UNIVERSITÄT
BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

Copernicus Workshop

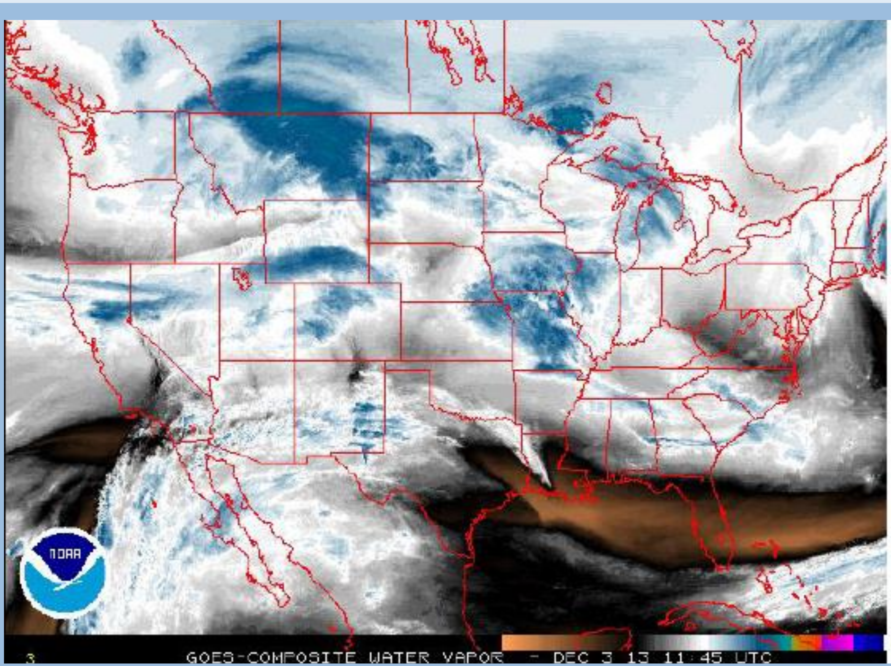
Data rescue and homogenization requirements

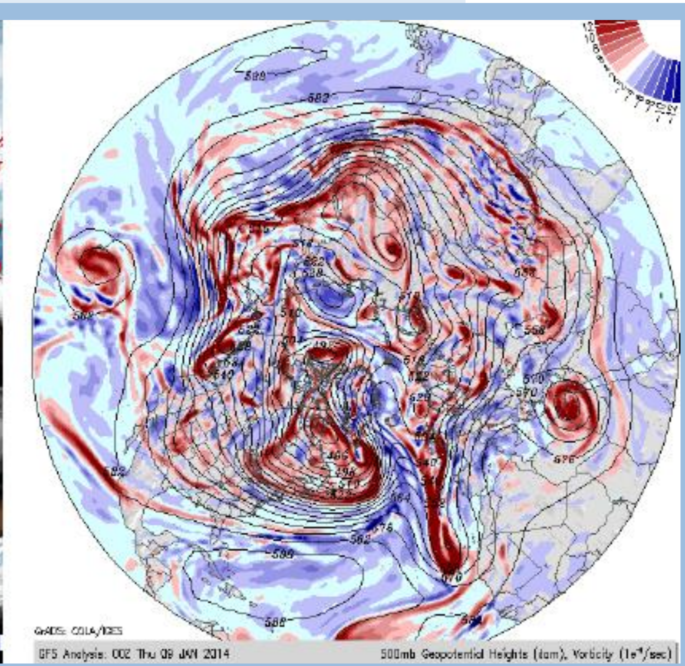
Stefan Brönnimann
Oeschger Centre for Climate Change Research and
Institute of Geography, University of Bern

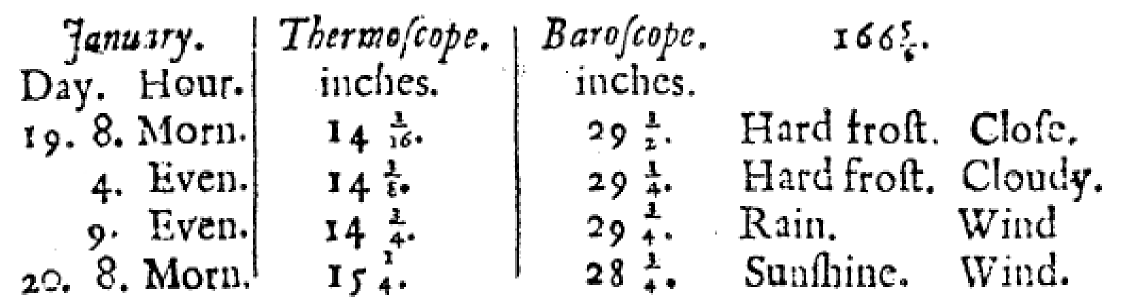
Outline

- Introduction
- Data rescue
- Homogenization
- Conclusions and outlook

Introduction





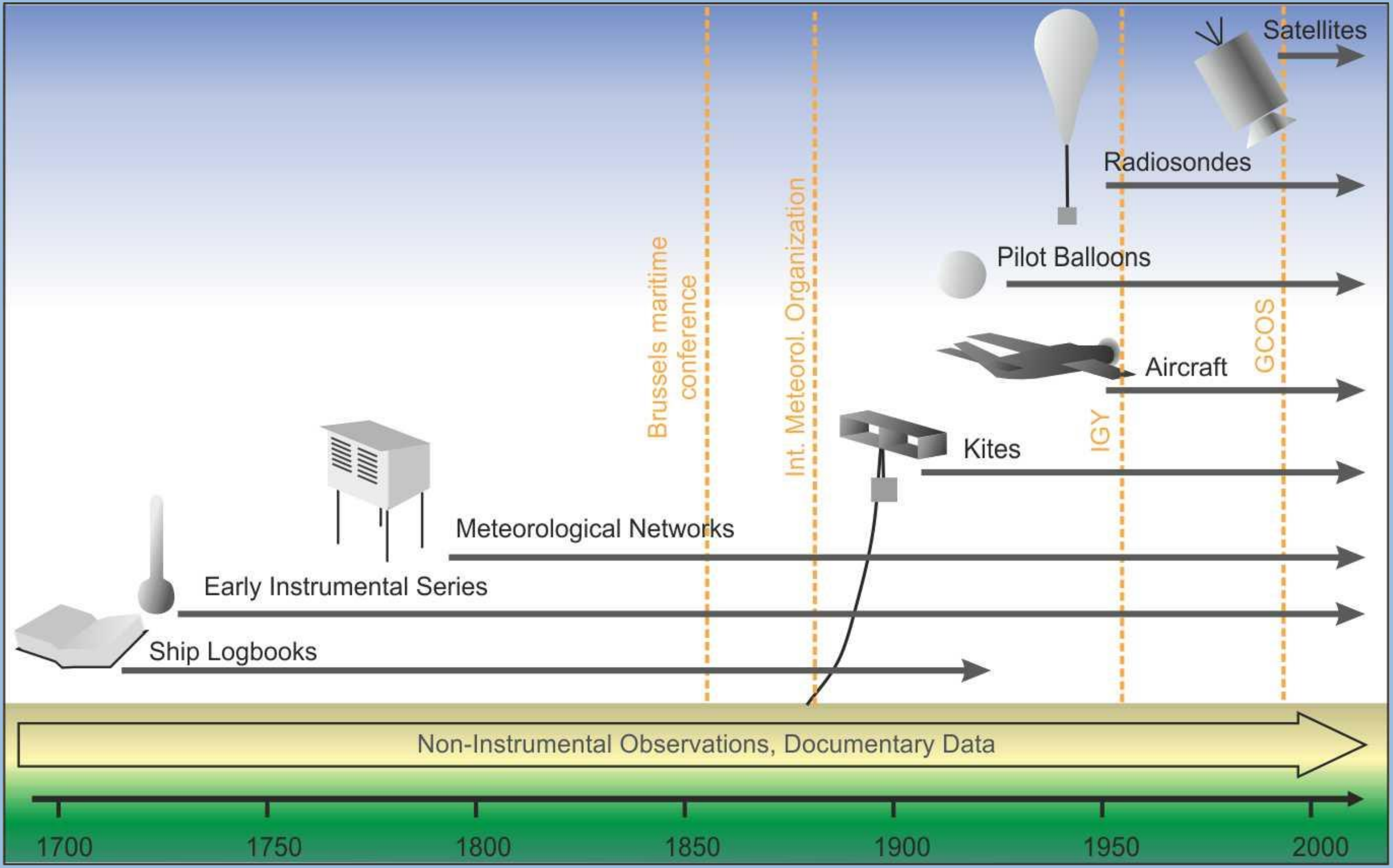


January.		*Thermoſcope.*	*Baroſcope.*	166$\frac{5}{6}$.	
Day.	Hour.	inches.	inches.		
19.	8. Morn.	14 $\frac{1}{16}$.	29 $\frac{1}{2}$.	Hard froſt.	Cloſe.
	4. Even.	14 $\frac{1}{8}$.	29 $\frac{1}{4}$.	Hard froſt.	Cloudy.
	9. Even.	14 $\frac{1}{4}$.	29 $\frac{1}{4}$.	Rain.	Wind
20.	8. Morn.	15 $\frac{1}{4}$.	28 $\frac{1}{4}$.	Sunſhine.	Wind.

Observing systems

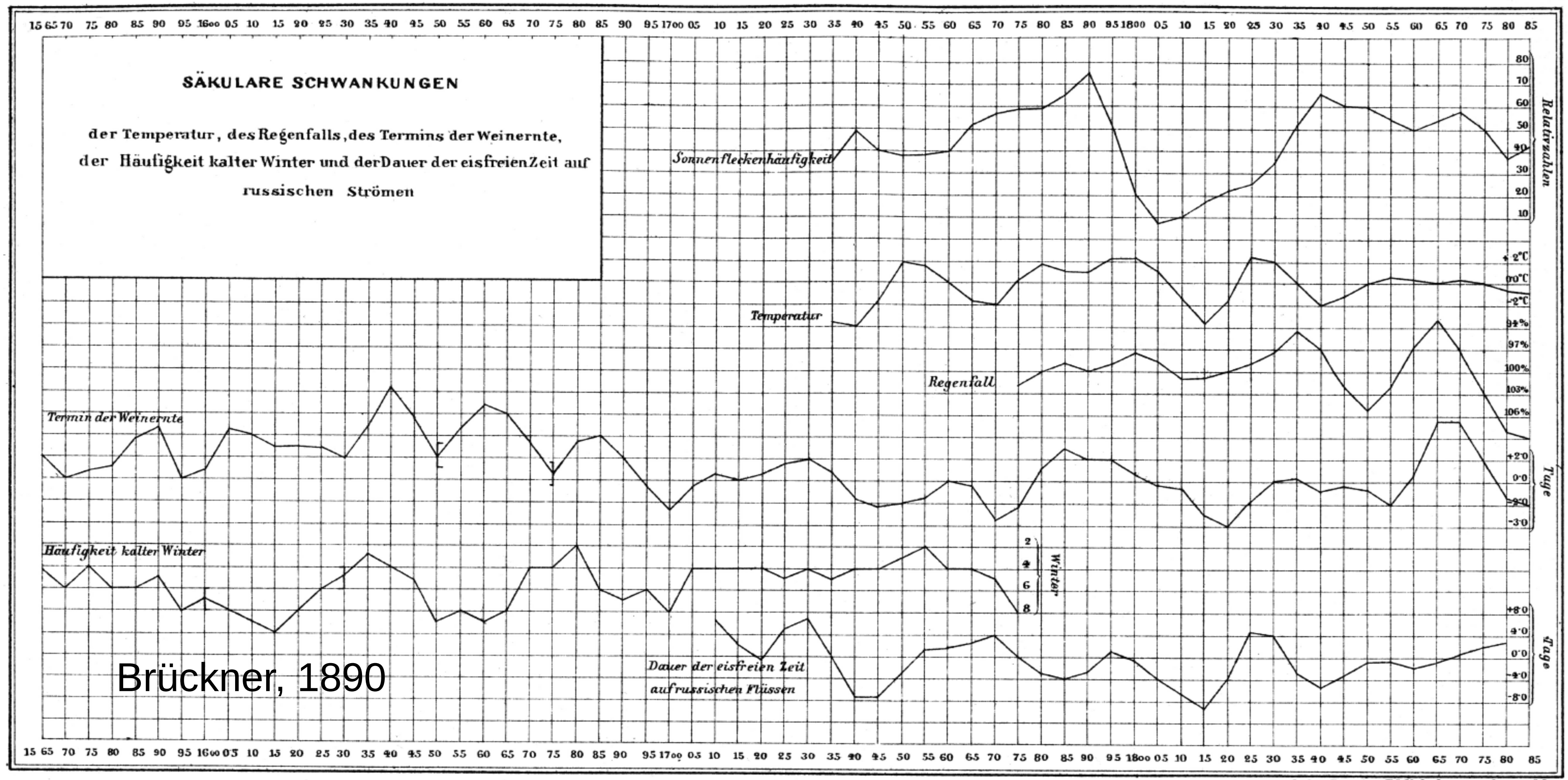
UNIVERSITÄT
BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH



Historical importance of climate data

- Some meteorological data have been compiled in the 19th century
- Some have made the transition to electronic format. Some not



Brückner, 1890

The problem

- A huge (unknown) fraction is not digitised

Why data rescue is an ongoing effort

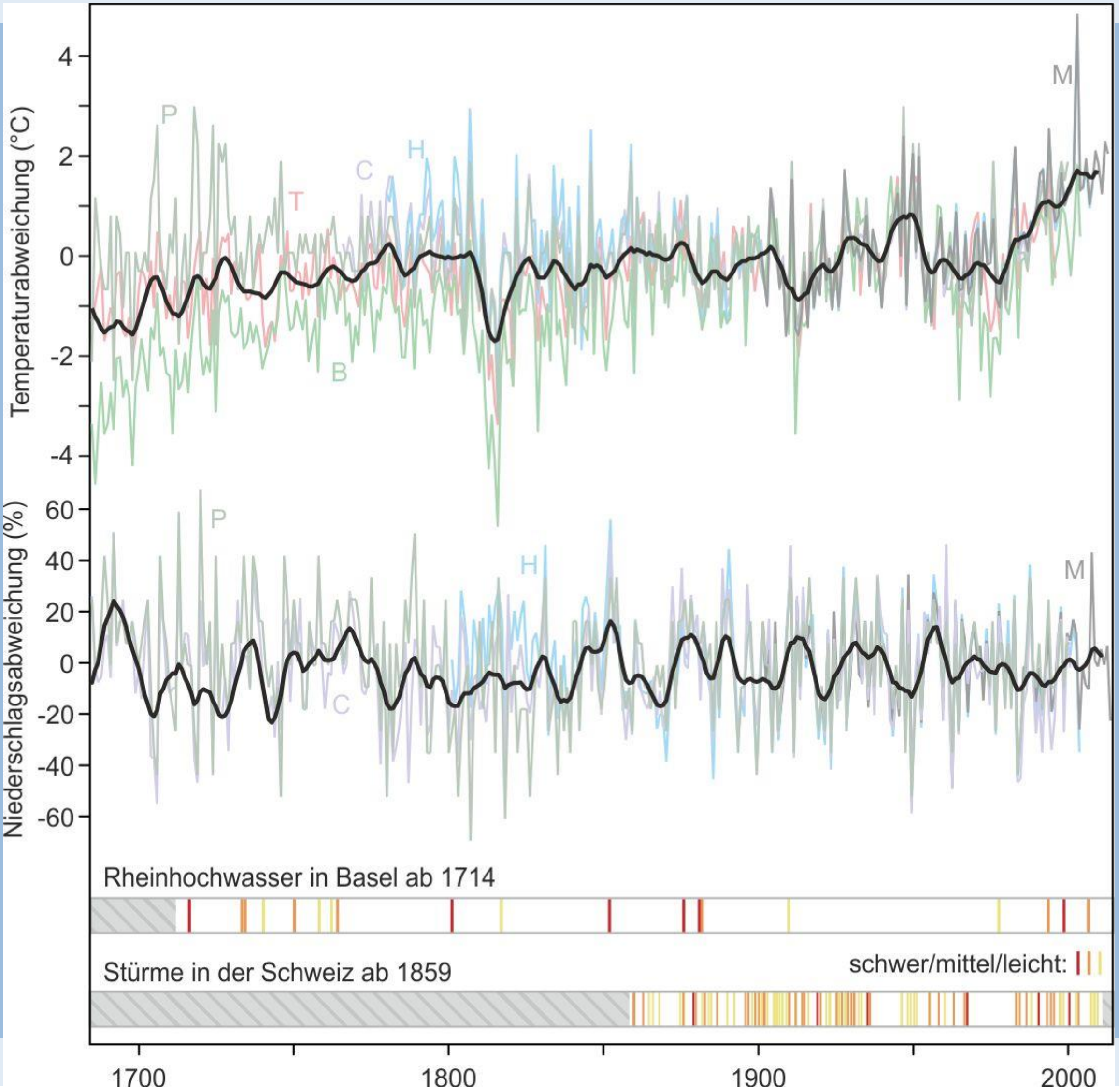
- The needs have changed: Climate science is no longer just about monthly means, but about daily weather and extrema. Climate also no longer is just about temperature, but about the water cycle etc.
  Metadata were often not digitised
- New methods create new opportunities: Data assimilation allows exploiting historical observations in a completely new manner. Pressure data have become important etc.
- An obligation: Meteorological observations are also part of the cultural heritage. Paper data are often in bad shape and need to be preserved.

Therefore: Data rescue is a perpetual task.

Role in climate monitoring and climate services

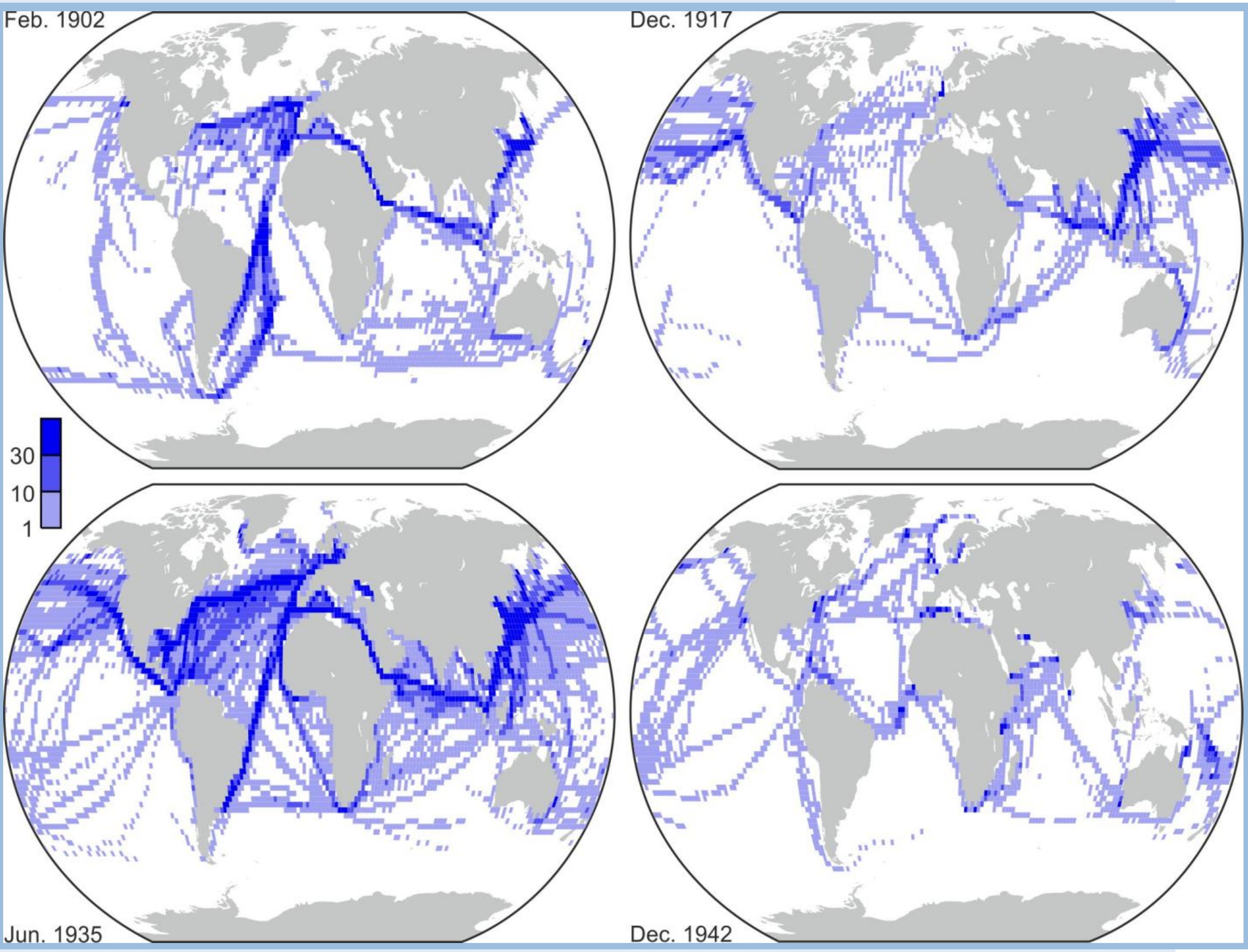
- Long time series (supplement)
- Input for data products (reanalyses and others) fill coverage gaps
- Validation

Long time series
(Summer temperature for Switzerland)

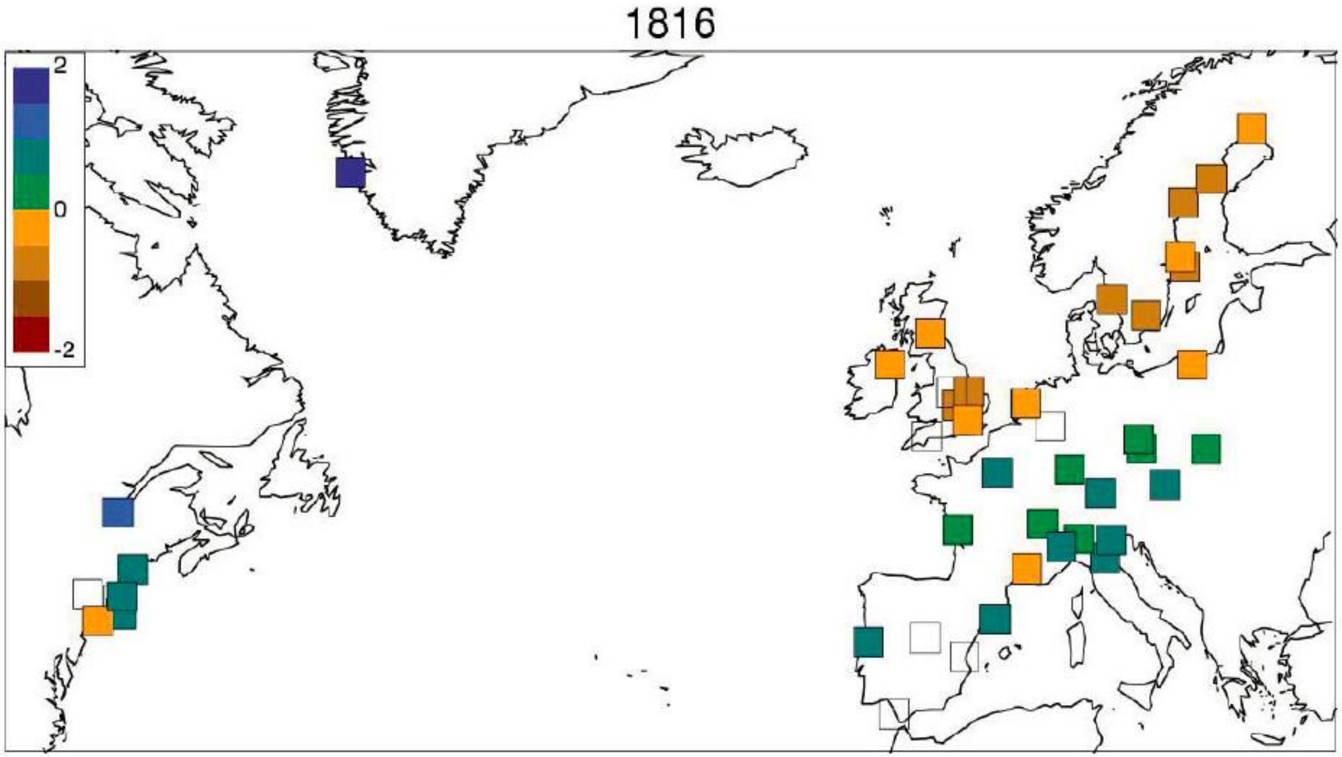


Fill gaps

(#obs in ICOADS SSTs)

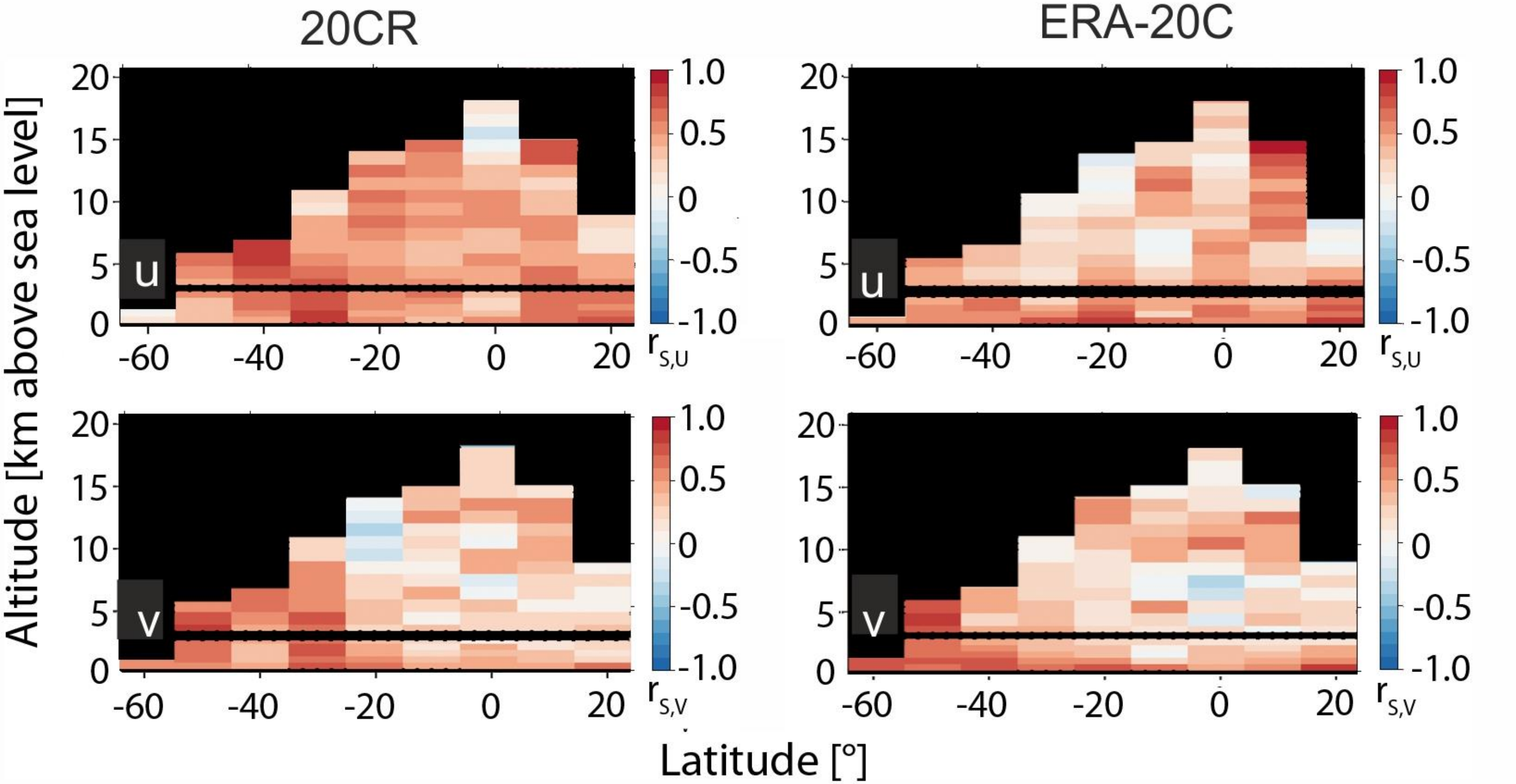


Input for data products
(Pressure data for 20CR-1816)



Brugnara et al. 2015

Independent validation
(Upper-air wind from ship „Meteor“, 1920s)



Stickler et al. (2015)

Who is currently doing the job?

Digitization

- Weather services (e.g. DWD)
- Research Projects: ERA-CLIM2, UERRA, etc.
- Foundations: IEDRO
- Users: Old Weather
- Development programmes (PPCR)

Coordination

- Atmospheric Circulation Reconstructions over the Earth (ACRE)
- WMO DARE (I-DARE talk: Peter Siegmund)
- International Surface Temperature Initiative (ISTI)

Input data, tools and activities needed

- Sustained expertise (community)
- Data repository for data rescue efforts
- Coordination activities
- Link to research projects
- Repository for paper data

Role of Copernicus

Maintaining data rescue infrastructure (incl. expertise, coordination, repository etc.) should become part of operational services

Copernicus should support data rescue that contributes to climate services

What should be digitised?

b
UNIVERSITÄT
BERN

OESCHGER CENTRE
CLIMATE CHANGE RESEARCH

- Data from the 18th and 19th century (examples):

  - Pressure data for reanalyses

  - The Mannheim collection

  - Europe and North America

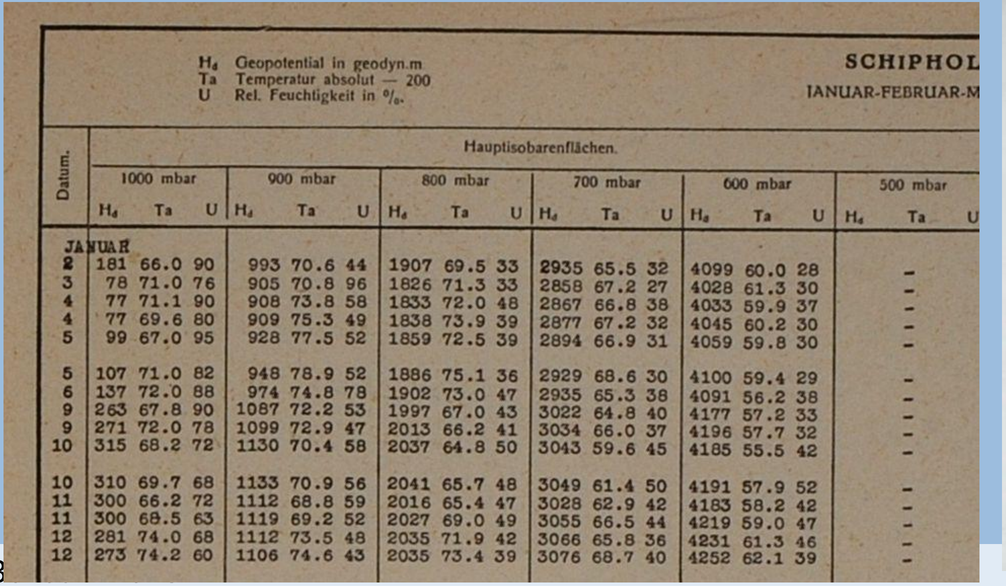
What should be digitised?

- Data from the 20th century:

  Precipitation data

  Irradiances/spectra

  Upper-air



H_d Geopotential in geodyn.m
Ta Temperatur absolut — 200
U Rel. Feuchtigkeit in %.

SCHIPHOL

IANUAR-FEBRUAR-M

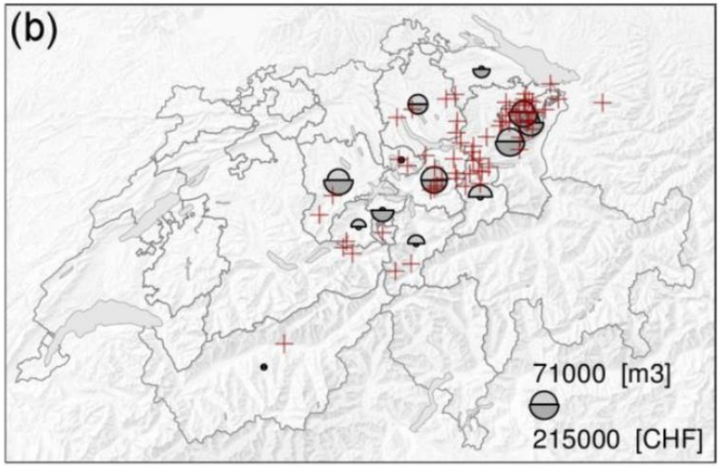
Datum.	Hauptisobarenflächen.																	
	1000 mbar			900 mbar			800 mbar			700 mbar			600 mbar			500 mbar		
	H_d	Ta	U	H_d	Ta	U	H_d	Ta	U	H_d	Ta	U	H_d	Ta	U	H_d	Ta	U
JANUAR																		
2	181	66.0	90	993	70.6	44	1907	69.5	33	2935	65.5	32	4099	60.0	28		-	
3	78	71.0	76	905	70.8	96	1826	71.3	33	2858	67.2	27	4028	61.3	30		-	
4	77	71.1	90	908	73.8	58	1833	72.0	48	2867	66.8	38	4033	59.9	37		-	
4	77	69.6	80	909	75.3	49	1838	73.9	39	2877	67.2	32	4045	60.2	30		-	
5	99	67.0	95	928	77.5	52	1859	72.5	39	2894	66.9	31	4059	59.8	30		-	
5	107	71.0	82	948	78.9	52	1886	75.1	36	2929	68.6	30	4100	59.4	29		-	
6	137	72.0	88	974	74.8	78	1902	73.0	47	2935	65.3	38	4091	56.2	38		-	
9	263	67.8	90	1087	72.2	53	1997	67.0	43	3022	64.8	40	4177	57.2	33		-	
9	271	72.0	78	1099	72.9	47	2013	66.2	41	3034	66.0	37	4196	57.7	32		-	
10	315	68.2	72	1130	70.4	58	2037	64.8	50	3043	59.6	45	4185	55.5	42		-	
10	310	69.7	68	1133	70.9	56	2041	65.7	48	3049	61.4	50	4191	57.9	52		-	
11	300	66.2	72	1112	68.8	59	2016	65.4	47	3028	62.9	42	4183	58.2	42		-	
11	300	68.5	63	1119	69.2	52	2027	69.0	49	3055	66.5	44	4219	59.0	47		-	
12	281	74.0	68	1112	73.5	48	2035	71.9	42	3066	65.8	36	4231	61.3	46		-	
12	273	74.2	60	1106	74.6	43	2035	73.4	39	3076	68.7	40	4252	62.1	39		-	

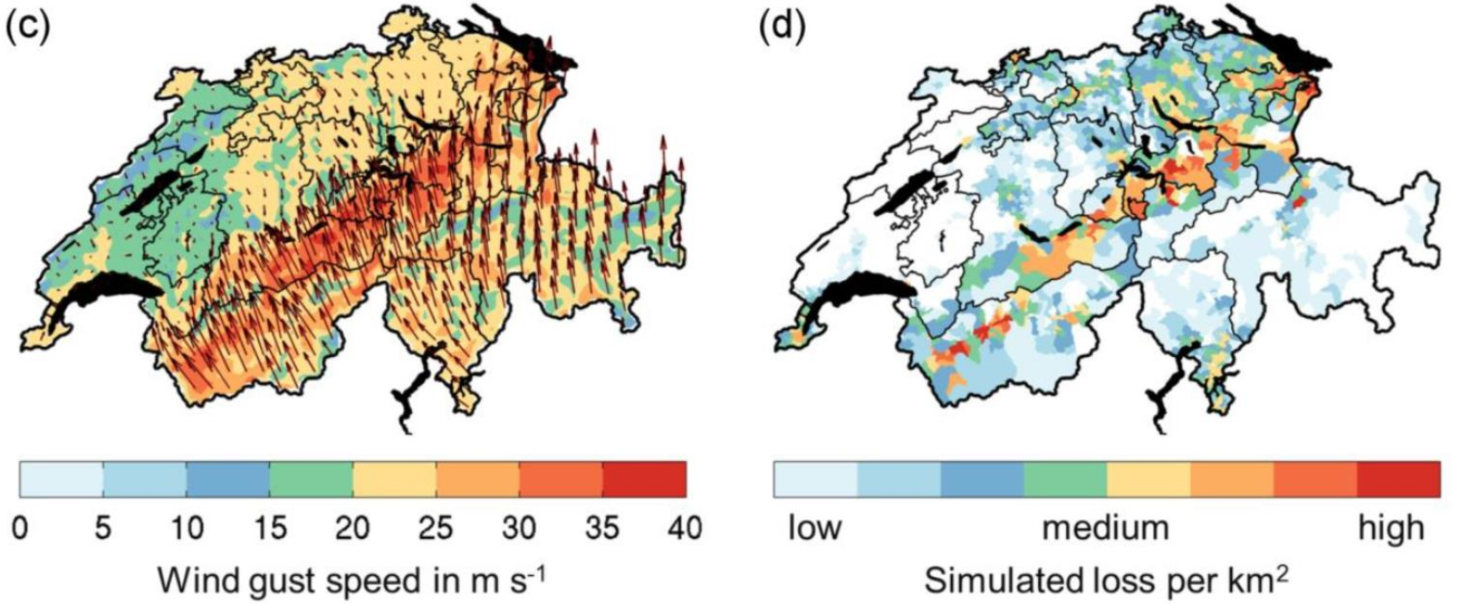
What should be digitised?

- Beyond weather data:

  - Documentary information

  - Damage data



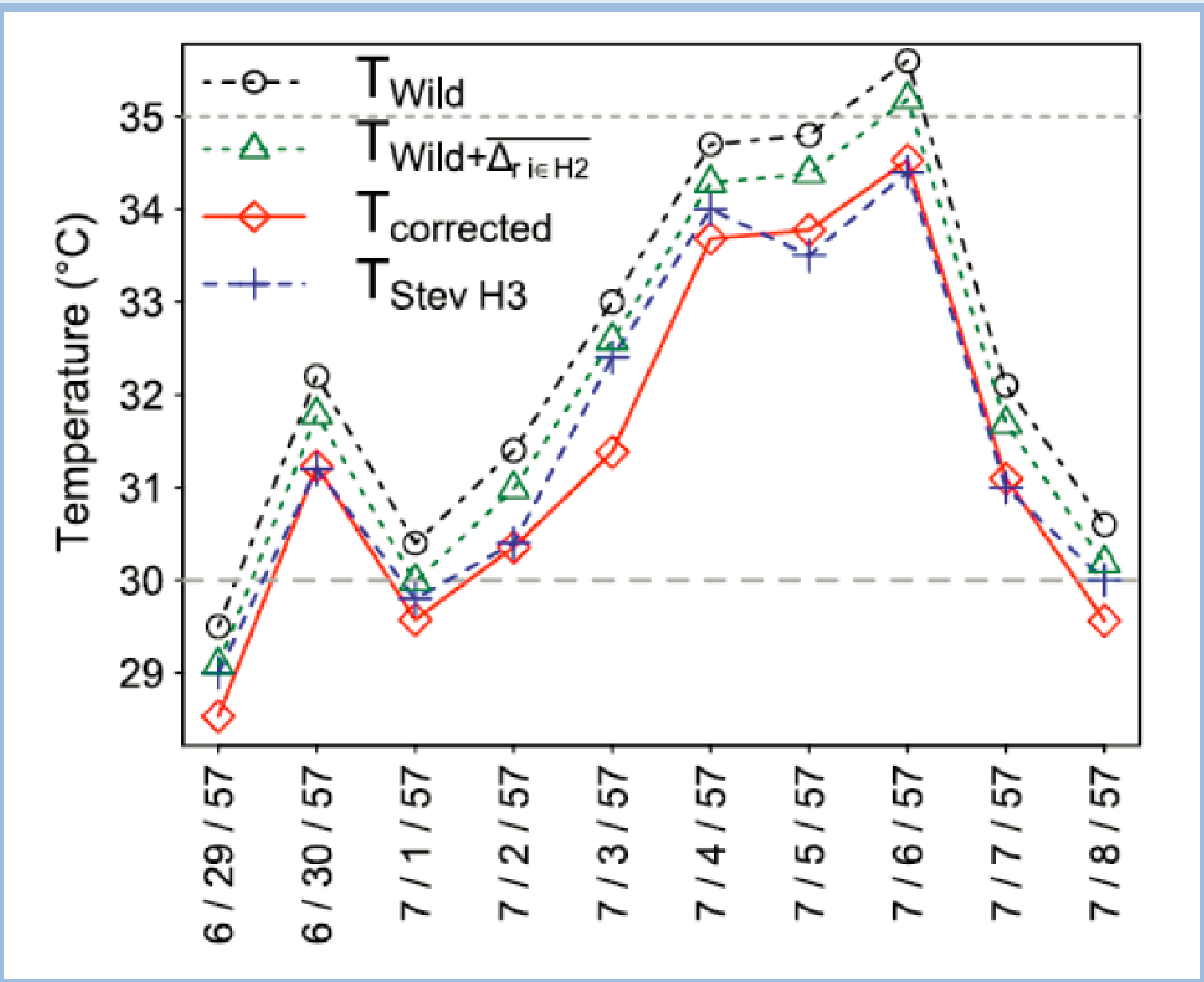


Stucki et al. (2015)

Homogenization

- Operational (weather services)
- Research needs: Daily homogenization, precipitation, physics-based homogenization
- Parallel measurements
- Cost Action HOME
- ISTI Group, EUSTACE

Relevance for extremes



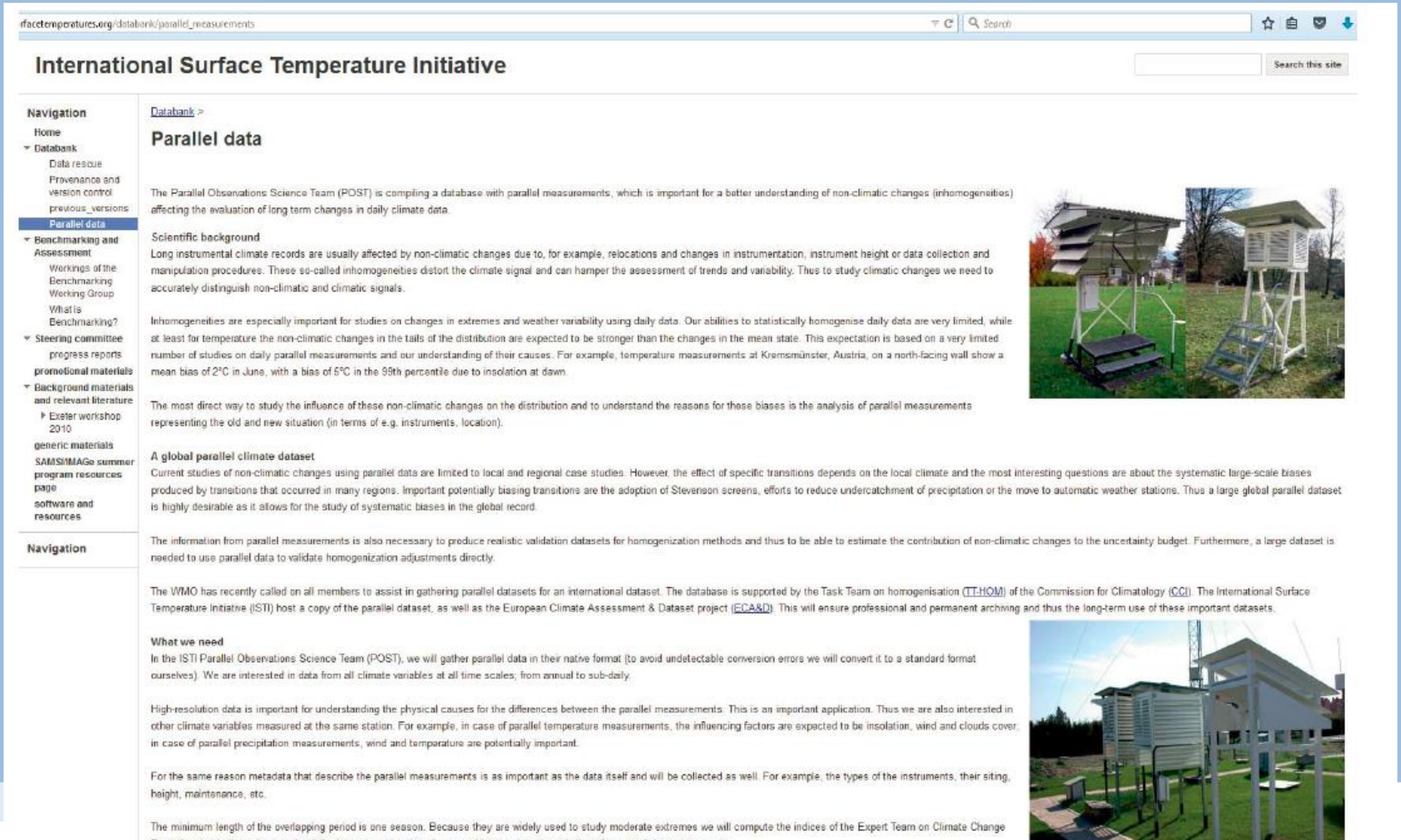
Auchmann and Brönimann (2012)

Change of screen

Auchmann and Brönimann (2012)

Parallel measurements

- Initiative within ISTI



International Surface Temperature Initiative

Parallel data

The Parallel Observations Science Team (POST) is compiling a database with parallel measurements, which is important for a better understanding of non-climatic changes (inhomogeneities) affecting the evaluation of long term changes in daily climate data.

Scientific background

Long instrumental climate records are usually affected by non-climatic changes due to, for example, relocations and changes in instrumentation, instrument height or data collection and manipulation procedures. These so-called inhomogeneities distort the climate signal and can hamper the assessment of trends and variability. Thus to study climatic changes we need to accurately distinguish non-climatic and climatic signals.

Inhomogeneities are especially important for studies on changes in extremes and weather variability using daily data. Our abilities to statistically homogenise daily data are very limited, while at least for temperature the non-climatic changes in the tails of the distribution are expected to be stronger than the changes in the mean state. This expectation is based on a very limited number of studies on daily parallel measurements and our understanding of their causes. For example, temperature measurements at Kremsmünster, Austria, on a north-facing wall show a mean bias of 2°C in June, with a bias of 5°C in the 99th percentile due to insolation at dawn.

The most direct way to study the influence of these non-climatic changes on the distribution and to understand the reasons for these biases is the analysis of parallel measurements representing the old and new situation (in terms of e.g. instruments, location).

A global parallel climate dataset

Current studies of non-climatic changes using parallel data are limited to local and regional case studies. However, the effect of specific transitions depends on the local climate and the most interesting questions are about the systematic large-scale biases produced by transitions that occurred in many regions. Important potentially biasing transitions are the adoption of Stevenson screens, efforts to reduce undercatchment of precipitation or the move to automatic weather stations. Thus a large global parallel dataset is highly desirable as it allows for the study of systematic biases in the global record.

The information from parallel measurements is also necessary to produce realistic validation datasets for homogenization methods and thus to be able to estimate the contribution of non-climatic changes to the uncertainty budget. Furthermore, a large dataset is needed to use parallel data to validate homogenization adjustments directly.

The WMO has recently called on all members to assist in gathering parallel datasets for an international dataset. The database is supported by the Task Team on homogenisation (TT-HOM) of the Commission for Climatology (CCl). The International Surface Temperature Initiative (ISTI) host a copy of the parallel dataset, as well as the European Climate Assessment & Dataset project (ECA&D). This will ensure professional and permanent archiving and thus the long-term use of these important datasets.

What we need

In the ISTI Parallel Observations Science Team (POST), we will gather parallel data in their native format (to avoid undetectable conversion errors we will convert it to a standard format ourselves). We are interested in data from all climate variables at all time scales; from annual to sub-daily.

High-resolution data is important for understanding the physical causes for the differences between the parallel measurements. This is an important application. Thus we are also interested in other climate variables measured at the same station. For example, in case of parallel temperature measurements, the influencing factors are expected to be insolation, wind and clouds cover, in case of parallel precipitation measurements, wind and temperature are potentially important.

For the same reason metadata that describe the parallel measurements is as important as the data itself and will be collected as well. For example, the types of the instruments, their siting, height, maintenance, etc.

The minimum length of the overlapping period is one season. Because they are widely used to study moderate extremes we will compute the indices of the Expert Team on Climate Change

Input data, tools and activities needed

- Coordinated Benchmarking and Homogenization Activities
- Metadata Data Coordination
- Parallel Measurements Coordination
- Link to Research Projects

Role of Copernicus

Maintaining homogenisation infrastructure (incl. expertise, coordination, repository etc.) should become part of operational services

Conclusions/outlook

- Data rescue:
  - Historical weather data continue to be an underused resource
  - Needs and opportunities change: Back to the archive!
  - Role of Copernicus: Support the infrastructure (expertise, coordination, repositories)
  - Role of Copernicus: Recover data that contribute to climate services (example: Mannheim compilation)
  - Beyond weather data

- Homogenization
  - Homogenization of daily data and precipitation
  - Parallel data collections for homogenization
  - Copernicus could support the infrastructure (expertise, coordination, repositories)