



u^b

**UNIVERSITÄT
BERN**

**OESCHGER CENTRE
CLIMATE CHANGE RESEARCH**

Data Rescue, Digitisation, and related Data Products in the Climatology Group at Univ. Bern

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

Outline

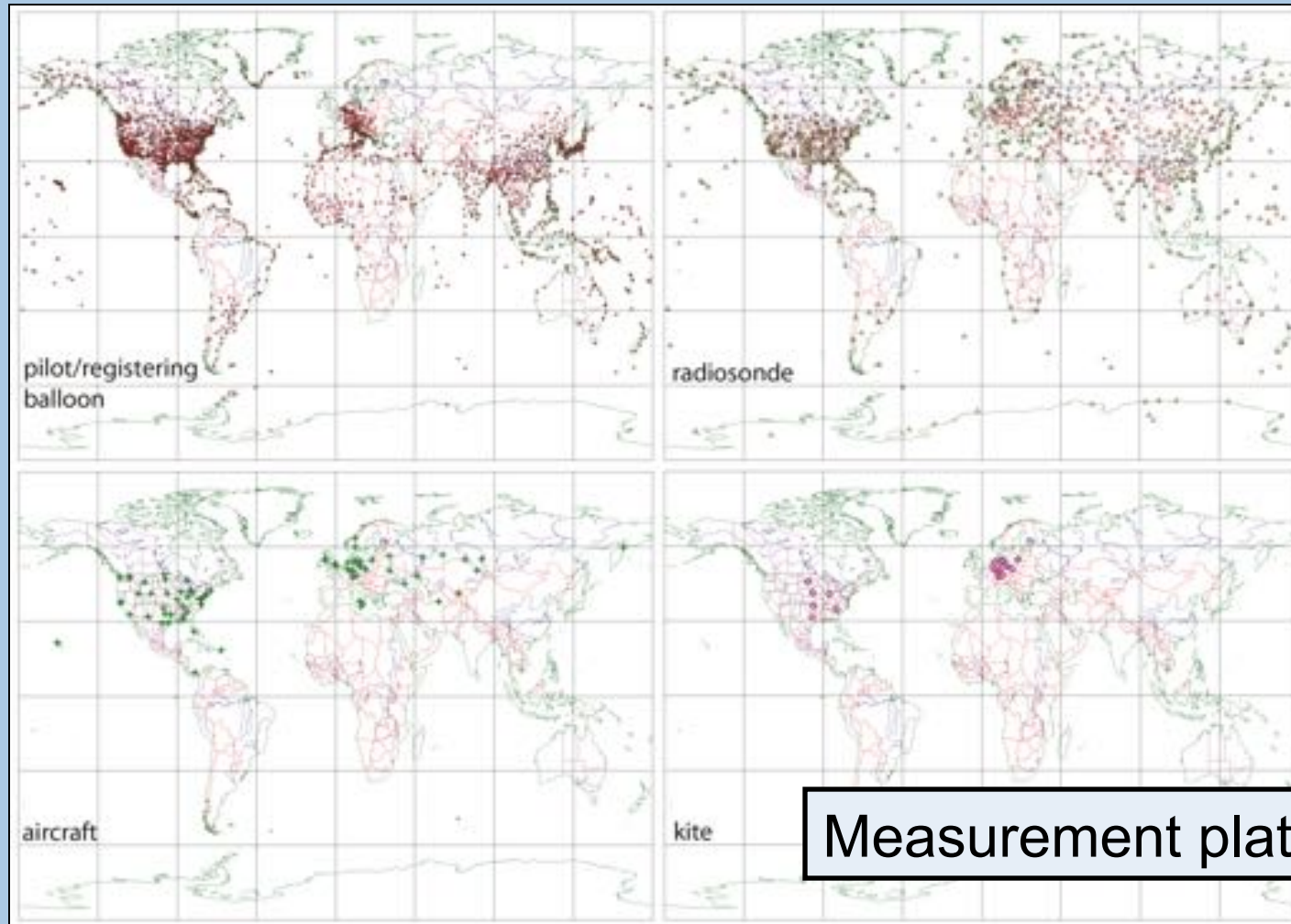
- > Recent data rescue and digitisation efforts and related products in the climatology group at the Oeschger Centre / Univ. Bern (formerly at ETH Zurich)
 - CHUAN (Comprehensive Historical Upper Air Network, Stickler et al. 2010, *BAMS*)
 - Statistical reconstructions (Griesser et al. 2010, *J. Clim.*, Brönnimann et al. 2009, *Met. Z.*, 2010, subm. to *Clim. Dyn.*)
 - data.rescue@home

- > Example of application
 - Wind biases of NCEP/NCAR (NNR) and Twentieth Century (20CR) reanalyses w.r.t. CHUAN observations in the African and Asian monsoon regions (Stickler and Brönnimann 2010, subm. to *QJRMS*)

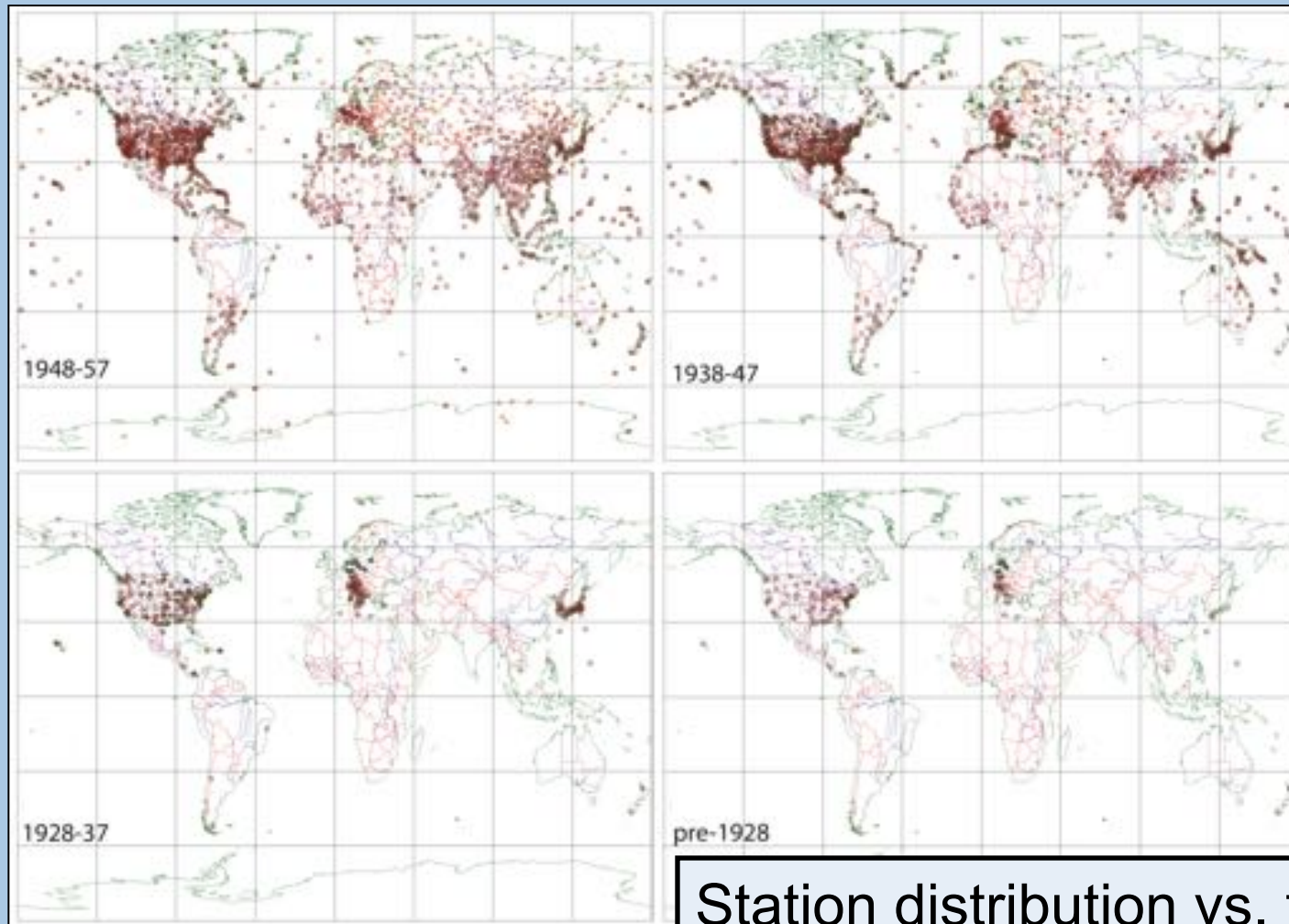
The Comprehensive Historical Upper Air Network (CHUAN)

- > Global upper air dataset
- > Going back to the first decades of the 20th century
- > Ca. 4000 records
- > > 12.5 mio observations before 1957, 5.3 mio before 1948
- > Early data mainly pilot balloon, but also radiosonde, aircraft, kite, registering balloon etc.
- > **Advantages:**
 - Brings available early upper air data into one simple format
 - Raw and corrected / quality controlled versions
 - Monthly means freely available: www.historicalupperair.org
 - Full dataset will be hosted by NCDC
 - Primary observational data, probably closest to “reality”
- > **Disadvantage:**
 - No spatially complete field, coverage strongly reduced in the early part

The Comprehensive Historical Upper Air Network (CHUAN)

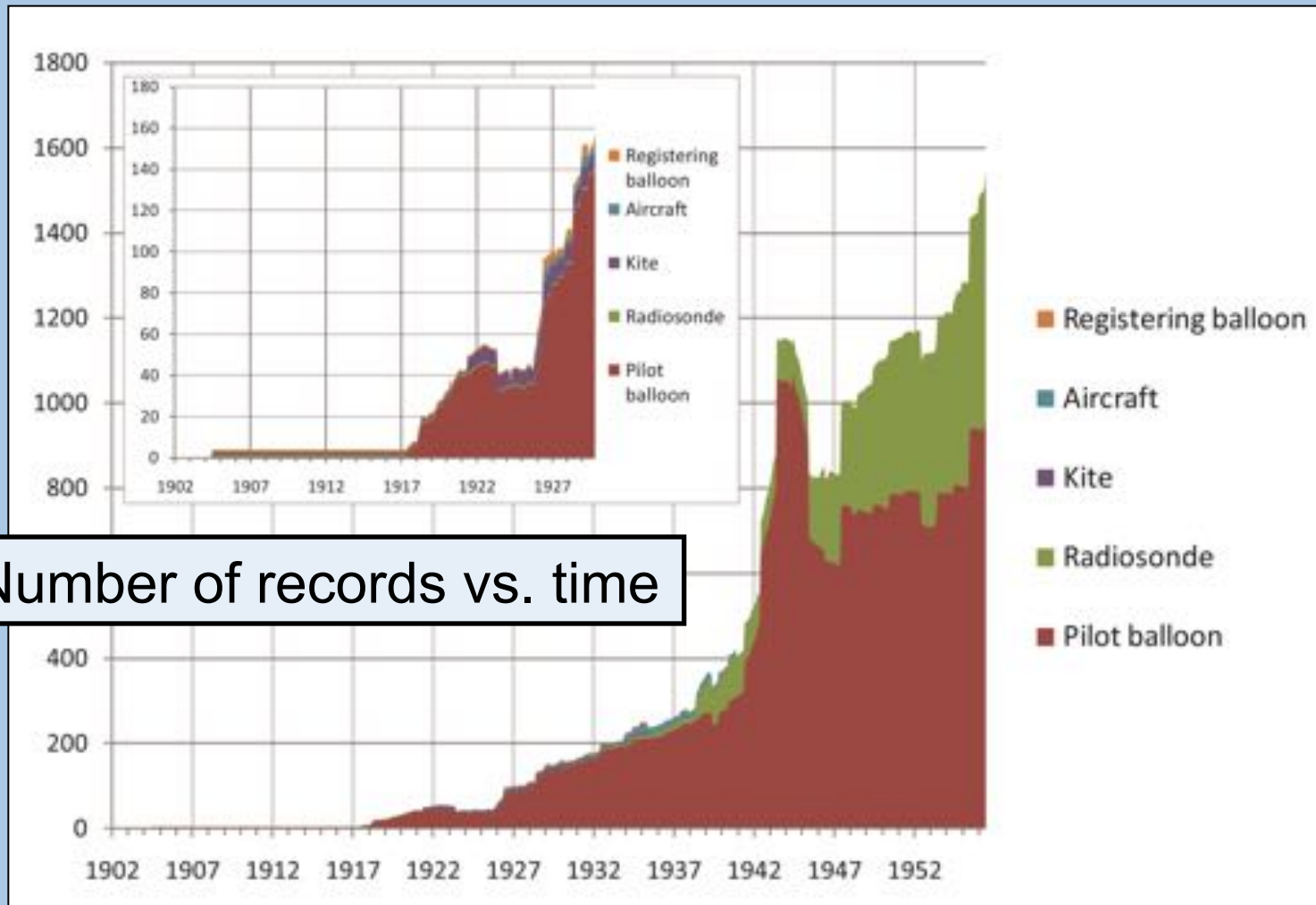


The Comprehensive Historical Upper Air Network (CHUAN)



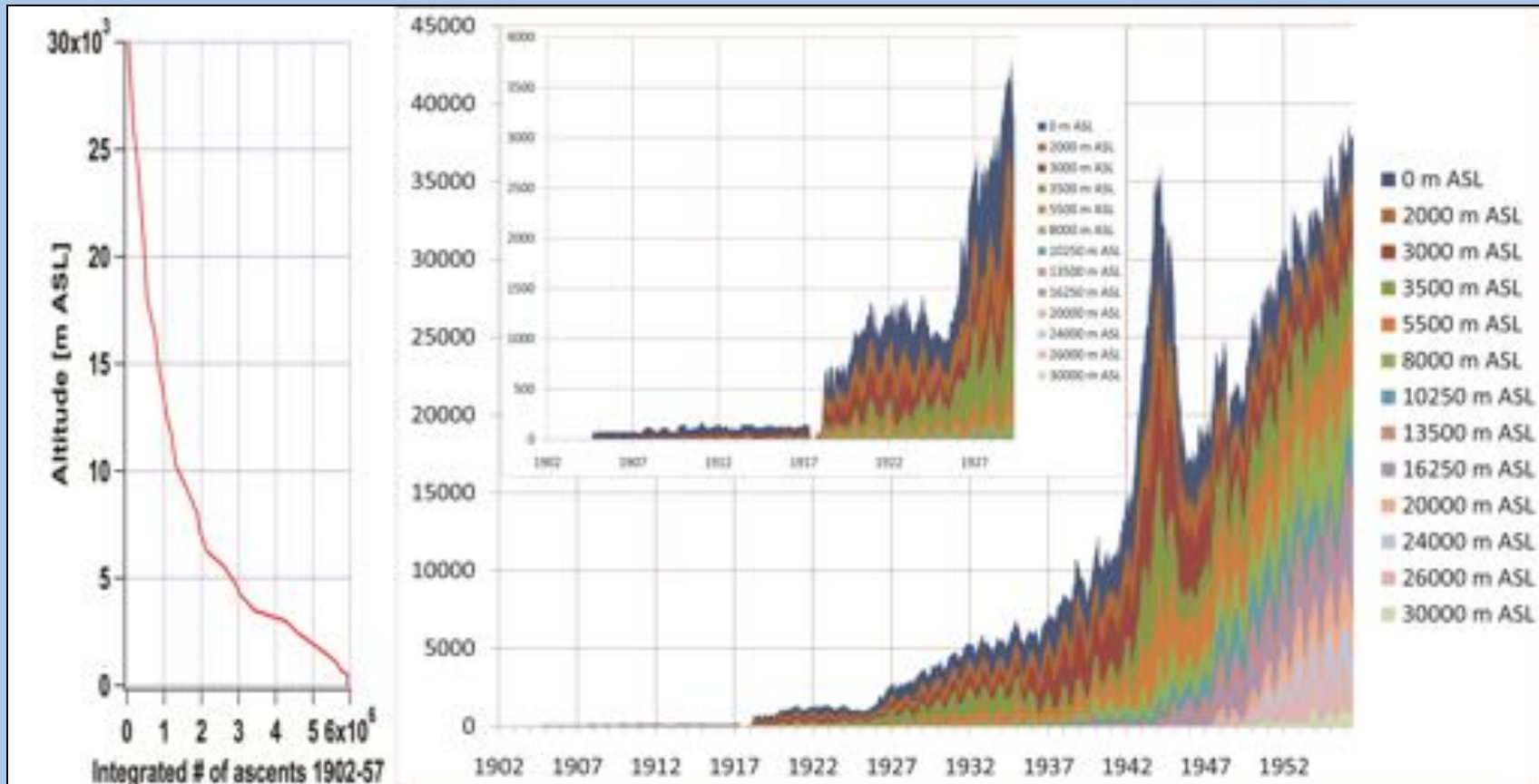
Station distribution vs. time

The Comprehensive Historical Upper Air Network (CHUAN)



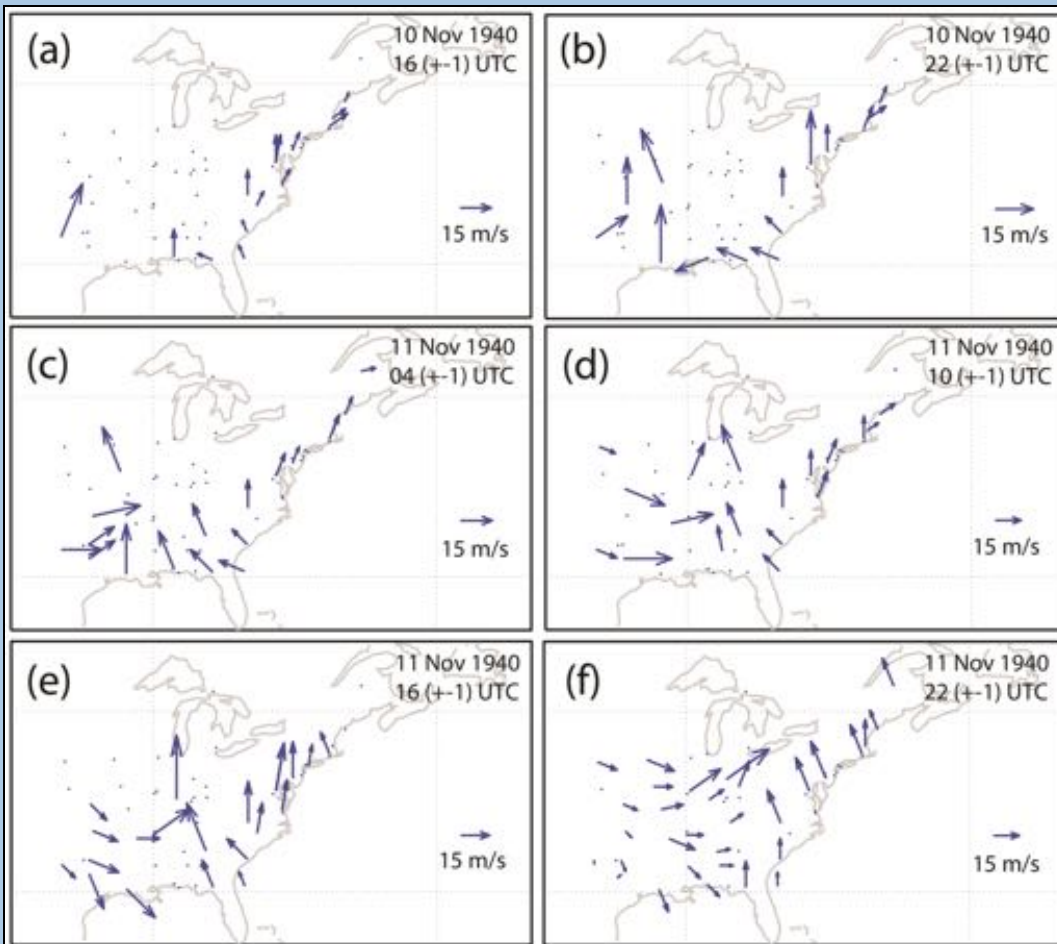
Number of records vs. time

The Comprehensive Historical Upper Air Network (CHUAN)



Altitude of ascents reached vs. time

The Comprehensive Historical Upper Air Network (CHUAN)



Data can even be used for analyses on synoptic timescales, e.g. of extreme events!

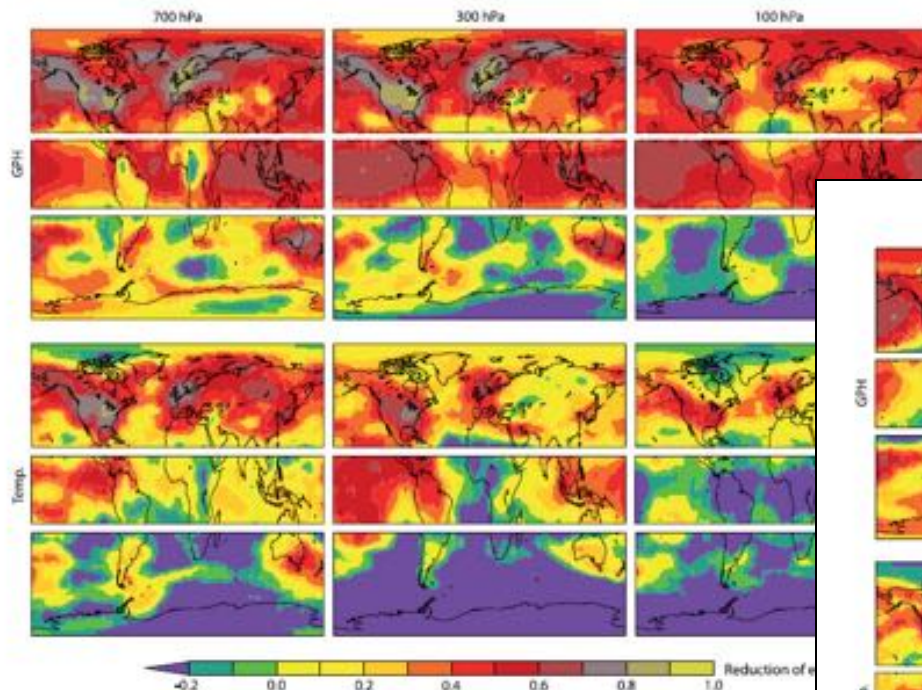
Armistice Day Storm
10/11 November 1940

Statistical reconstructions

- > REC1 (Griesser et al. 2010)
 - Also available on www.historicalupperair.org
 - Hemispheric and Tropics only PC and multiple linear regression
 - Using historical upper air and surface data
 - GPH and T up to 100 hPa level back to 1880
 - Monthly time resolution
 - **Advantage: spatial and temporal completeness**
 - **Potential disadvantage: Stationarity assumption of large-scale circulation patterns implied in PC analysis**
- > REC2 (Brönnimann et al. 2010)
 - As REC1, but:
 - Back to 1918 only
 - Local PC and multiple regression
 - **Advantage: only local pattern stationarity assumed**
 - **Disadvantage: spatially and temporally incomplete**

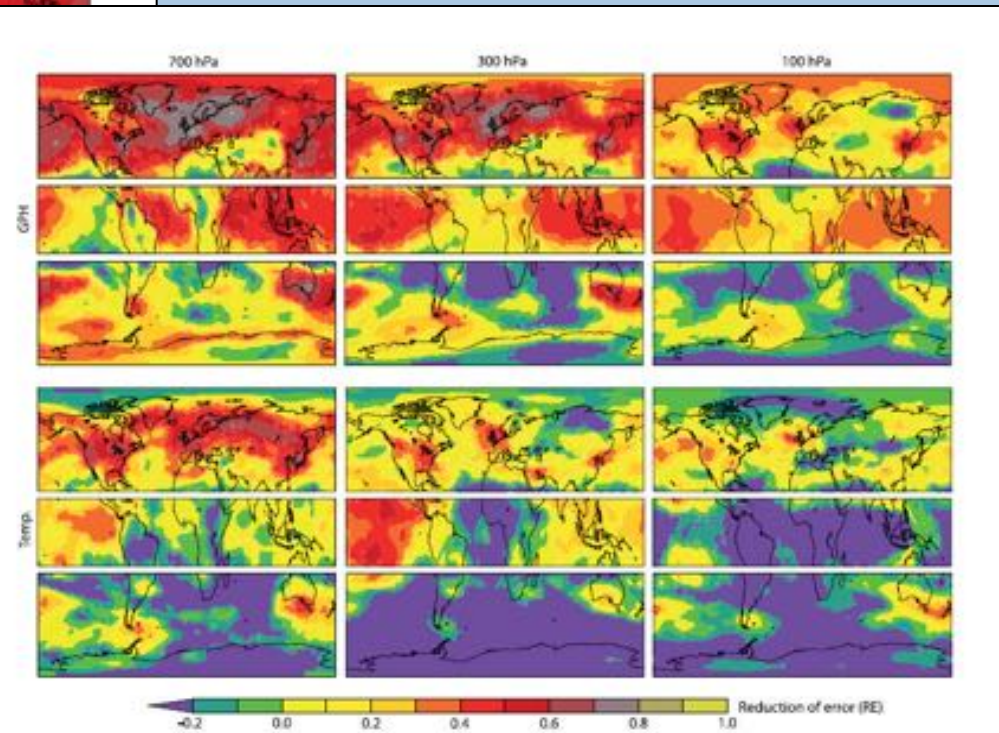
Statistical reconstructions

$$RE = 1 - \frac{\sum_t (x_{rec} - x_{obs})^2}{\sum_t (x_{null} - x_{obs})^2}$$



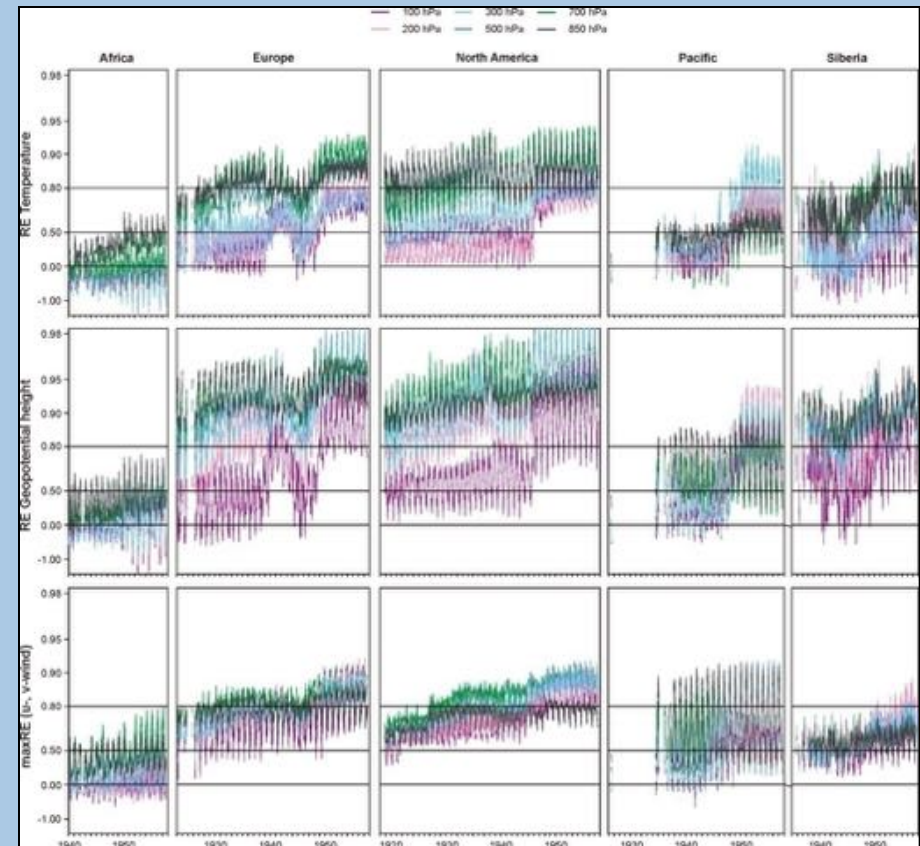
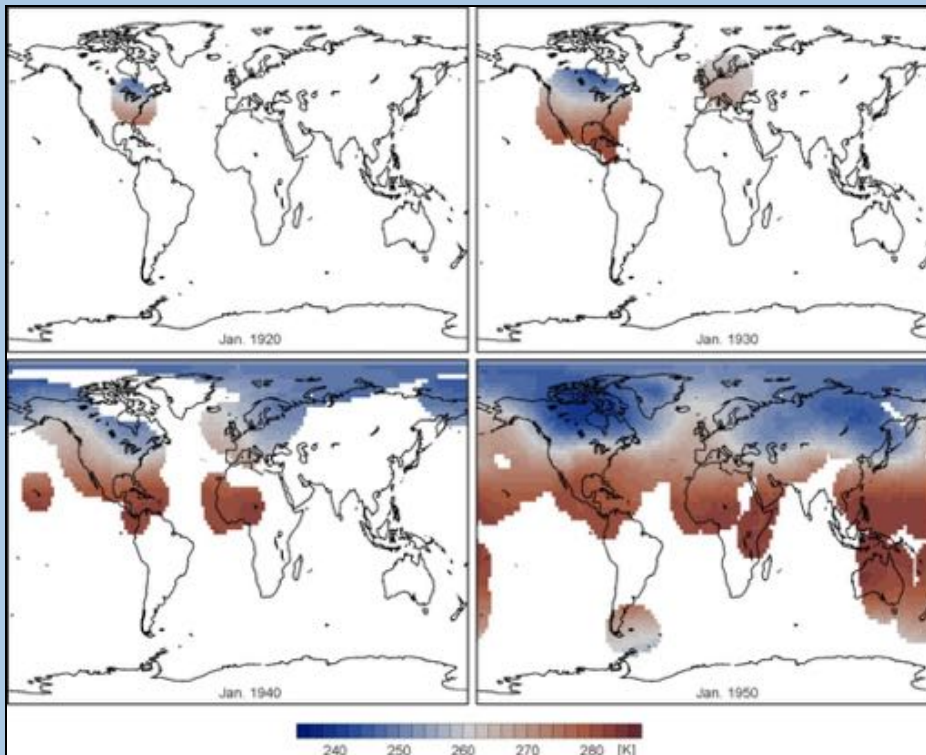
1940-1944

REC1: reconstruction skill



1900-1904

Statistical reconstructions



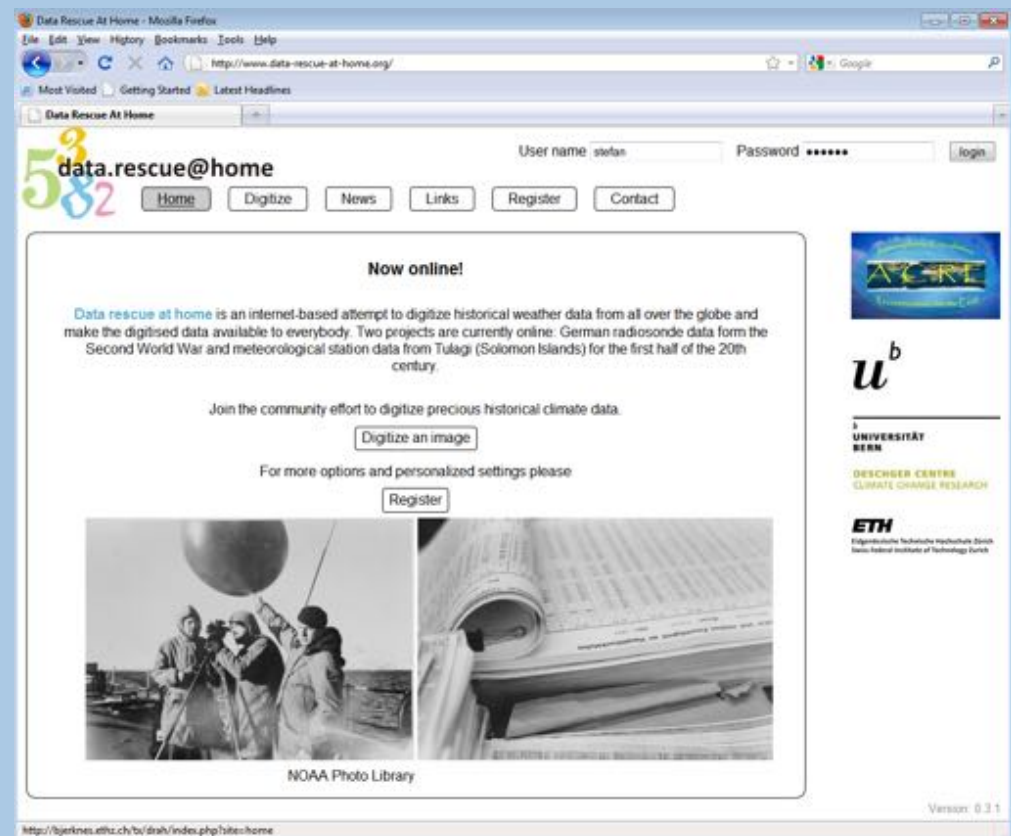
REC2

Coverage

RE skill

data.rescue@home

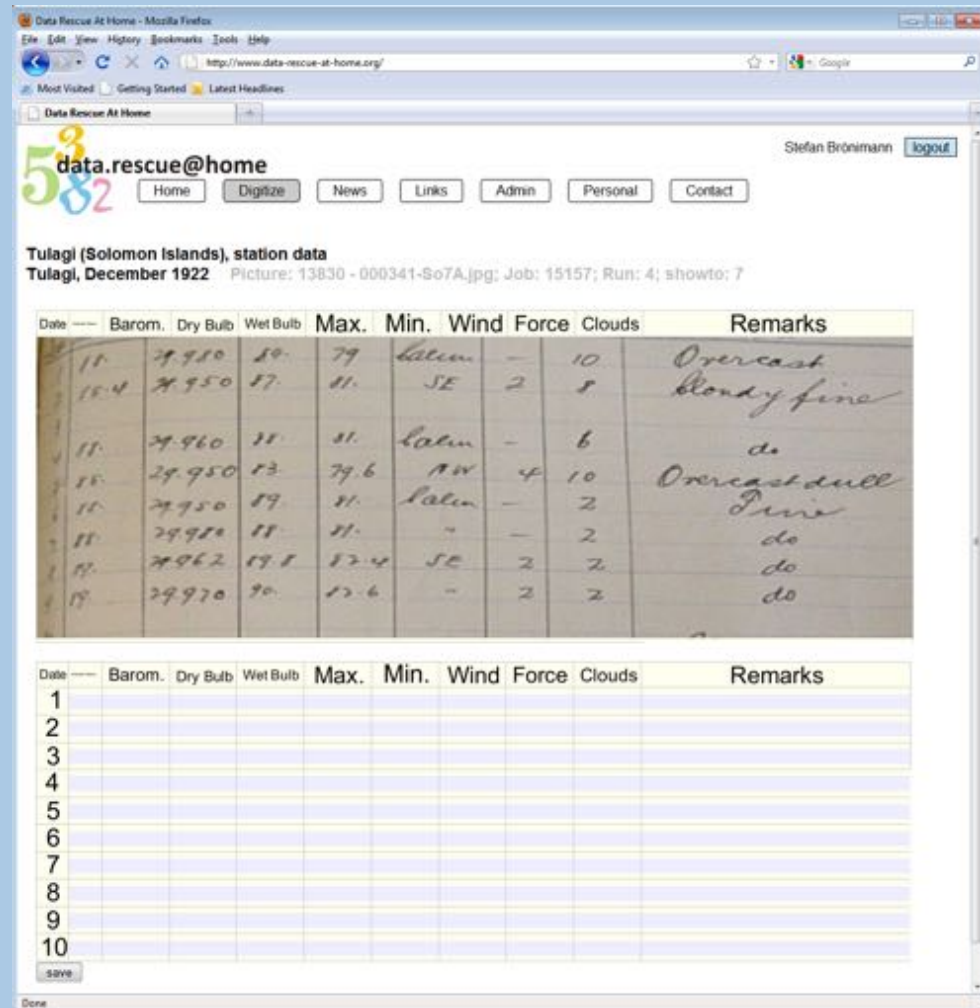
- > New data rescue project
- > <http://www.data-rescue-at-home.org/>
- > At the moment: 3600 photos of upper air data from
 - Tulagi (Solomon Islands, 1909-23, 1936-41)
 - WW II data from Germany and occupied countries
- > Possibility to register
 - Access to digitisation statistics
 - Planned: digitisation contest with ranking and web community / forum for people interested in historical weather data (cf. Zooniverse)



data.rescue@home

Main challenges:

1. Good quality of photos
→ professional digitisation equipment will be purchased (fixture for source and camera, fixed lighting equipment)
2. Many different source formats
→ entry masks need to be created manually



Tulagi (Solomon Islands), station data
Tulagi, December 1922 Picture: 13830 - 000341-So7A.jpg; Job: 15157; Run: 4; showto: 7

Date	Barom.	Dry Bulb	Wet Bulb	Max.	Min.	Wind Force	Clouds	Remarks
11.	29.910	89.	79	Calm	-	10	Overcast	
12.	29.950	87.	81.	SE	2	8	Bloody fine	
13.	29.960	88.	81.	Calm	-	6	do	
14.	29.950	83.	79.6	NW	4	10	Overcast dull	
15.	29.950	89.	81.	Calm	-	2	Fine	
16.	29.980	88.	81.	-	-	2	do	
17.	29.962	89.8	82.4	SE	2	2	do	
18.	29.970	90.	82.6	-	-	2	do	

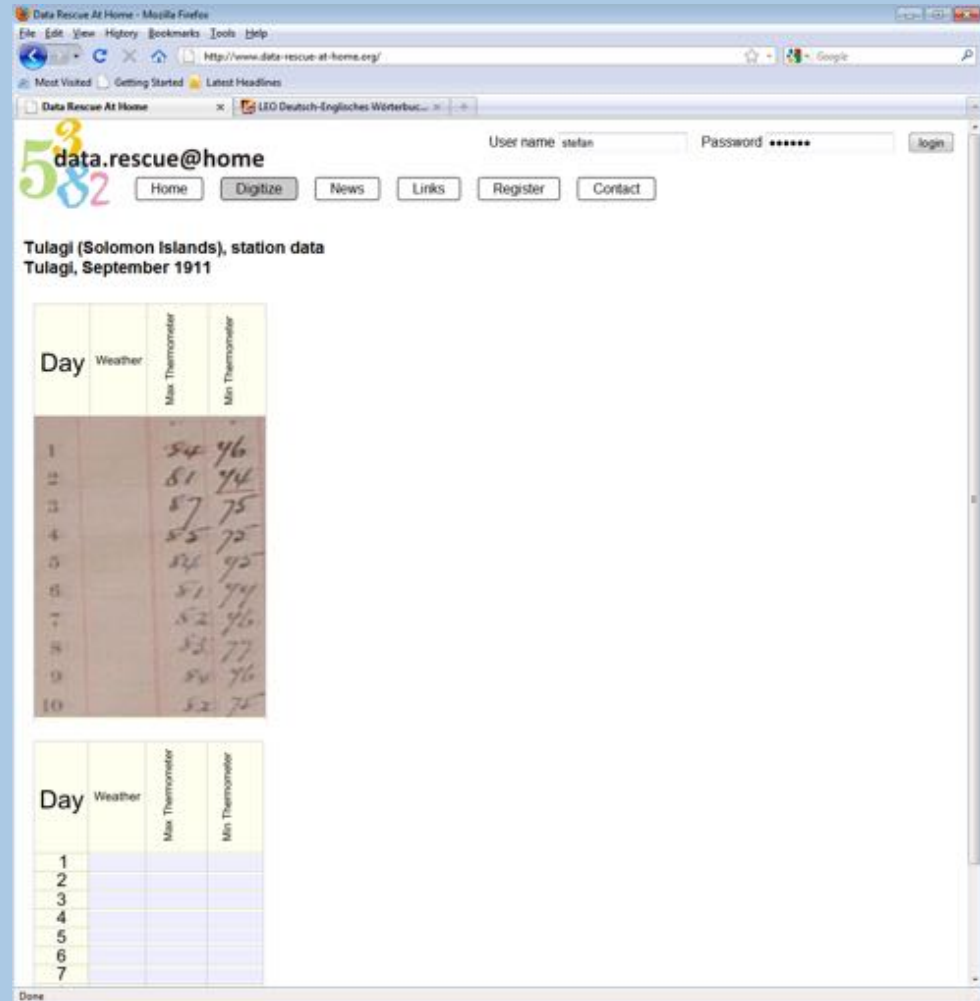
Date	Barom.	Dry Bulb	Wet Bulb	Max.	Min.	Wind Force	Clouds	Remarks
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								

save

data.rescue@home

Main challenges:

1. Good quality of photos
→ professional digitisation equipment will be purchased (fixture for source and camera, fixed lighting equipment)
2. Many different source formats
→ entry masks need to be created manually



The screenshot shows the 'data.rescue@home' website interface. At the top, there is a navigation bar with buttons for 'Home', 'Digitize', 'News', 'Links', 'Register', and 'Contact'. Below this, the site title 'data.rescue@home' is displayed. A login form is visible with fields for 'User name' (containing 'stata') and 'Password' (masked with asterisks), and a 'login' button. The main content area is titled 'Tulagi (Solomon Islands), station data' and 'Tulagi, September 1911'. It features two data entry tables. The first table is a photograph of a handwritten data sheet with columns for 'Day', 'Weather', 'Max Thermometer', and 'Min Thermometer'. The data is as follows:

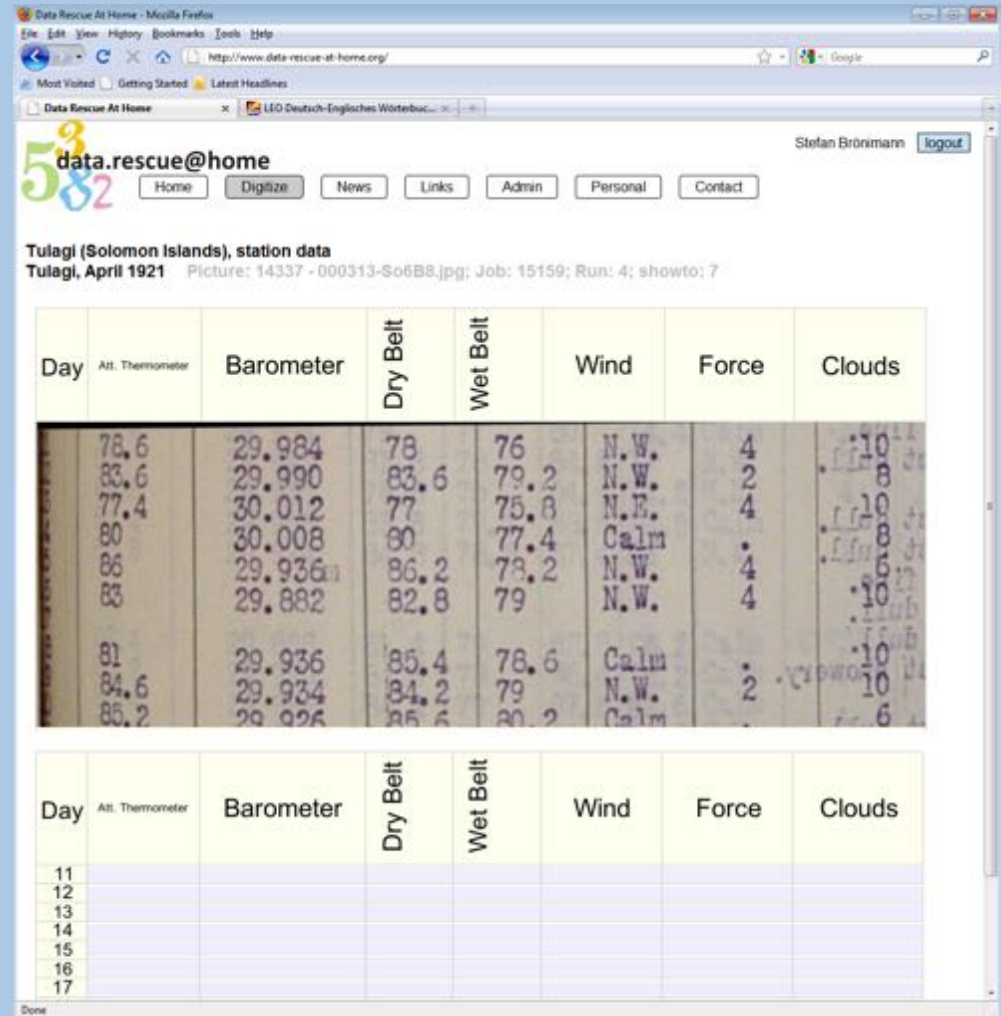
Day	Weather	Max Thermometer	Min Thermometer
1		84	76
2		81	74
3		87	75
4		85	70
5		84	70
6		81	74
7		82	76
8		83	77
9		84	76
10		82	75

The second table is a digital data entry form with the same columns and rows (1-7), currently empty.

data.rescue@home

Main challenges:

1. Good quality of photos
→ professional digitisation equipment will be purchased (fixture for source and camera, fixed lighting equipment)
2. Many different source formats
→ entry masks need to be created manually



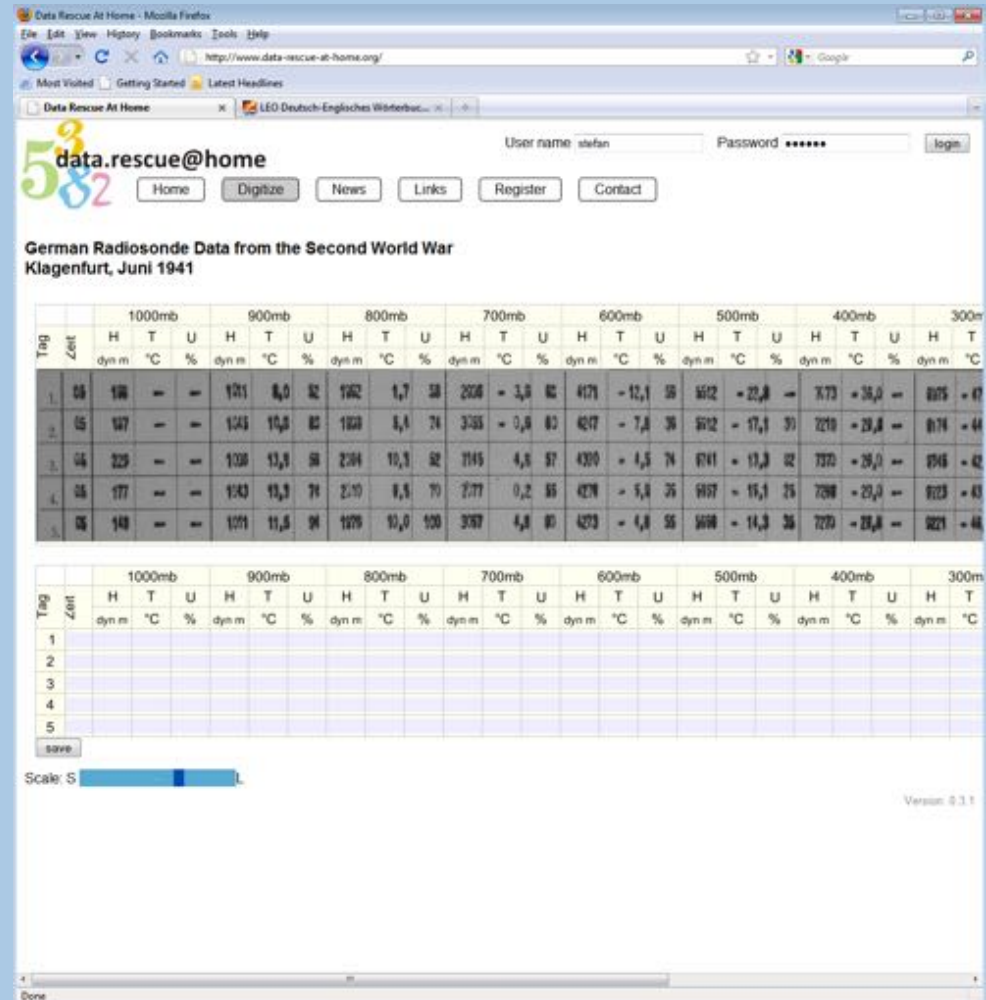
Tulagi (Solomon Islands), station data
Tulagi, April 1921 Picture: 14337 - 000313-So6B8.jpg; Job: 15159; Run: 4; showto: 7

Day	Alt. Thermometer	Barometer	Dry Belt	Wet Belt	Wind	Force	Clouds
	78.6	29.984	78	76	N.W.	4	10
	83.6	29.990	83.6	79.2	N.W.	2	8
	77.4	30.012	77	75.8	N.E.	4	10
	80	30.008	80	77.4	Calm	.	8
	86	29.936	86.2	78.2	N.W.	4	6
	83	29.882	82.8	79	N.W.	4	10
	81	29.936	85.4	78.6	Calm	.	10
	84.6	29.934	84.2	79	N.W.	2	10
	86.2	29.926	85.6	80.2	Calm	.	6
11							
12							
13							
14							
15							
16							
17							

data.rescue@home

Main challenges:

1. Good quality of photos
→ professional digitisation equipment will be purchased (fixture for source and camera, fixed lighting equipment)
2. Many different source formats
→ entry masks need to be created manually



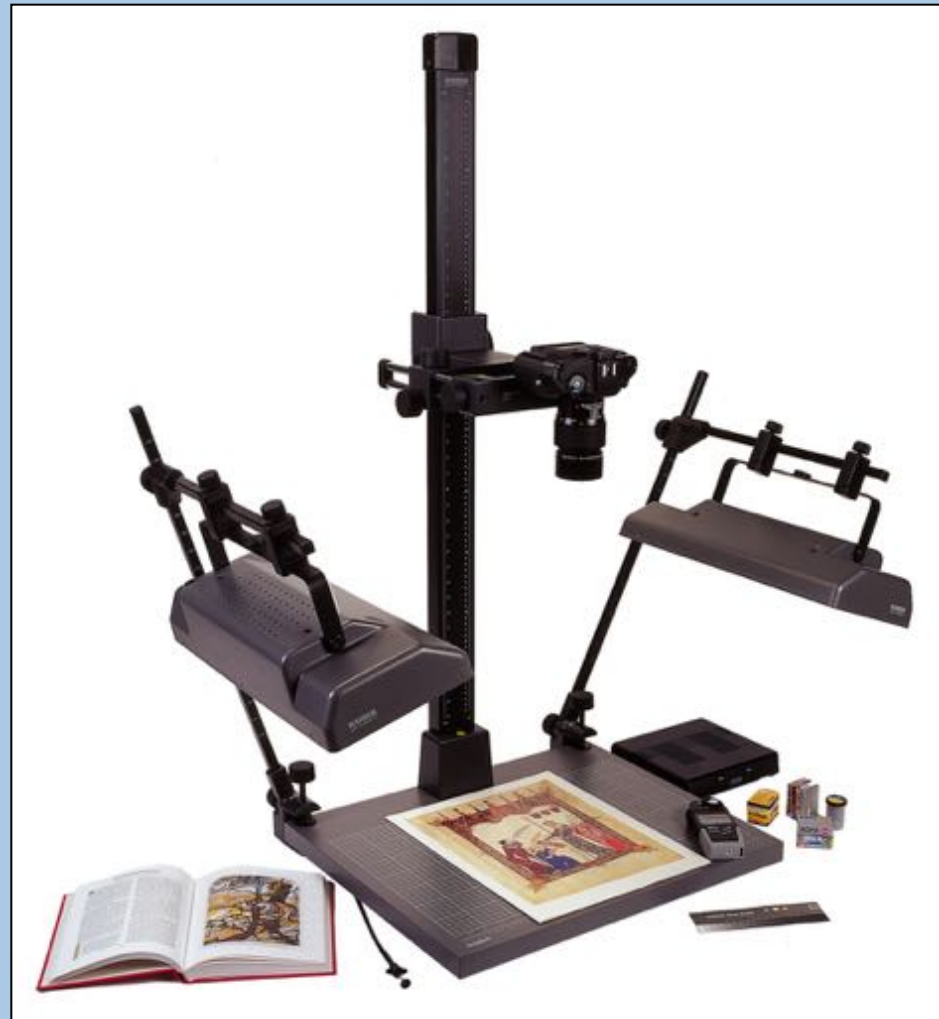
The screenshot shows the website interface for 'data.rescue@home'. It includes a login form with fields for 'User name' (stefan) and 'Password' (*****), and buttons for 'Home', 'Digitize', 'News', 'Links', 'Register', and 'Contact'. Below the login form is a section titled 'German Radiosonde Data from the Second World War' with a sub-header 'Klagenfurt, Juni 1941'. This section contains two data tables. The first table is a completed data entry for June 1941, showing atmospheric data at various pressure levels (1000mb to 300mb) for five different times of day (15:00, 16:00, 17:00, 18:00, 19:00). The second table is an empty data entry form for the same location and time, with columns for 'Tag', 'Zeit', and various atmospheric parameters (H, T, U, %).

Tag	Zeit	1000mb			900mb			800mb			700mb			600mb			500mb			400mb			300mb		
		H	T	U	H	T	U	H	T	U	H	T	U	H	T	U	H	T	U	H	T	U	H	T	U
1	15	198	-	-	1011	15,0	82	1062	1,7	58	2008	- 3,8	82	4171	- 12,1	58	6612	- 22,8	-	1073	- 35,0	-	8875	- 47	
2	16	187	-	-	1043	15,0	82	1058	1,1	71	3355	- 0,8	80	4217	- 7,8	58	6612	- 17,1	50	7219	- 29,8	-	8174	- 44	
3	17	229	-	-	1038	15,3	58	2384	10,3	82	1145	4,8	57	4300	- 4,5	74	6741	- 17,3	82	7370	- 26,0	-	8748	- 42	
4	18	177	-	-	1043	15,3	71	2110	1,5	70	2777	0,2	58	4278	- 5,4	26	6637	- 15,1	75	7288	- 23,0	-	8123	- 43	
5	19	148	-	-	1051	11,8	94	1876	10,0	100	3787	4,8	80	4273	- 4,8	58	5688	- 14,3	58	7270	- 28,8	-	8221	- 48	

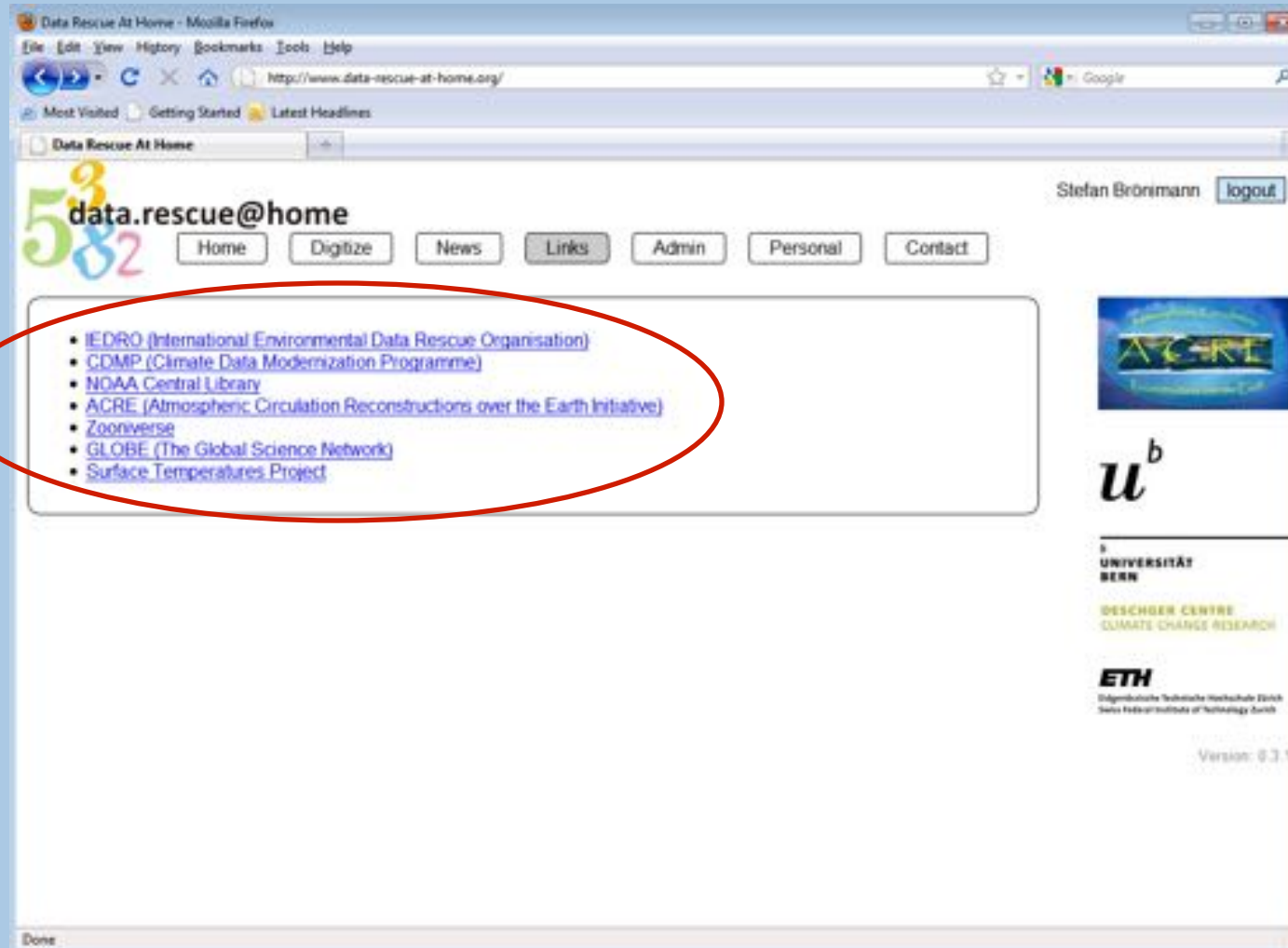
Tag	Zeit	1000mb			900mb			800mb			700mb			600mb			500mb			400mb			300mb		
		H	T	U	H	T	U	H	T	U	H	T	U	H	T	U	H	T	U	H	T	U	H	T	U
1																									
2																									
3																									
4																									
5																									

data.rescue@home

Planned:
More professional
digitisation equipment



data.rescue@home



Data Rescue At Home - Mozilla Firefox
http://www.data-rescue-at-home.org/

data.rescue@home

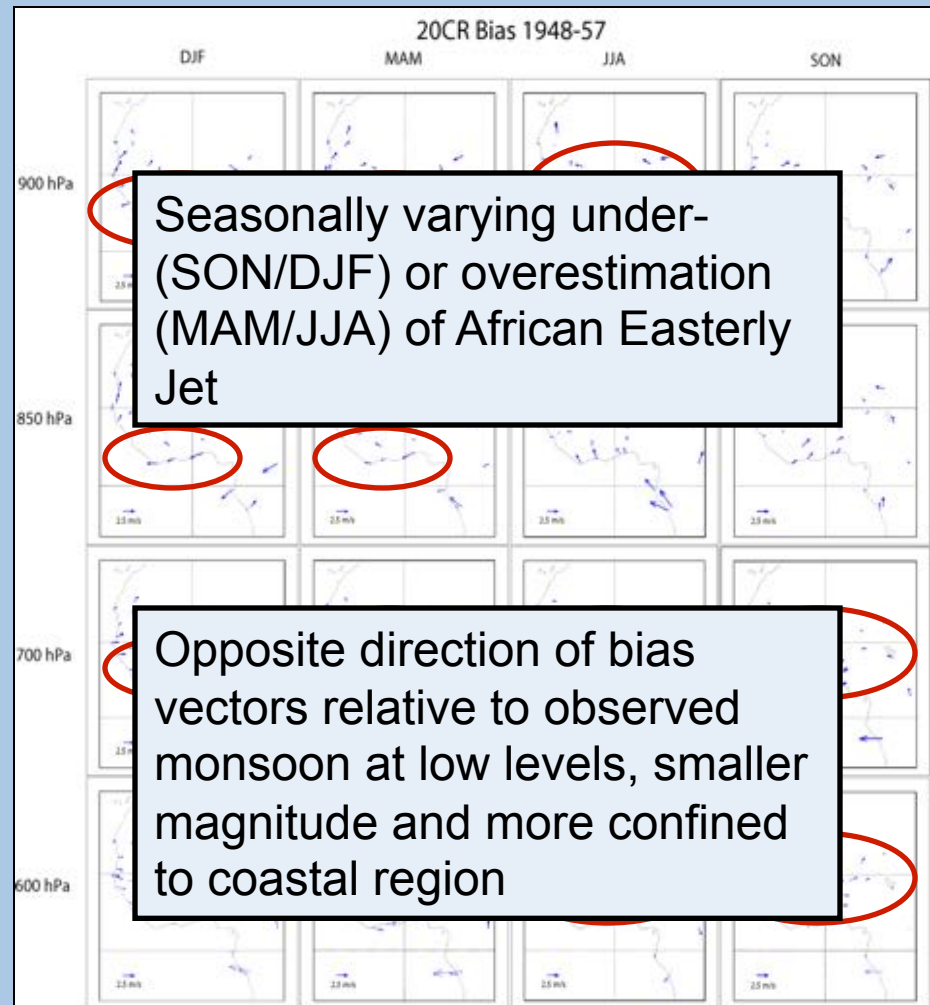
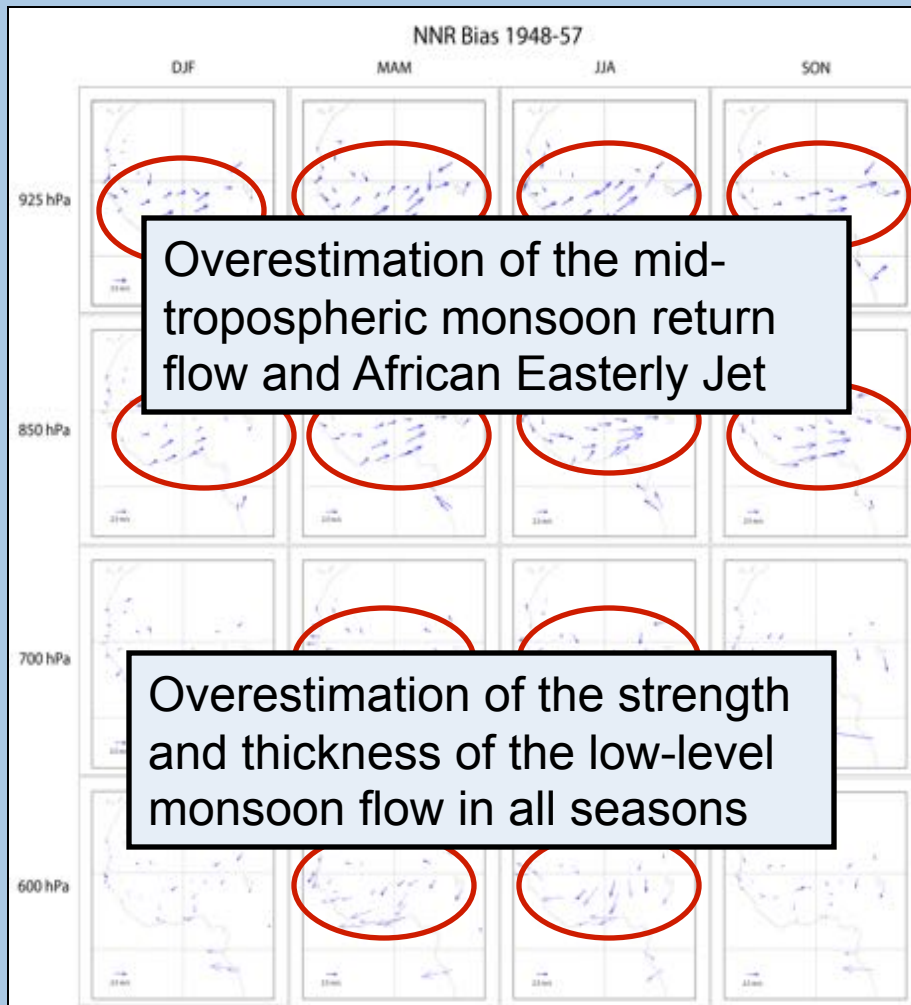
Stefan Brörmann [logout](#)

Home Digitize News **Links** Admin Personal Contact

- [IEDRO \(International Environmental Data Rescue Organisation\)](#)
- [CDMP \(Climate Data Modernization Programme\)](#)
- [NOAA Central Library](#)
- [ACRE \(Atmospheric Circulation Reconstructions over the Earth Initiative\)](#)
- [Zooniverse](#)
- [GLOBE \(The Global Science Network\)](#)
- [Surface Temperatures Project](#)

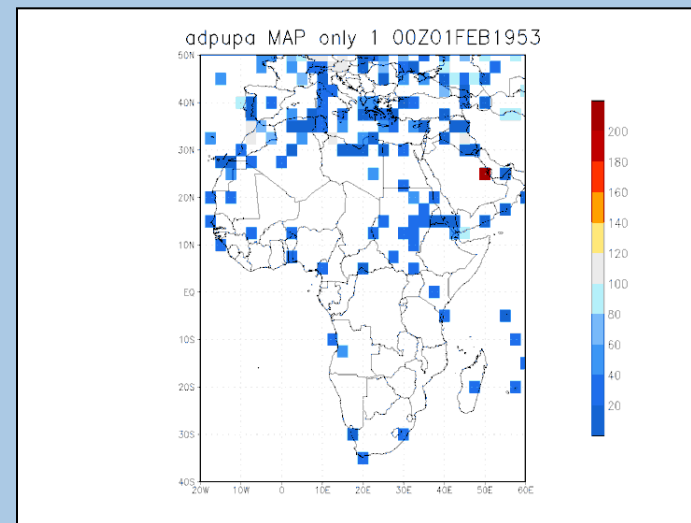
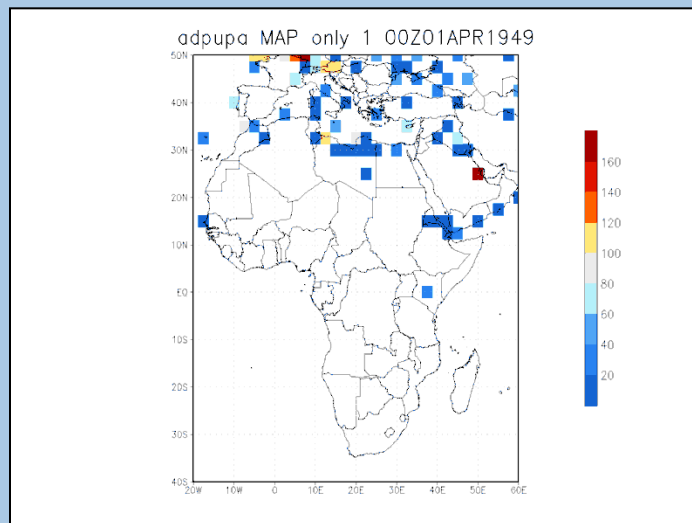
u^b
UNIVERSITÄT
BERN
OESCHGER CENTRE
CLIMATE CHANGE RESEARCH
ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich
Version: 0.3.1

Wind biases of NNR and 20CR w.r.t. CHUAN observations



Wind biases of NNR and 20CR w.r.t. CHUAN observations

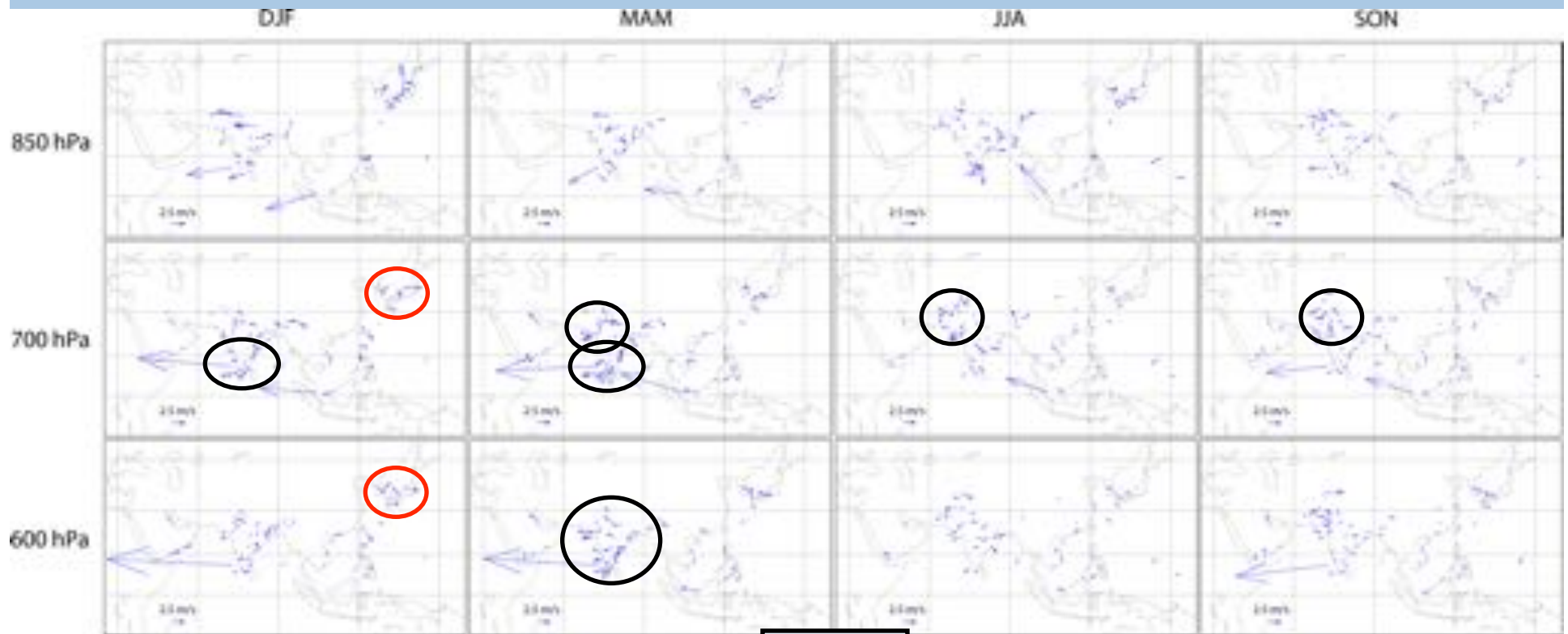
- > Significant, seasonally varying and spatially coherent biases
- > Interestingly weaker for 20CR than for NNR in the West African monsoon region
- > Up to > 5 m/s, similar or larger than seasonal cycle
- > Spatial structure suggests deficiency of reanalyses rather than observation
 - Probably connected to small number of assimilated upper air data in the pre-IGY period (NNR)



Upper air obs assimilated into NCEP reanalysis

http://nomad3.ncep.noaa.gov/cgi-bin/pdisp_r1_obs.sh

Wind biases of NNR and 20CR w.r.t. CHUAN observations



NNR

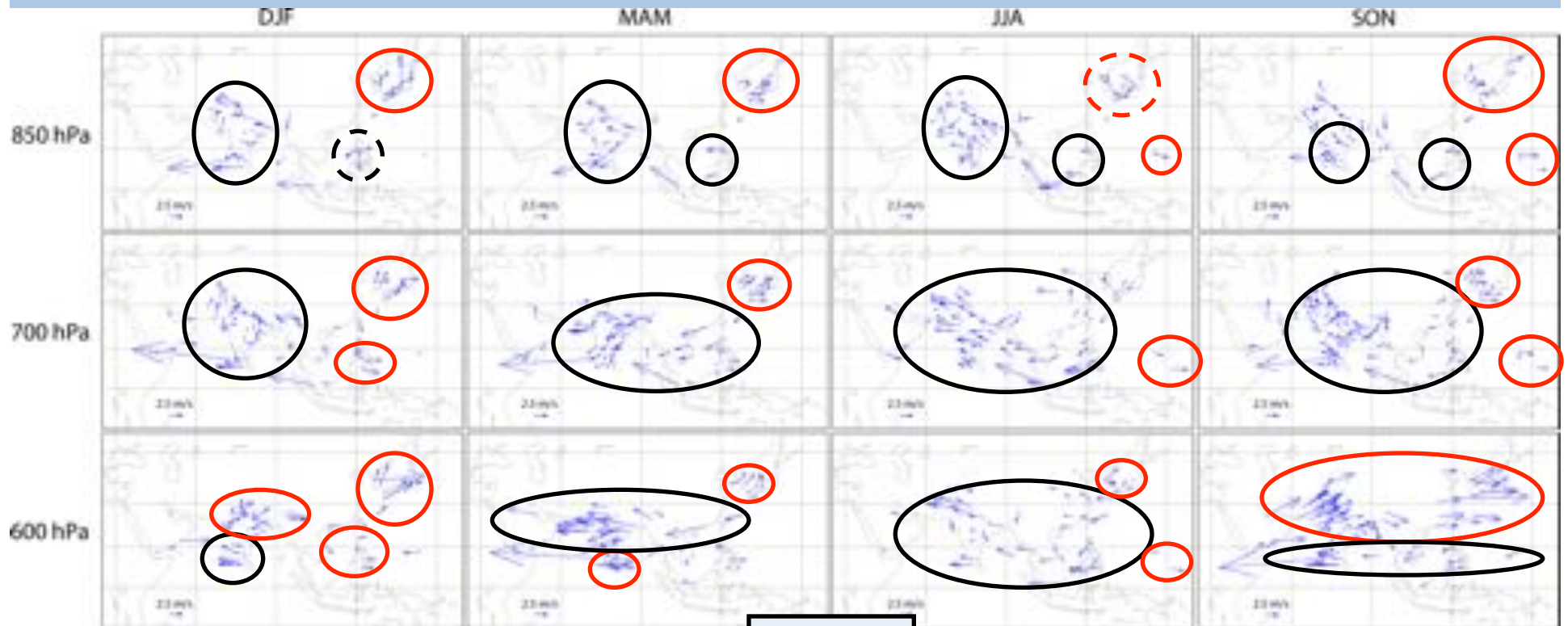


W bias



E bias

Wind biases of NNR and 20CR w.r.t. CHUAN observations



20CR



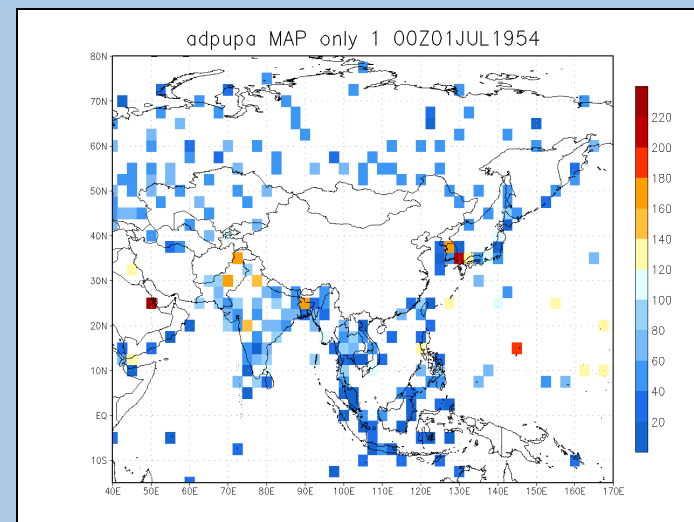
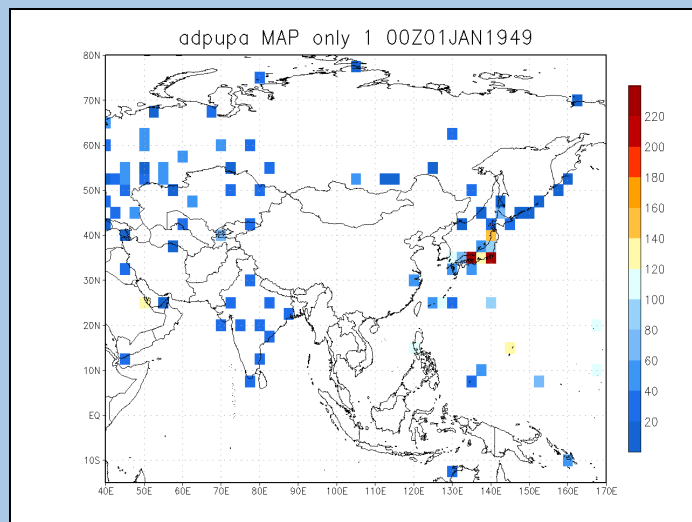
W bias



E bias

Wind biases of NNR and 20CR w.r.t. CHUAN observations

- > Similar, spatially organised biases
- > For the Asian monsoon region, 20CR shows stronger deviations than NNR where CHUAN data are available, especially at higher levels
- > Available data seem to have been assimilated in the pre-IGY period (NNR), but for China, Indochina, Malaysia and Indonesia CHUAN lacks data too



Upper air obs assimilated into NCEP reanalysis

http://nomad3.ncep.noaa.gov/cgi-bin/pdisp_r1_obs.sh

Conclusions

- > Much more data are available, but not yet digitised (e.g. NOAA Central Library (already photographed) and other sources on paper)
- > We will concentrate on early upper air data especially from the Tropics and Southern Hemisphere
- > Digitisation can be accelerated by making use of potential voluntary collaborators on the web, but preparative work has to be done manually
- > Intercomparison of different datasets is useful and reveals the respective strengths and weaknesses of the data (see also Brönnimann et al. 2009: Variability of large-scale atmospheric circulation indices for the Northern Hemisphere during the past 100 years)



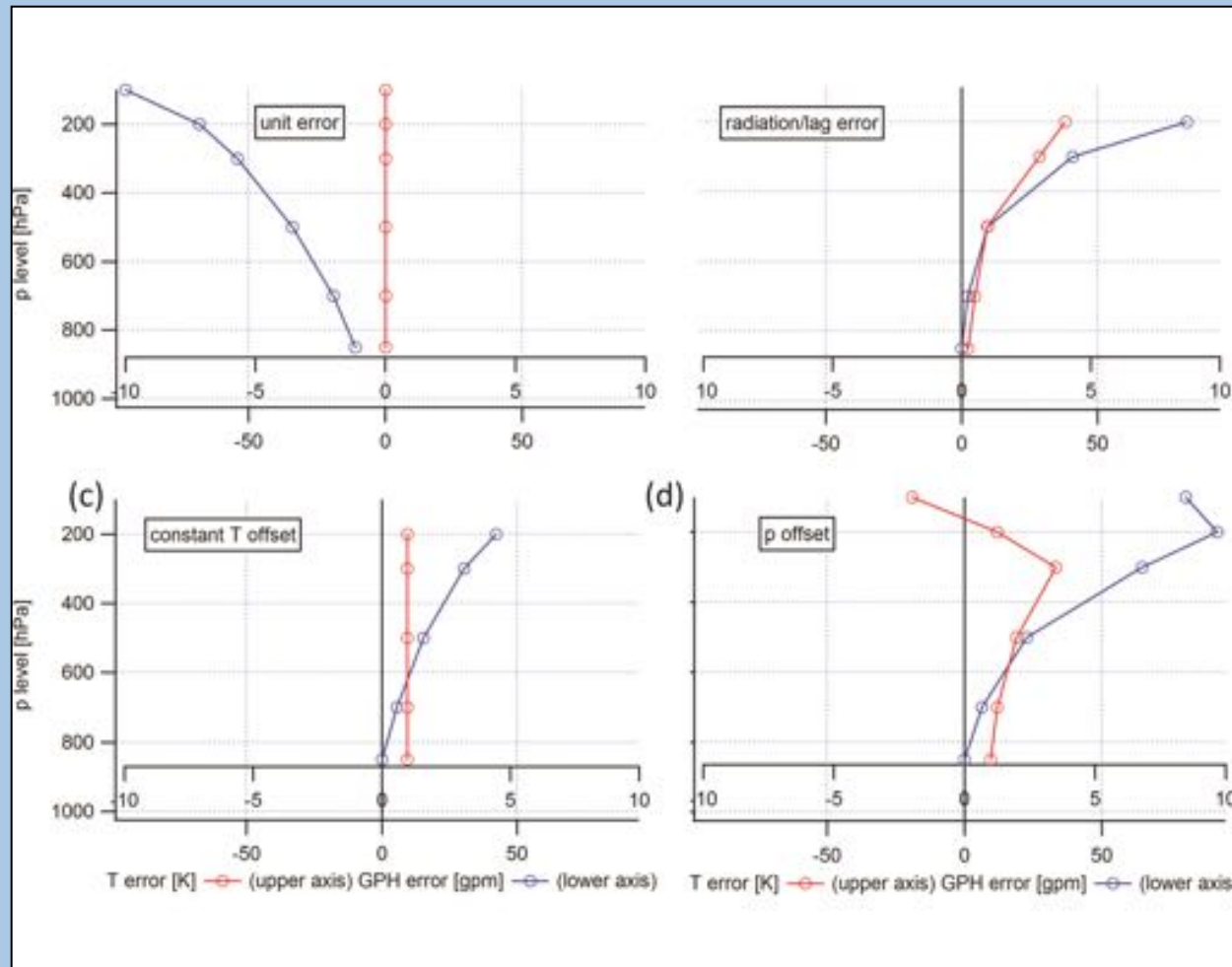
u^b

u^b
**UNIVERSITÄT
BERN**

**OESCHGER CENTRE
CLIMATE CHANGE RESEARCH**

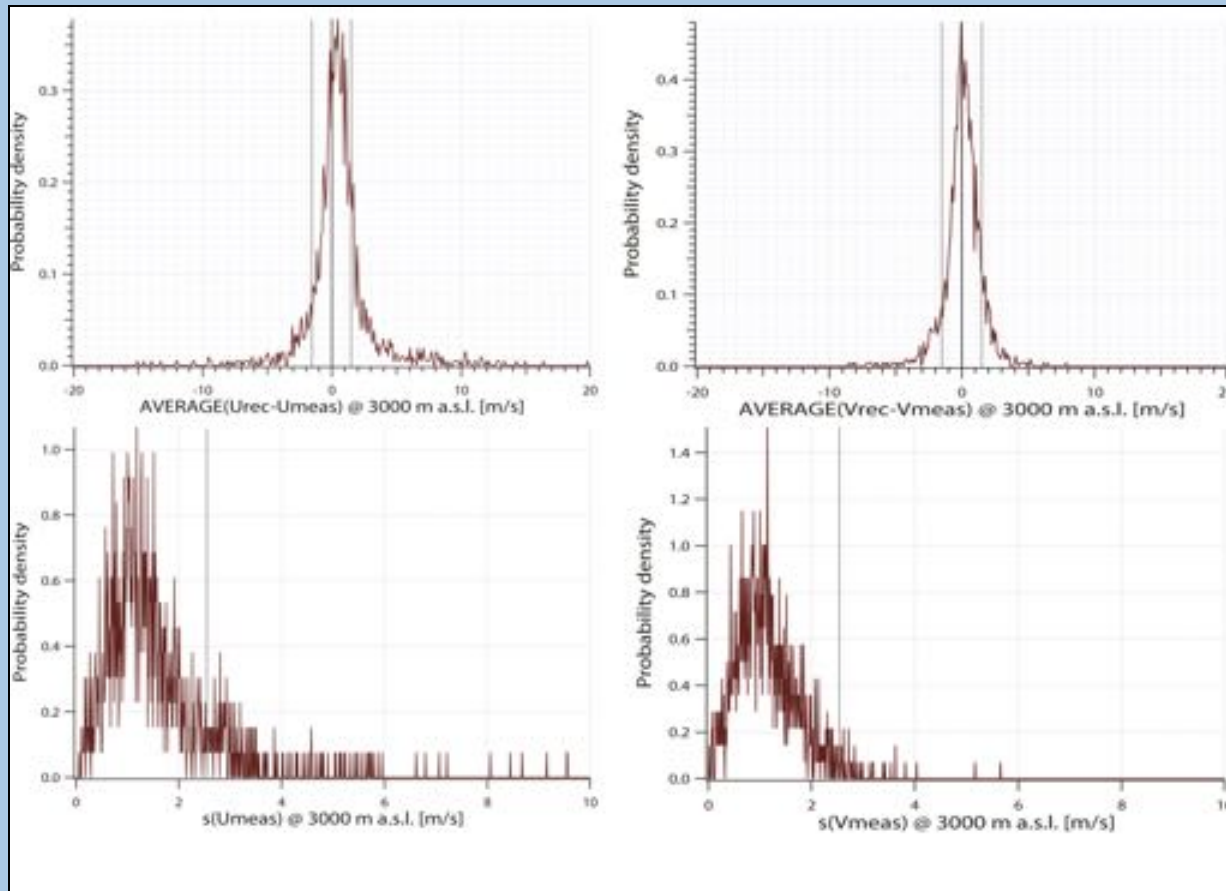
Thank you!

The Comprehensive Historical Upper Air Network (CHUAN)



Physics based correction of radiosonde data (Grant et al. 2009, *J. Clim.*)

The Comprehensive Historical Upper Air Network (CHUAN)



PDFs of mean and SD of the difference of all wind records from statistical reconstructions calibrated with ERA-40