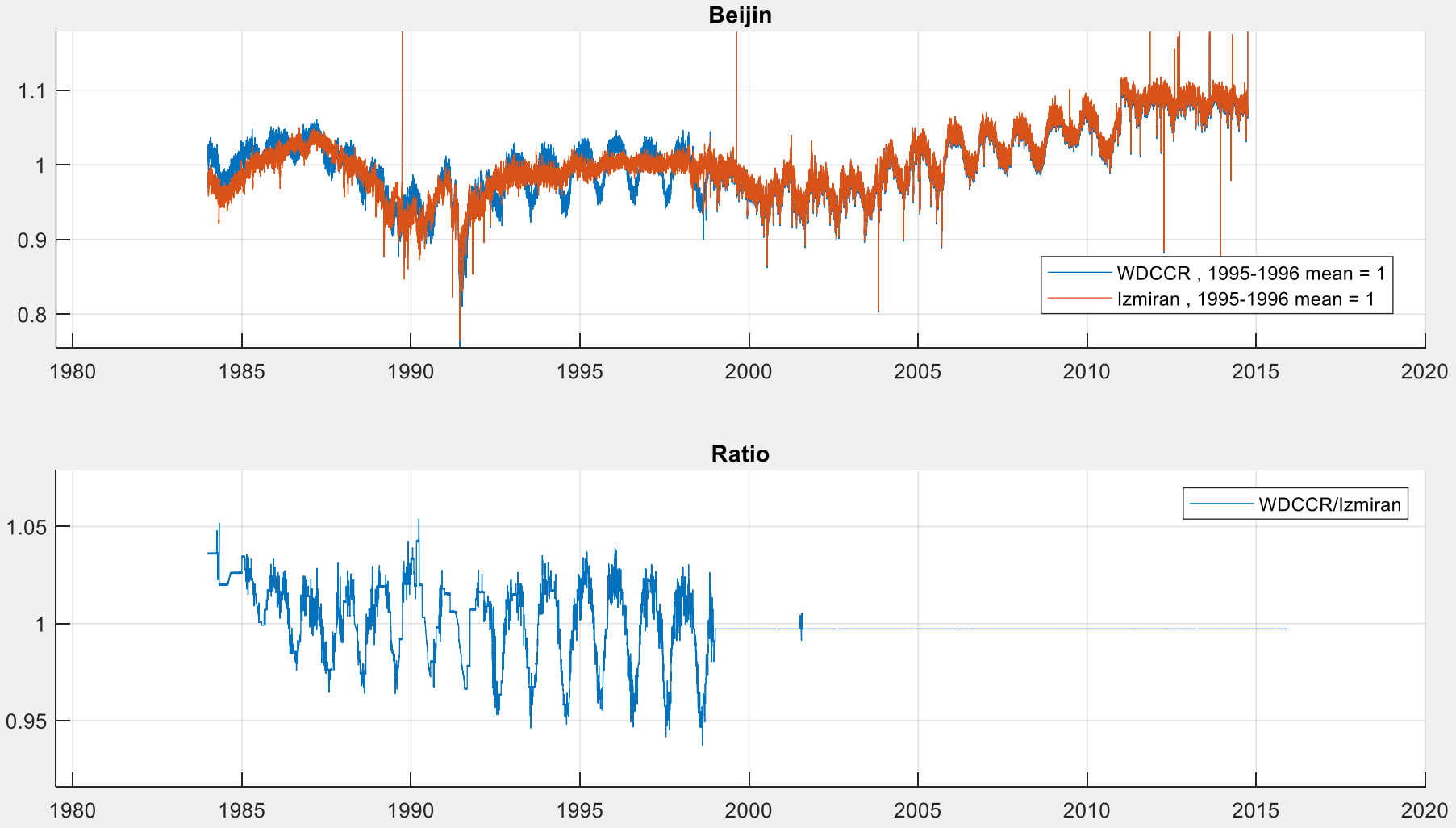


Come see the poster

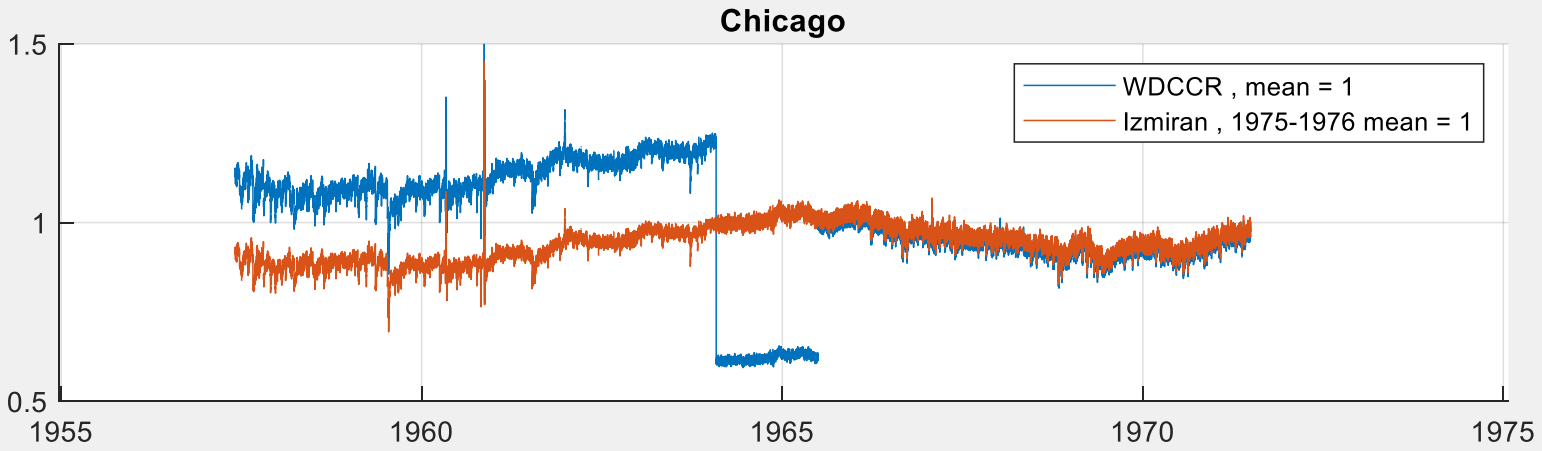
"Compiled "multi-NM" recommended dataset of global NM network" #32

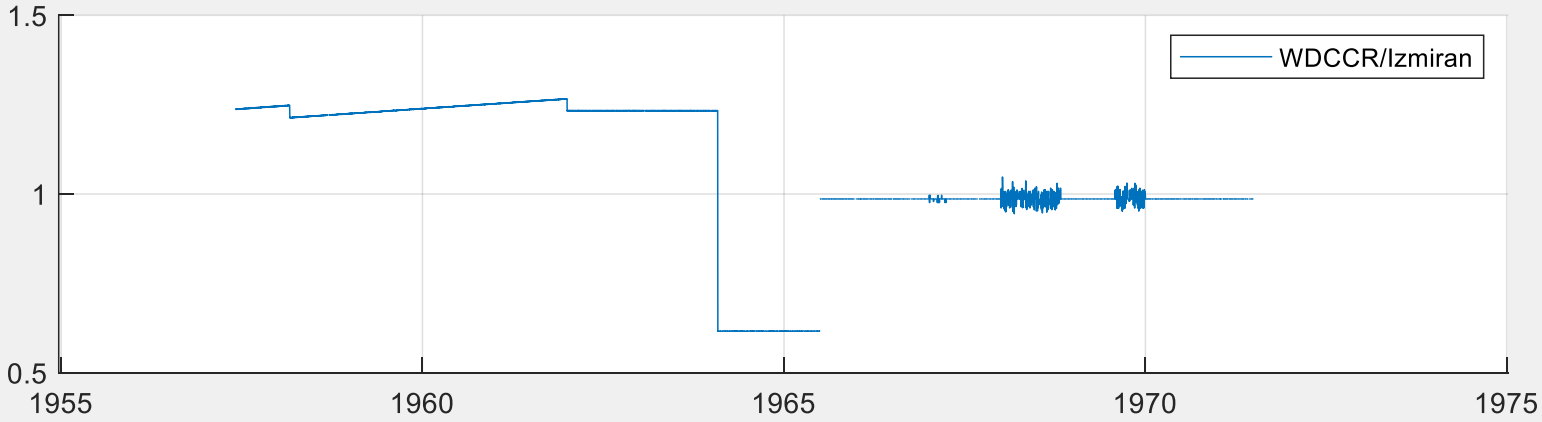
to see what the full dataset looks like

Examples – Izmiran corrections

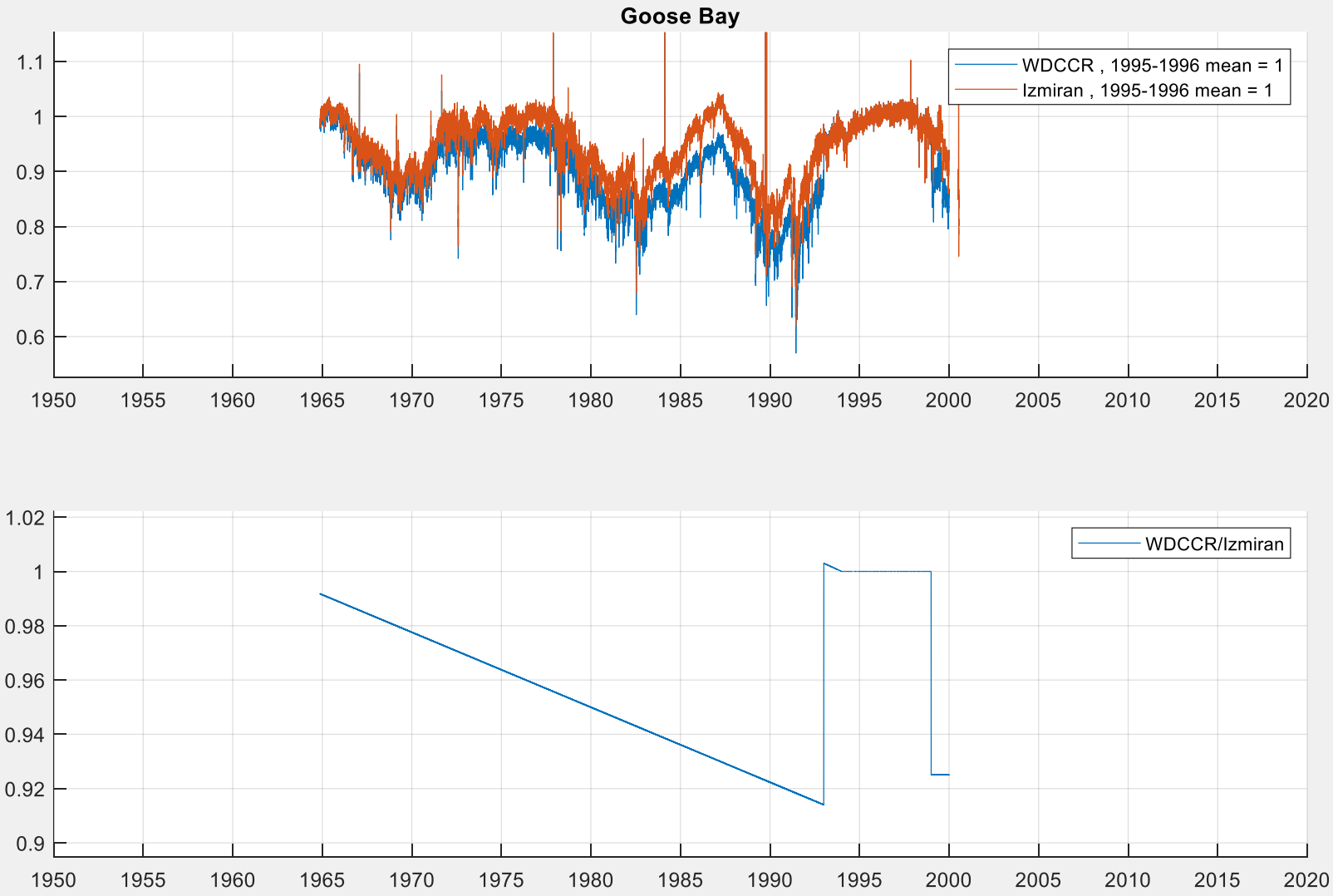


Examples – Izmiran corrections

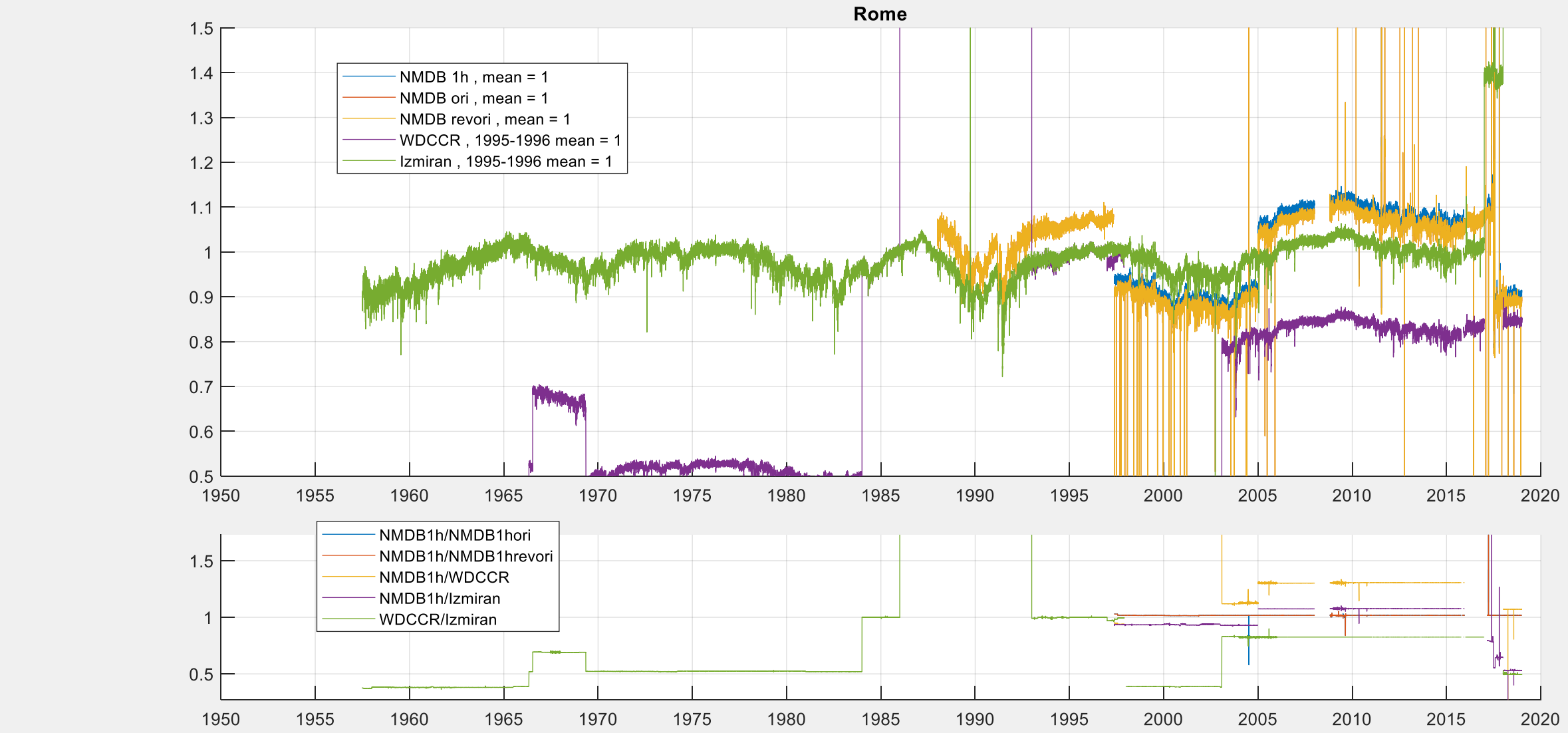




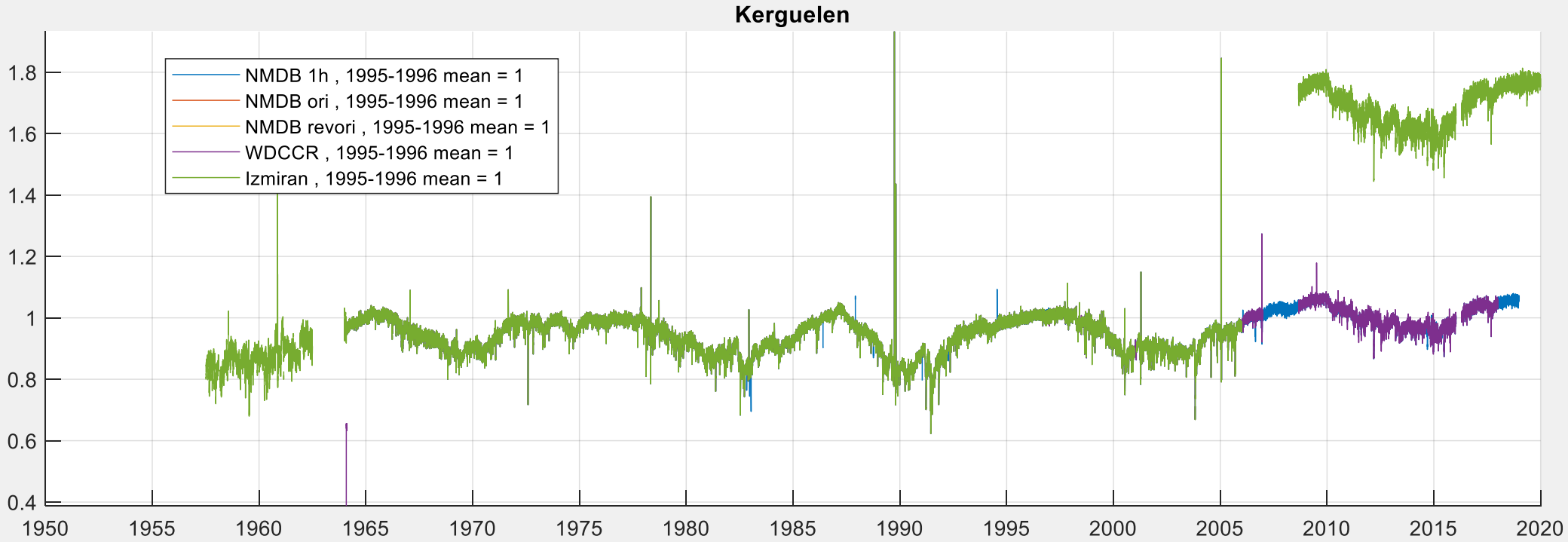
Examples – Izmiran corrections

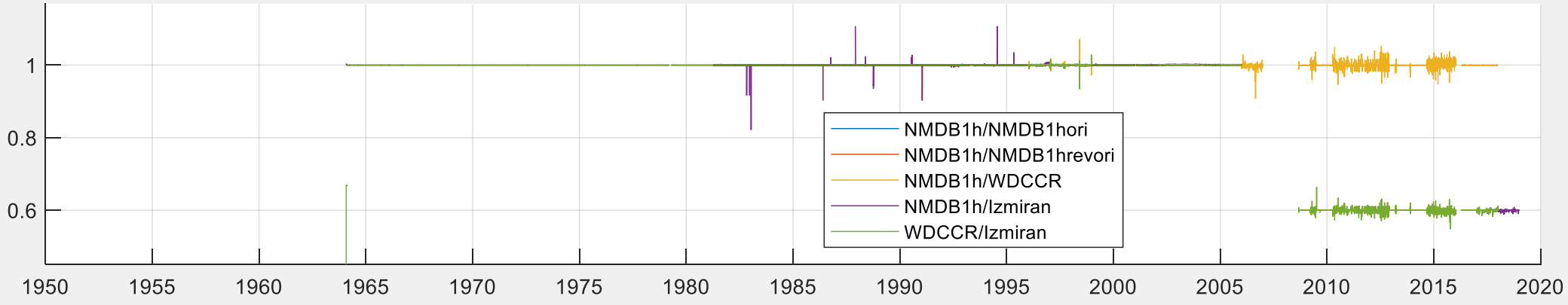


Examples – Izmiran corrections

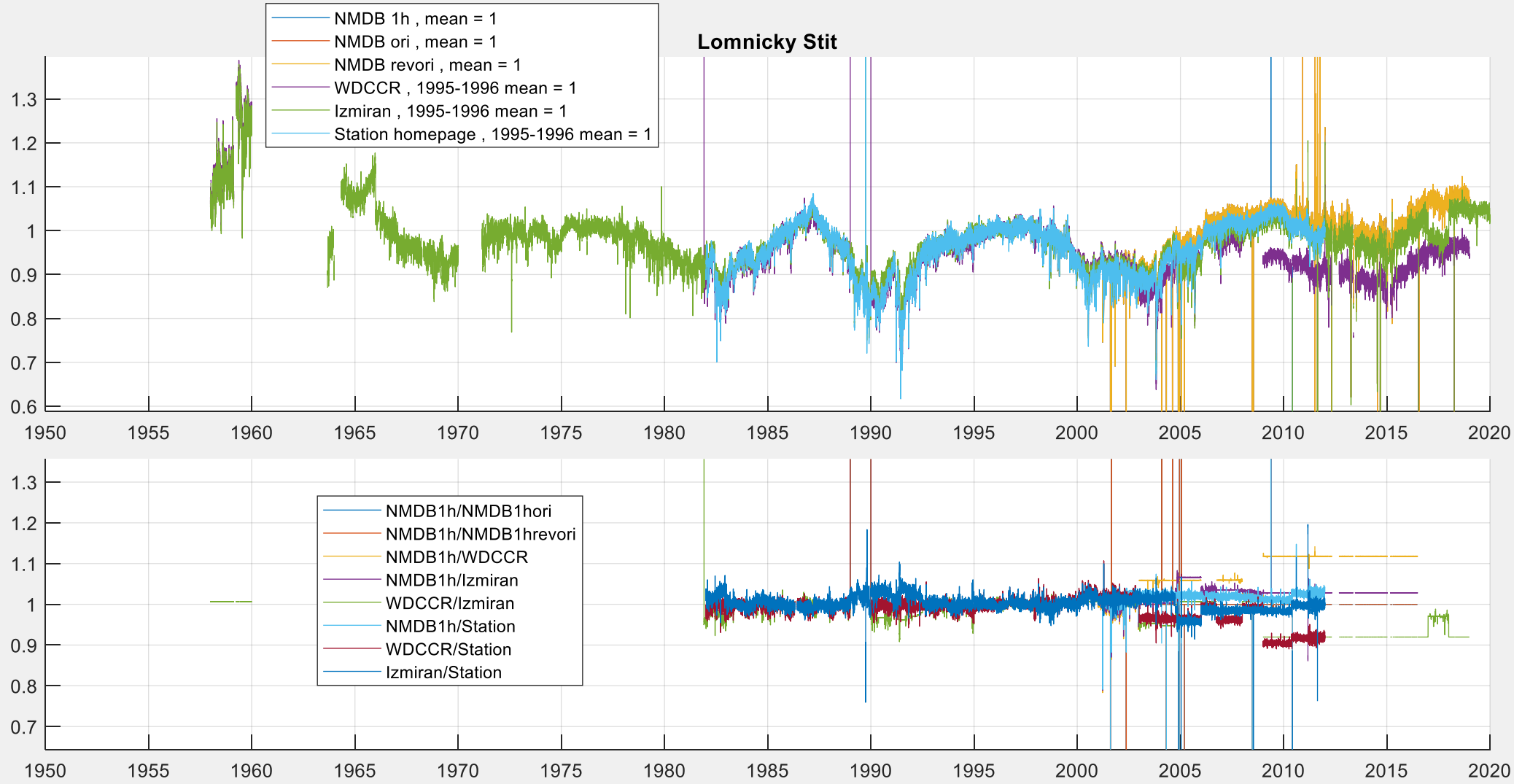


Examples – Multiple sources

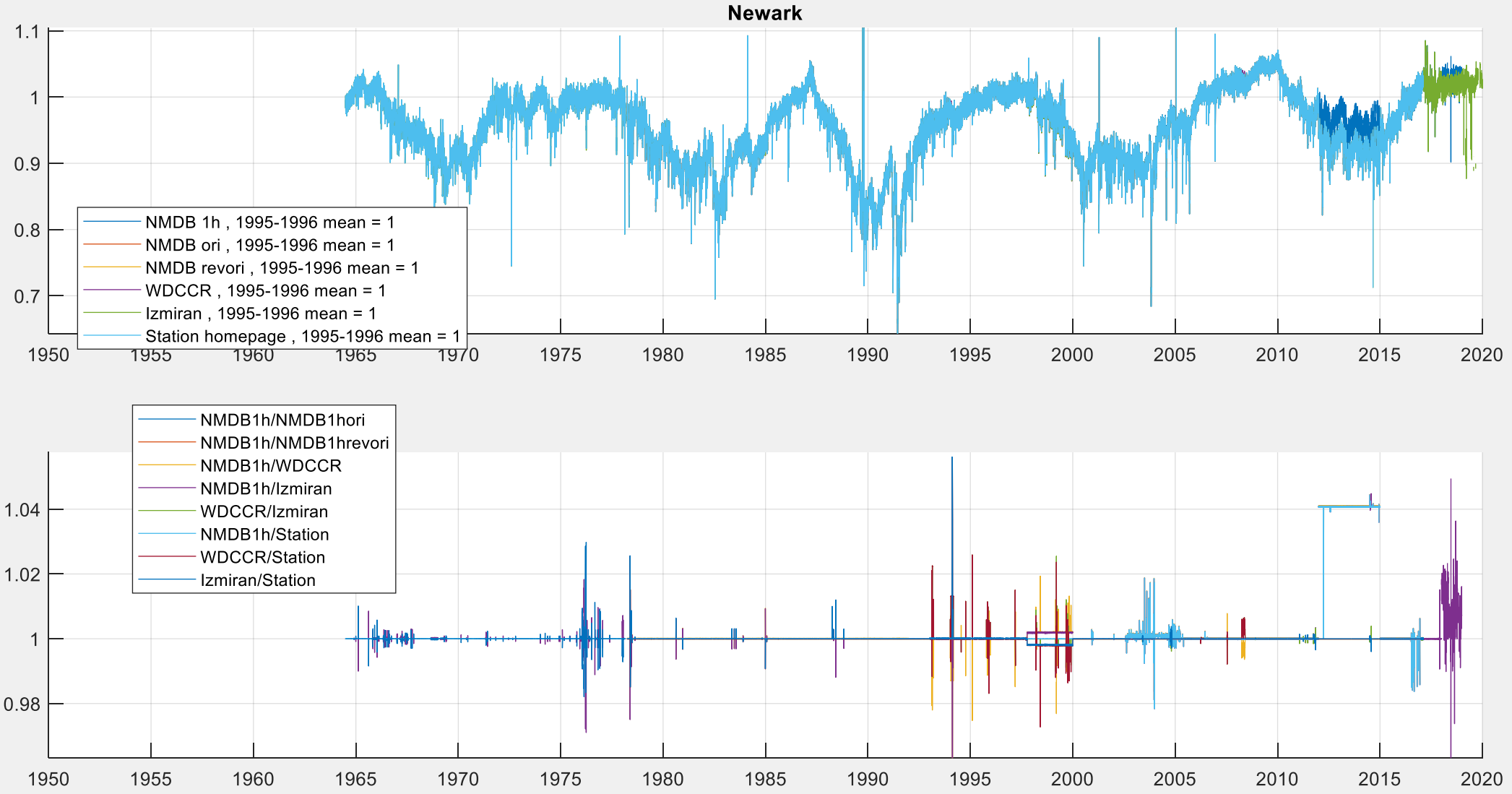




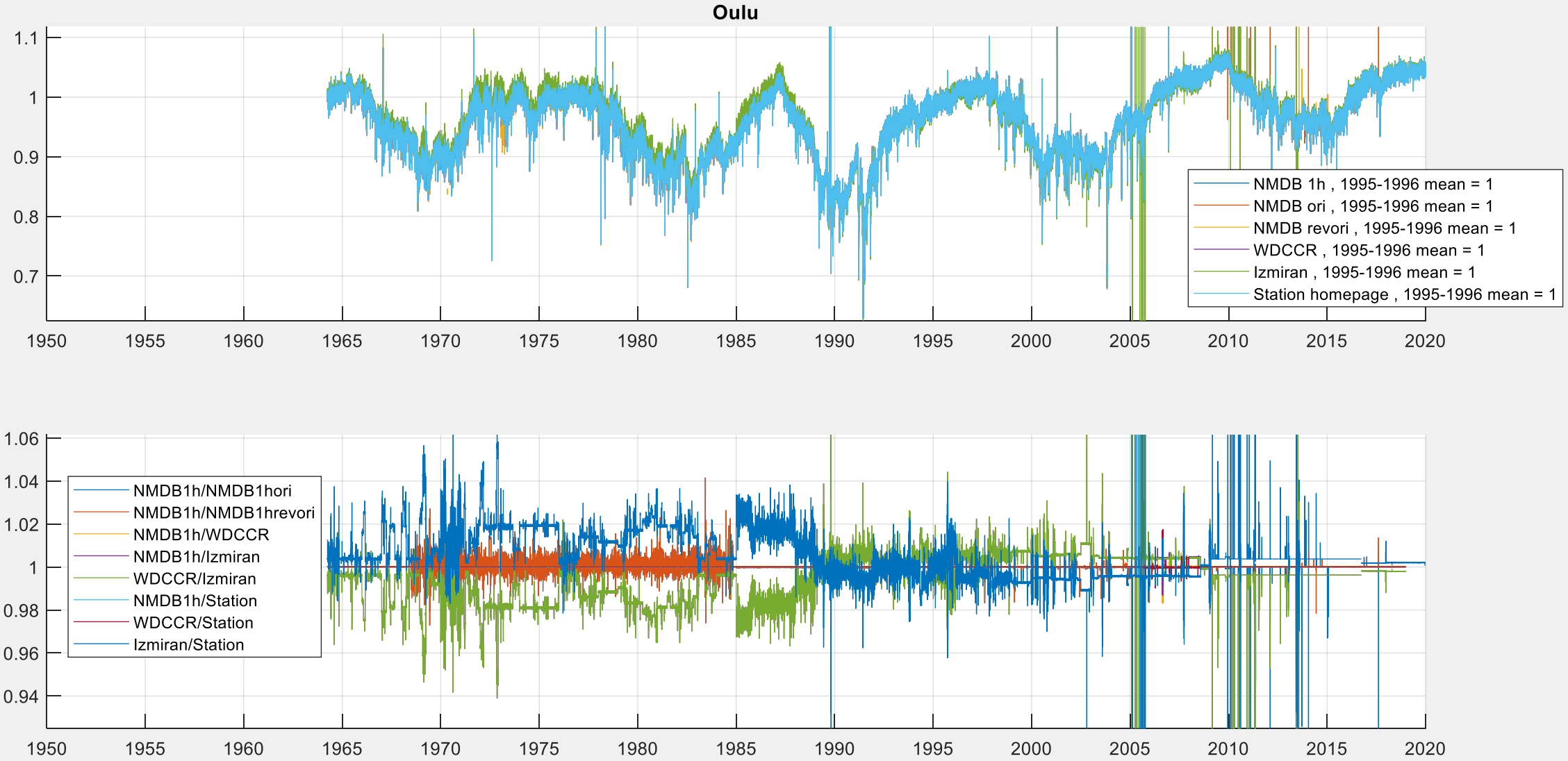
Examples – Multiple sources



Examples – Small differences



Examples – Small differences



Examples – Small differences