Sparse data rescue

Future US Earth System Reanalysis Pre-Workshop Webinar Series

8 March 2022

Dr Linden Ashcroft |@lindenashcroft

School of Geography, Earth and Atmospheric Sciences

Image: State Library of QLD

Sparse data rescue

Future US Earth System Reanalysis Pre-Workshop Webinar Series

8 March 2022

Dr Linden Ashcroft

School of Geography, Earth and Atmospheric Sciences

Image: Darren England/AAP

Data Rescue

- Recovery of data from non-useable format to a a state that can be used for scientific analysis
- Includes not just identification and recovery of data sources, but rescue of metadata as well
- Finding, scanning, imaging, transcribing, compiling, formatting, quality controlling: time consuming, messy, laborious
- BUT, a cheaper way to get ‘new’ data than building a super computer, or launching a satellite
- Also a popular engagement activity for volunteers and citizen scientists

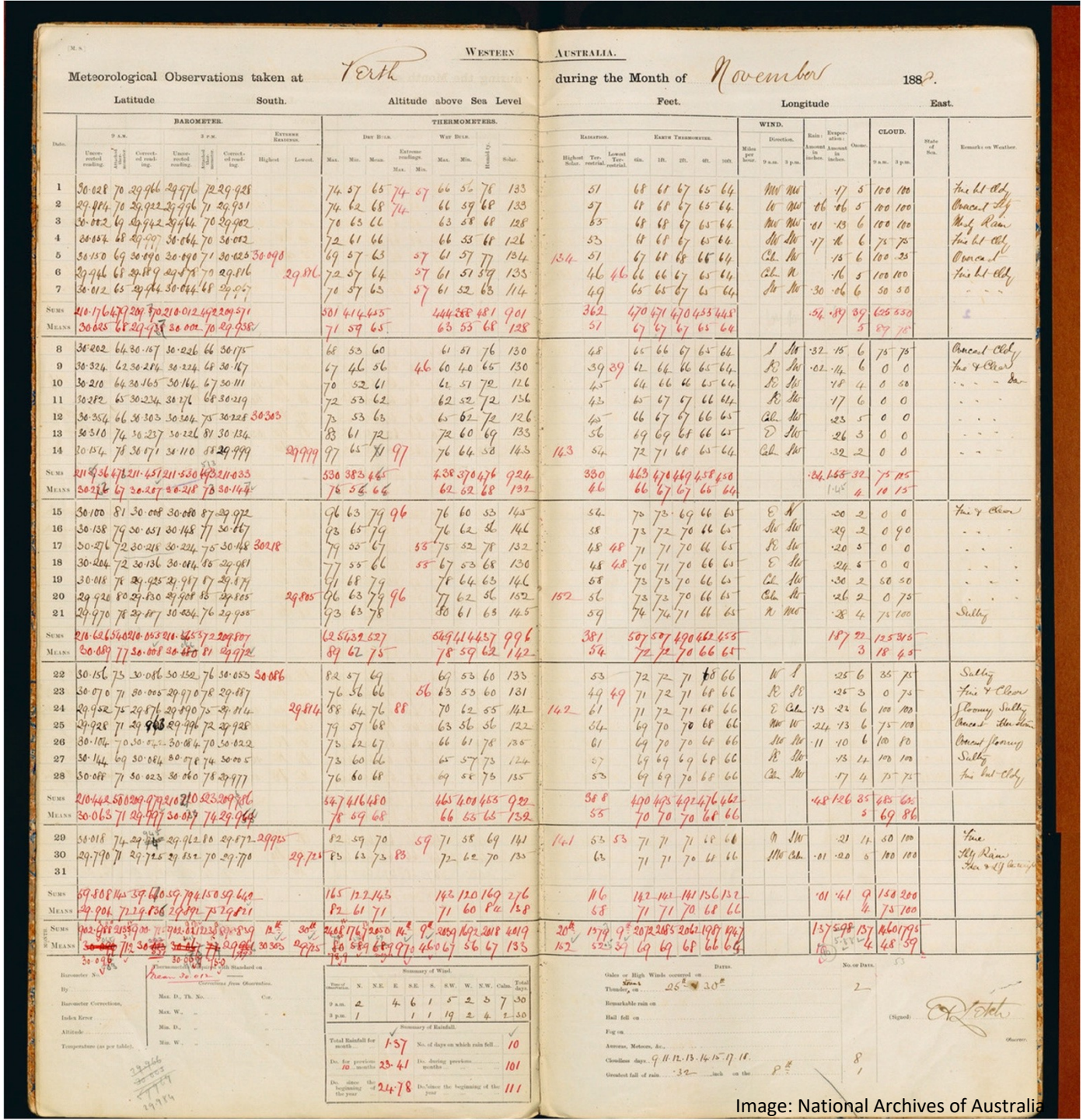
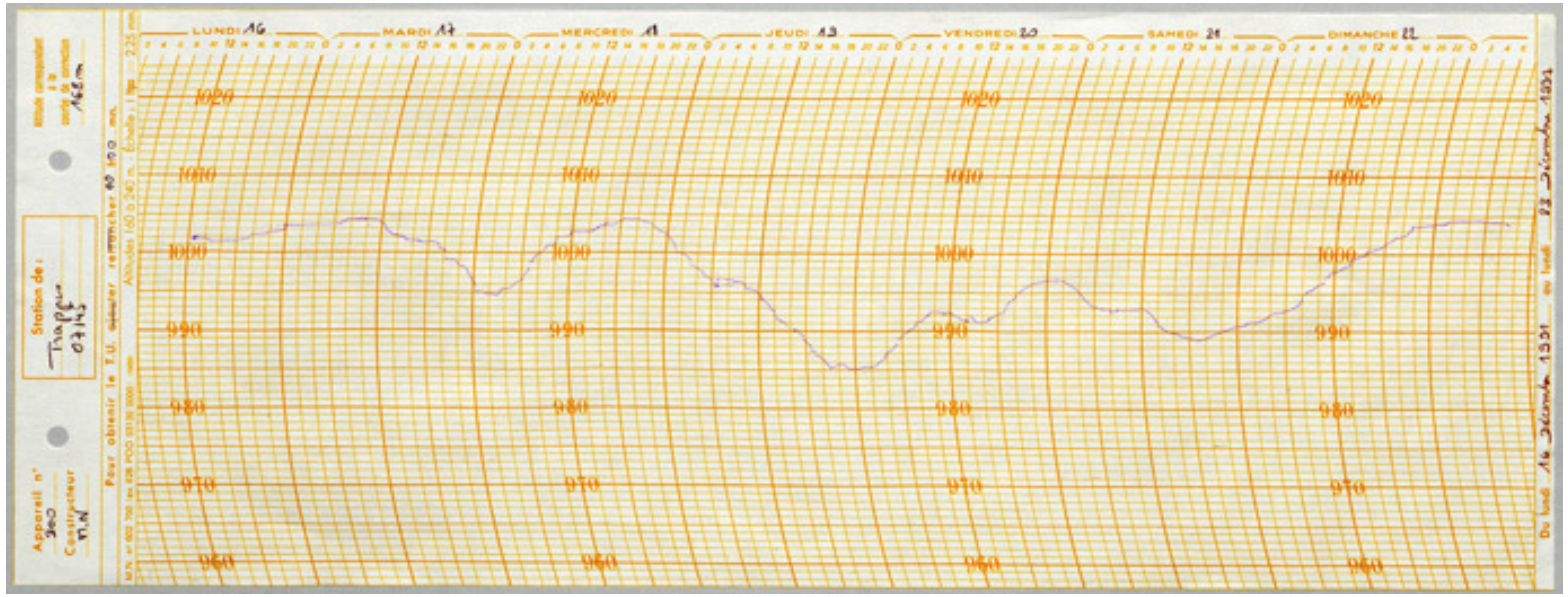


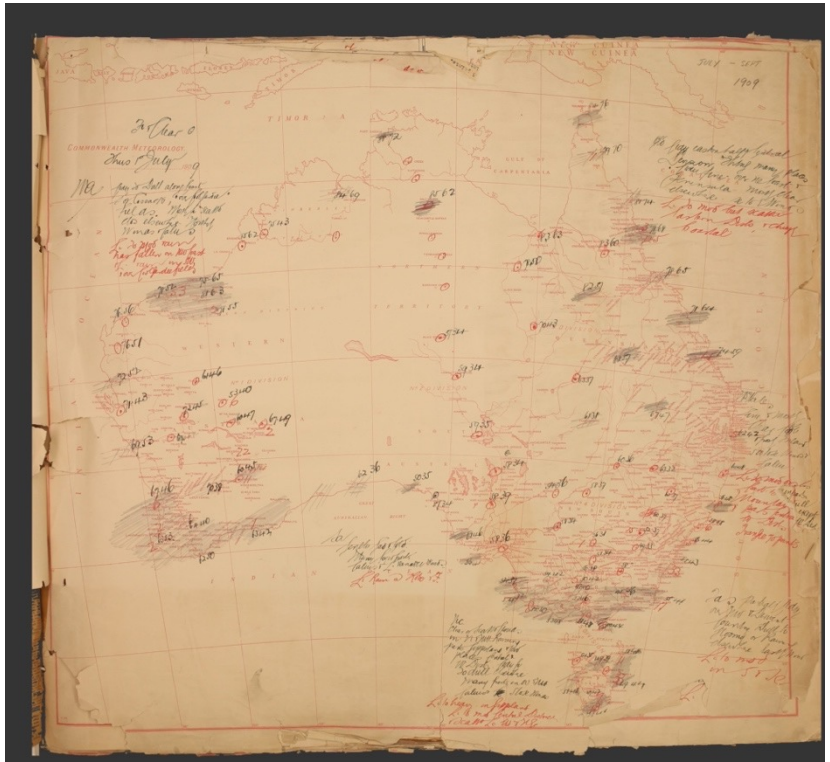
Image: National Archives of Australia

It is not just about discovering and digitising the data, but also involves bringing those data together in a consistent way and making them available to all. It is a huge task.

-ECMWF

Data rescue





Images: ECMWF, ACRE, CC

Data rescue

Image: WMO, CODATA

Old Weather

Help scientists transcribe Arctic and worldwide weather observations recorded in ship's logs since the mid-19th century.

First voyage *Zooniverse projects*

Old Weather: WWII

Recover hidden weather data collected by the Navy during World War II

The deep *More challenging projects*

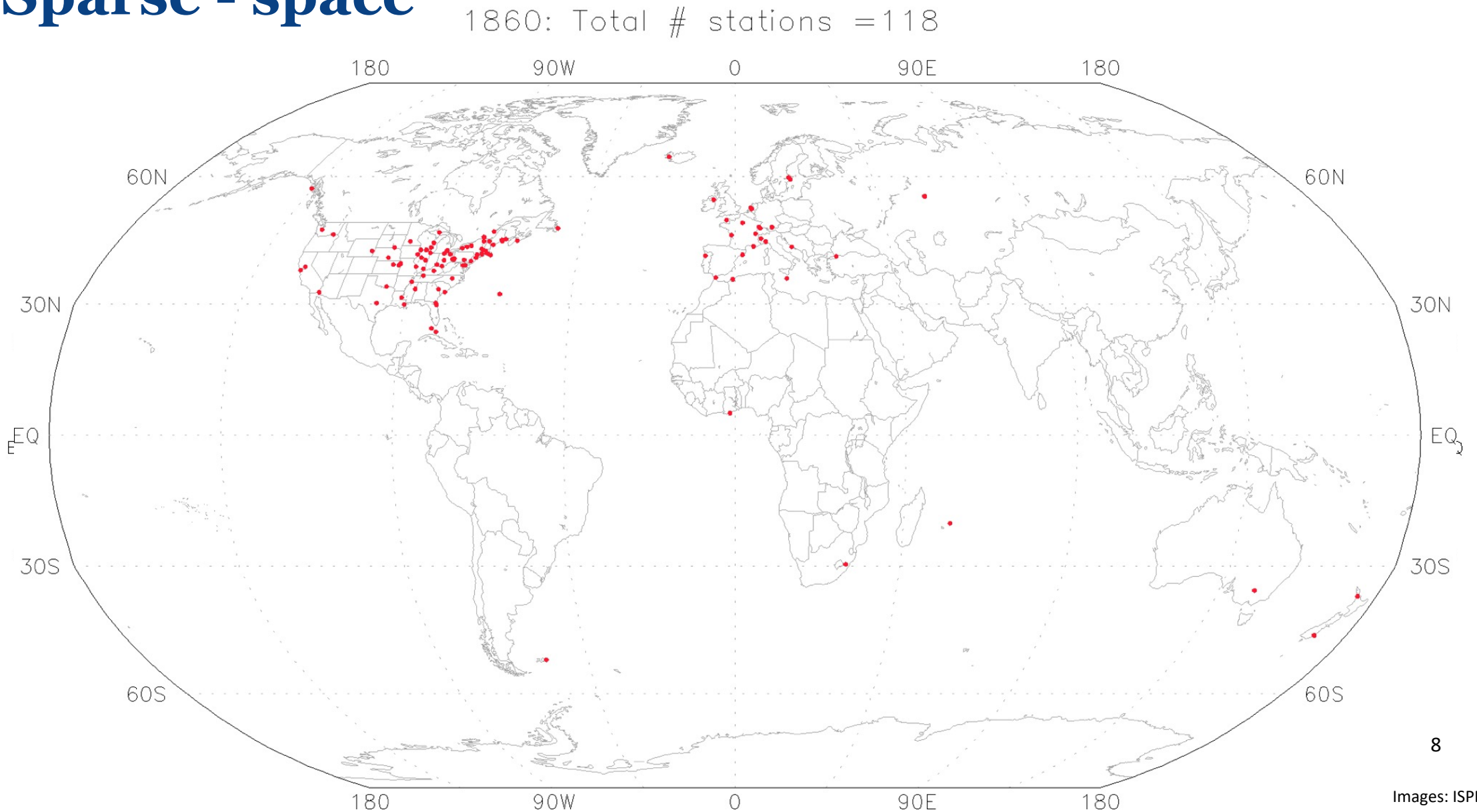
Old Weather: Whaling

Explore the Arctic of the past from the deck of a whaling ship

Old Weather: Arctic

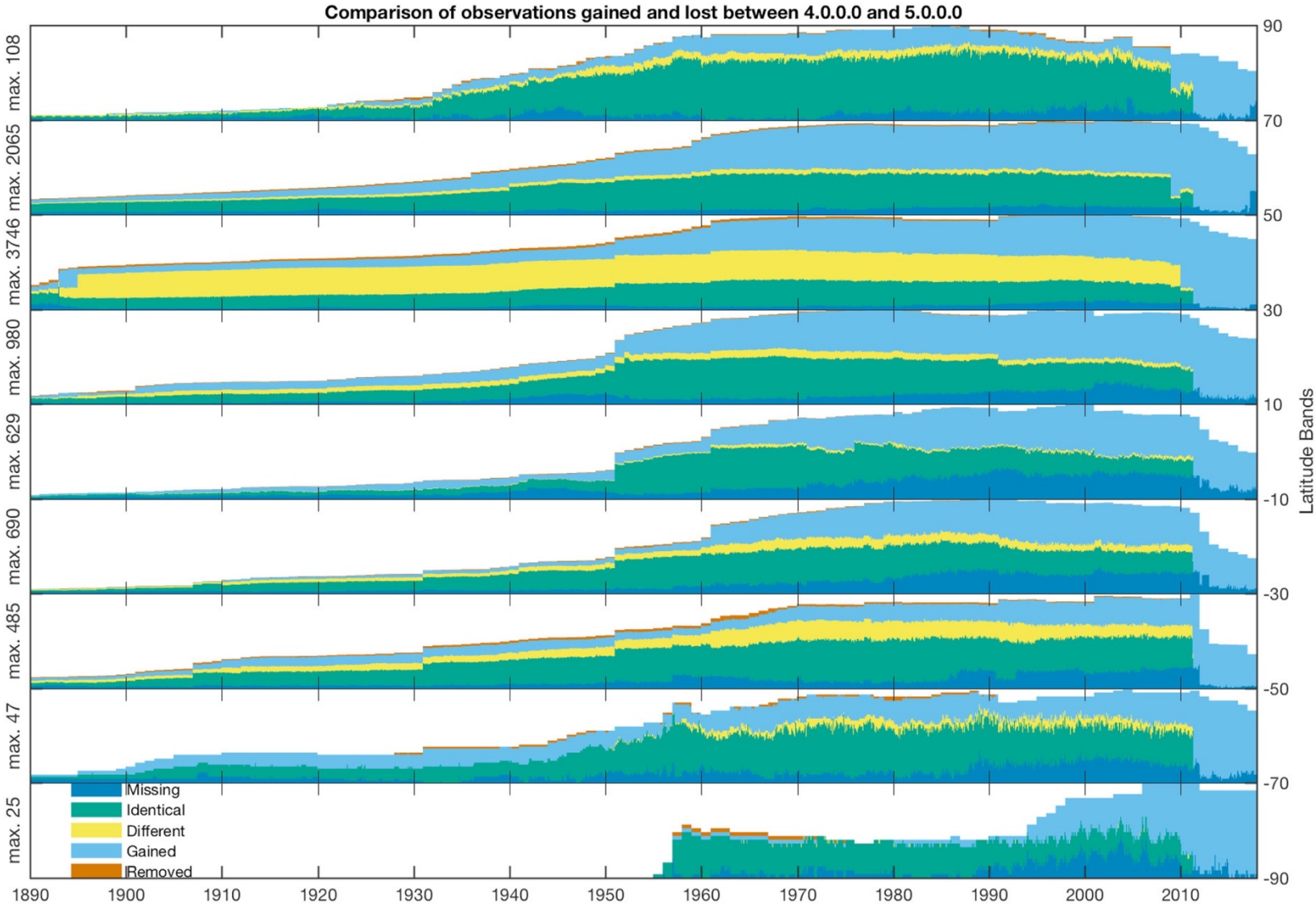
Rediscover the historic Arctic voyages of the U.S. Navy and Coast Guard

Sparse - space



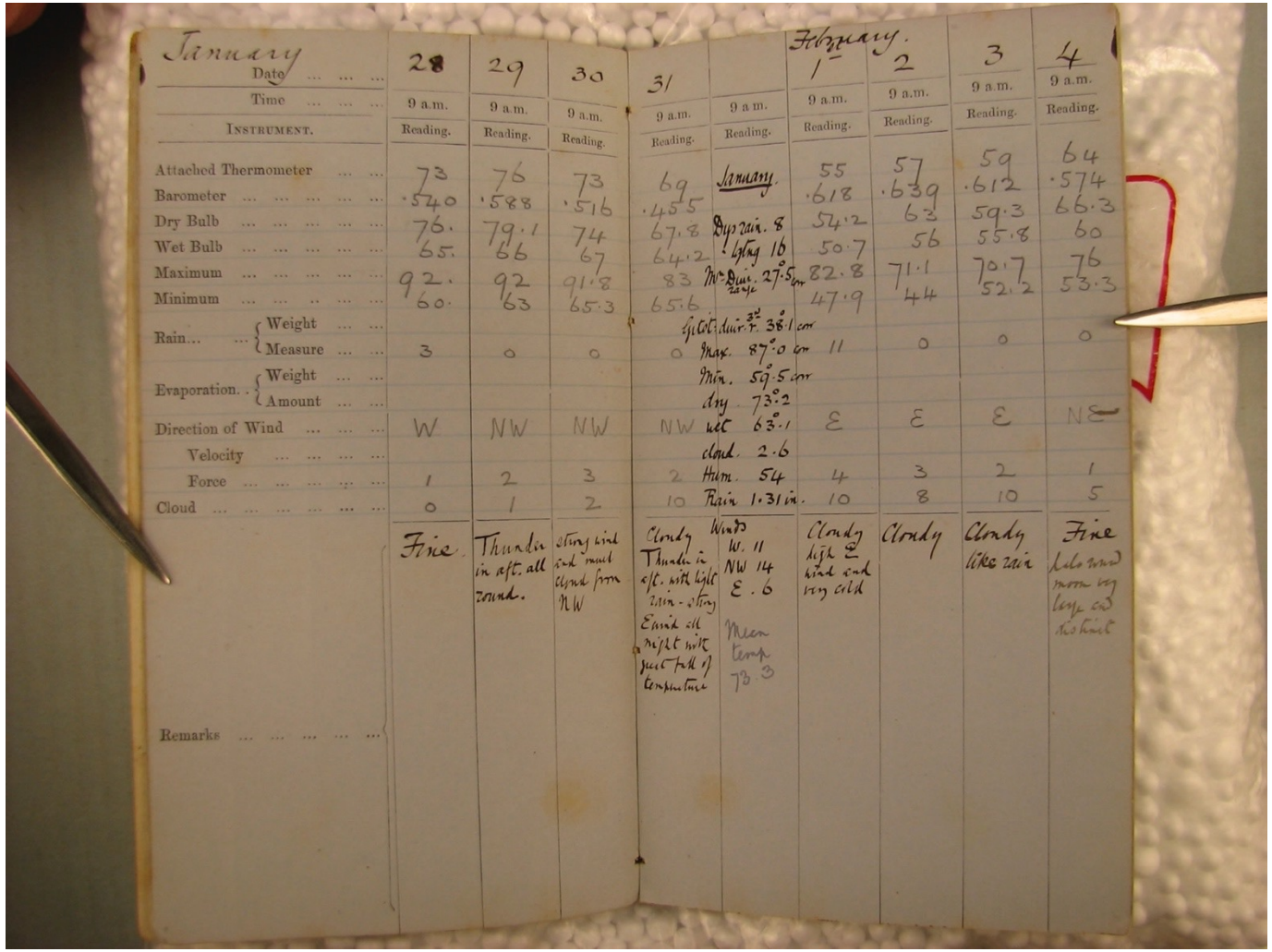
Images: ISPD

Sparse - time



CRUTEMP4.0 vs CRUTEMP5.0 input data (Osborn et al. 2021)

Sparse - variable



January

Date	28	29	30	31		February 1	2	3	4
Time	9 a.m.	9 a.m.	9 a.m.	9 a.m.	9 a.m.	9 a.m.	9 a.m.	9 a.m.	9 a.m.
Instrument.	Reading.	Reading.	Reading.	Reading.	Reading.	Reading.	Reading.	Reading.	Reading.
Attached Thermometer	73	76	73	69	January.	55	57	59	64
Barometer	·540	·588	·516	·455		·618	·639	·612	·574
Dry Bulb	76.	79·1	74	67·8	Dyp rain. 8	54·2	63	59·3	66·3
Wet Bulb	65.	66	67	64·2	Lyhg 16	50·7	56	55·8	60
Maximum	92.	92	91·8	83	Mn Diur. 27·5	82·8	71·1	70·7	76
Minimum	60.	63	65·3	65·6		47·9	44	52·2	53·3
Rain... Weight					Syl Col. diur. 38·1 cm				
Rain... Measure	3	0	0	0	Max. 87°0 cm	11	0	0	0
Evaporation.. Weight					Min. 59°5 cm				
Evaporation.. Amount					dry 73°2				
Direction of Wind	W	NW	NW	NW	wet 63°1	E	E	E	NE
Velocity					cloud 2·6				
Force	1	2	3	2	Hum. 54	4	3	2	1
Cloud	0	1	2	10	Rain 1·31 in.	10	8	10	5
Remarks	Fine	Thunder in aft. all round.	strong wind and squall clouds from NW	Cloudy Thunder in aft. with light rain - strong E wind all night with great fall of temperature	Winds W. 11 NW 14 E. 6 Mean temp 73.3	Cloudy light E wind and very cold	Cloudy	Cloudy like rain	Fine Lake round moon very large and halo

Australian Government
Bureau of Meteorology

Basic Climatological Station Metadata
Current status

Station:	EVERSLEIGH			**Location:**	EVERSLEIGH			**State:**	NSW
Bureau No.:	056056	**WMO No.:**	?	**Aviation ID:**	NO ID	**Opened:**	01 Jan 1877	**Current Status:**	Closed
Latitude:	-30.5000	**Longitude:**	151.5000	**Elevation:**		**Barometer Elev:**		**Metadata compiled:**	27 JUL 2021

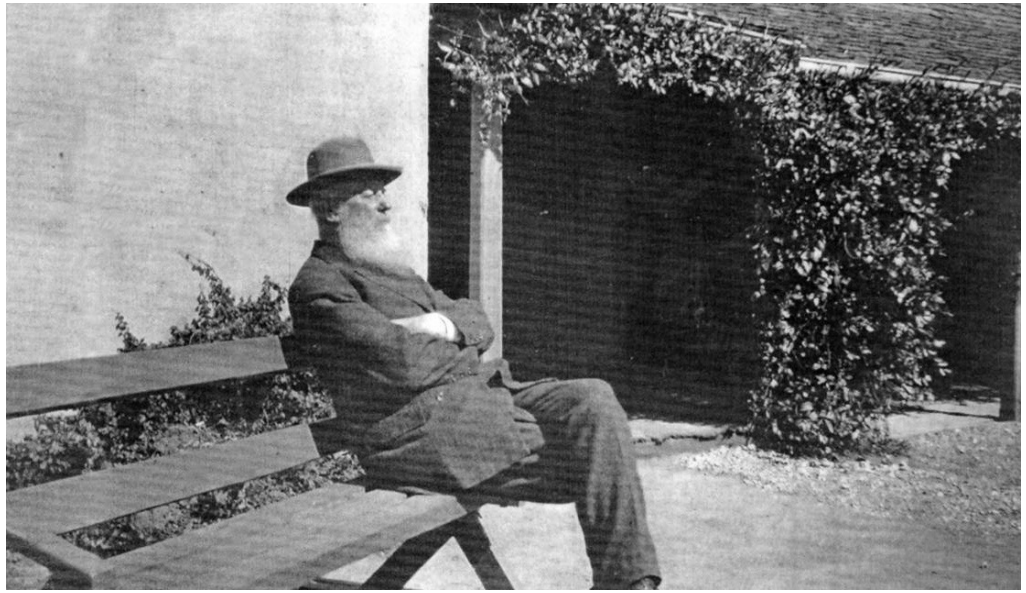
Observation summary

The table below indicates the approximate completeness of the record for individual element types within the Australian Data Archive for Meteorology. For elements not listed see the note below.

Completeness 0% → 10

DAILY DATA HOLDINGS

RVATION TYPE	FIRST MONTH	LAST MONTH	COMPLETENESS (% estimate)	SINGLE DAYS MISSED	FULL MONTHS MISSED
FALL	JUN 1877	JUL 1922	96	N/A	



Images: Belfield family, University of Newcastle, Bureau of Meteorology

TAKING THE SHADE TEMPERATURE.

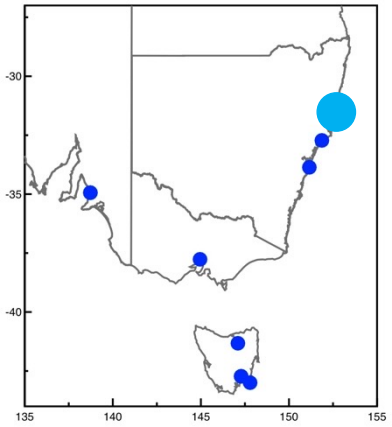
How can we trust sparse rescued data?

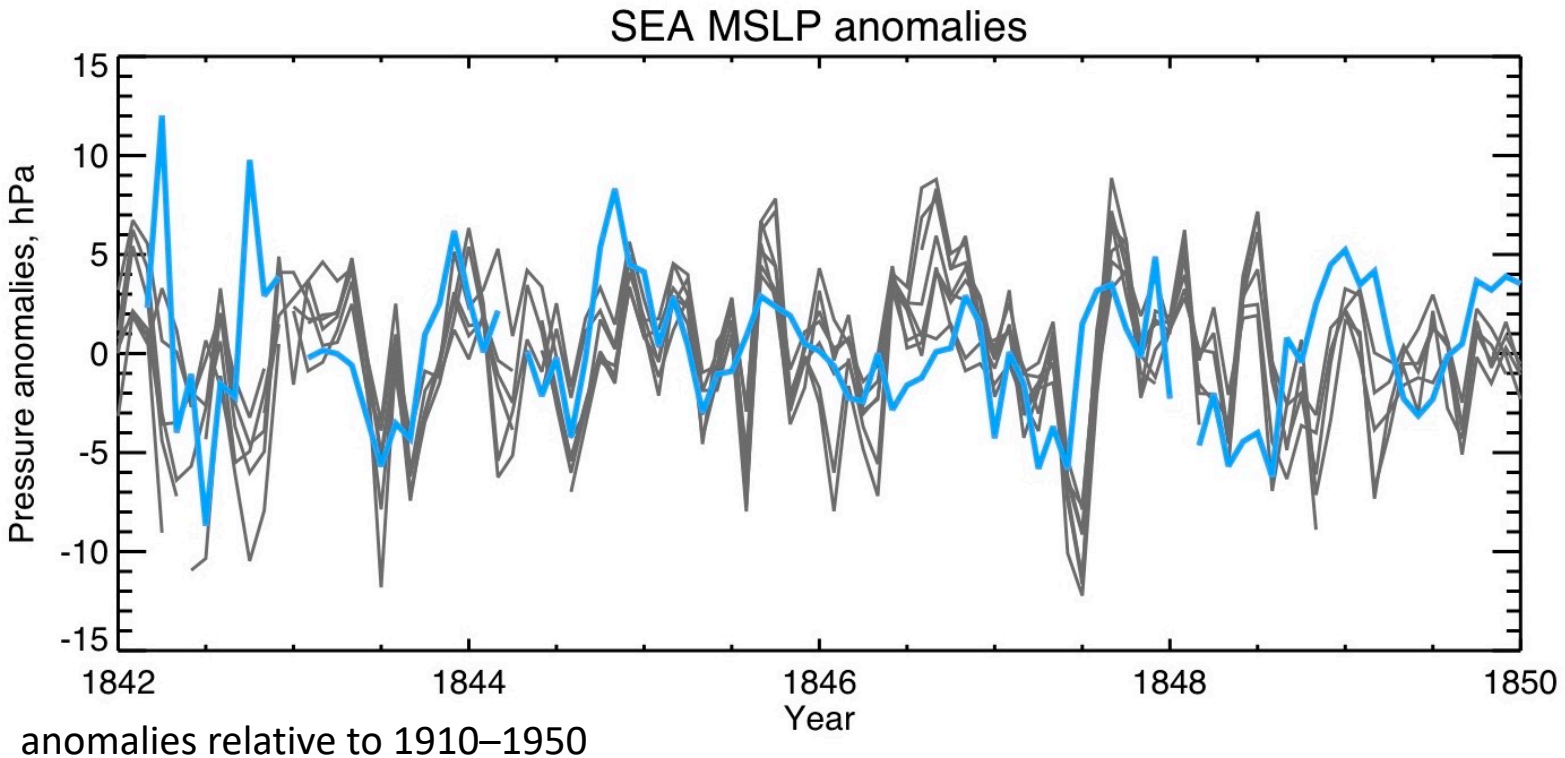
1. Rescuing as much as we can

- Often hard to do with budgets, project foci.
- But a full picture enables a better understanding of datasets
- Also ensures more complete quality assessment
- Citizen science is helpful here – one person's boring information is another person's fascinating story!
- (Although don't underestimate the time and effort it takes to set up a citizen science project)

Image: Belfield family

2. Using neighbours if they exist (even if they're far away)





anomalies relative to 1910–1950

3. Comparing with qualitative data

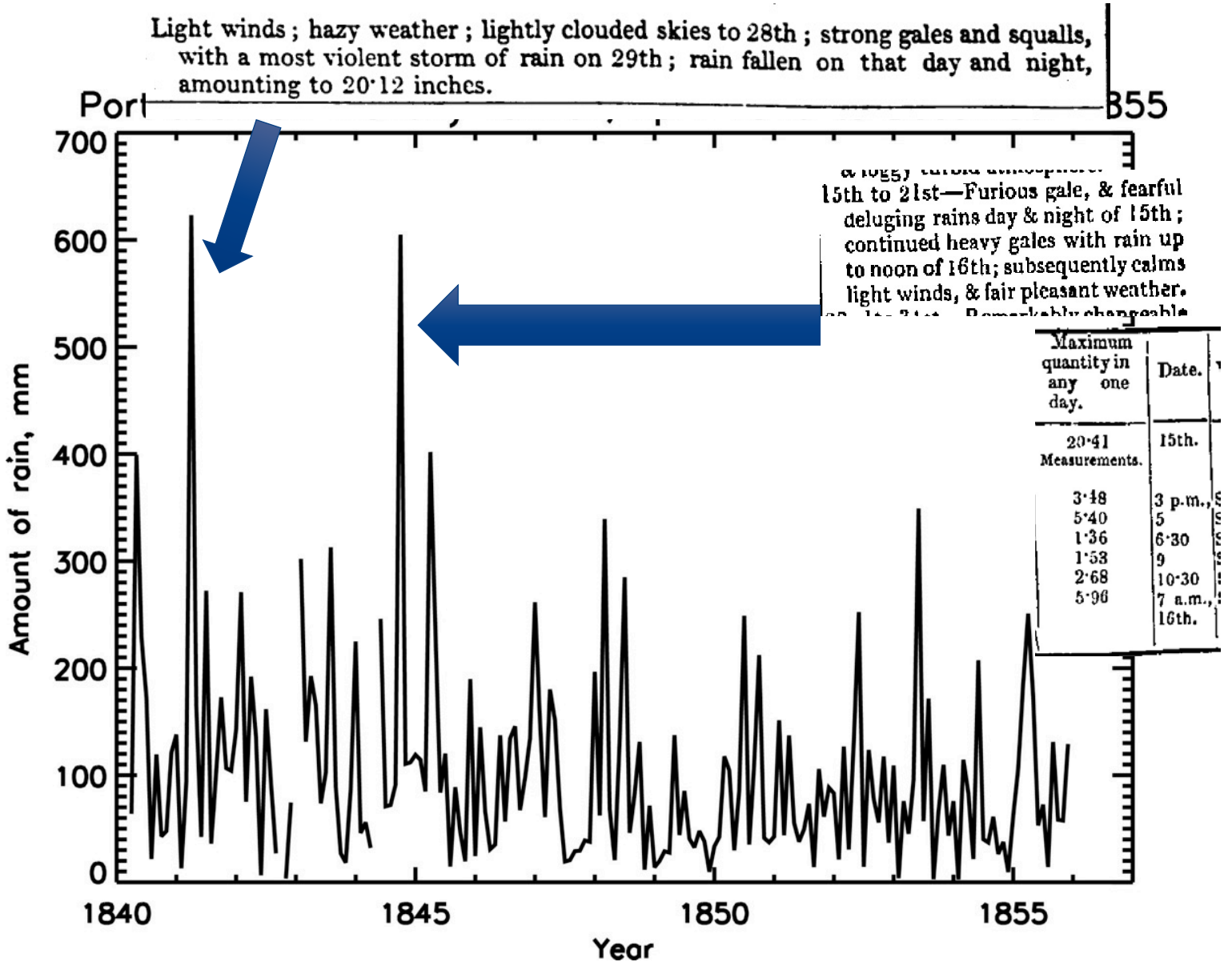


Image: SA Zoo

3. Comparing with qualitative data

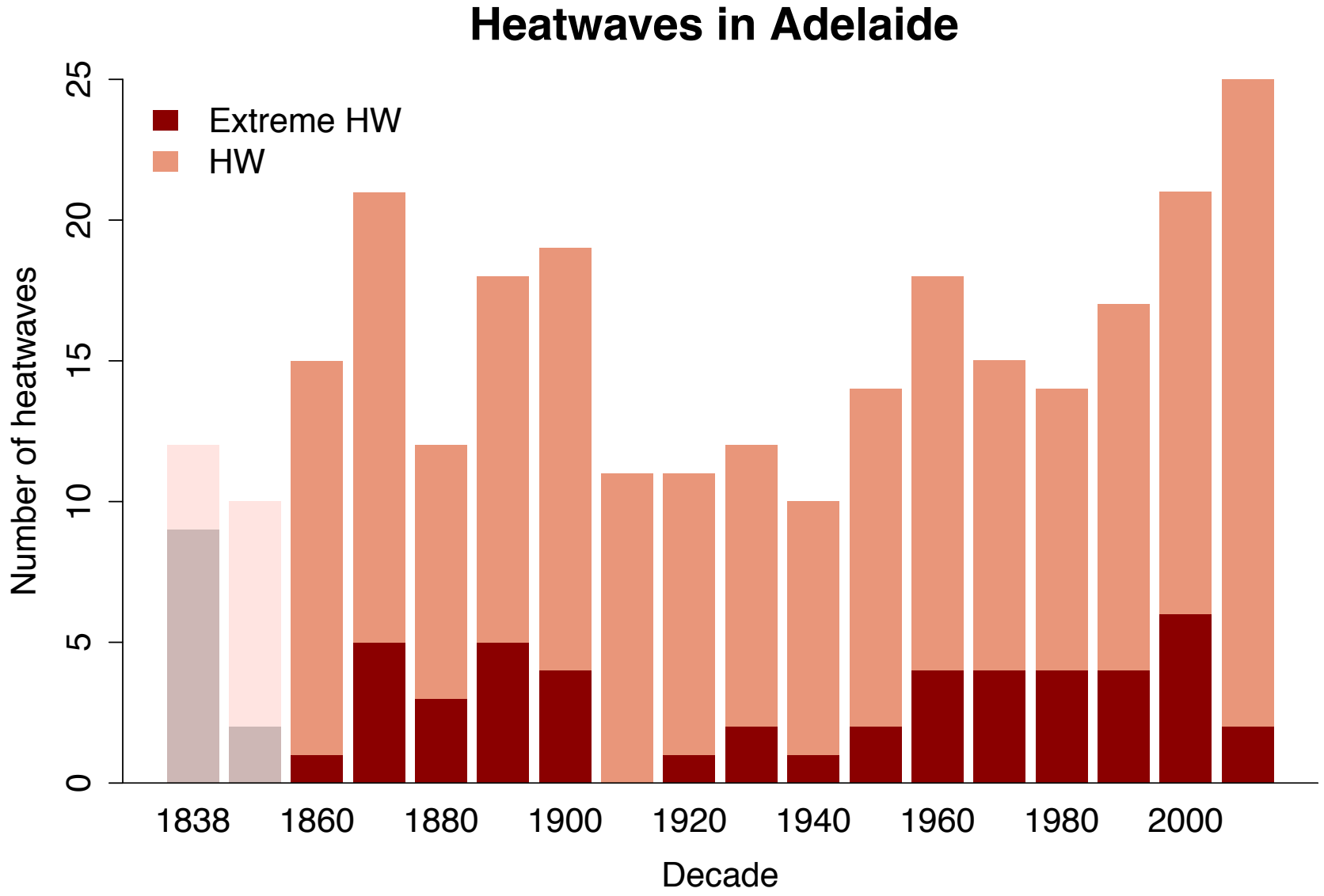
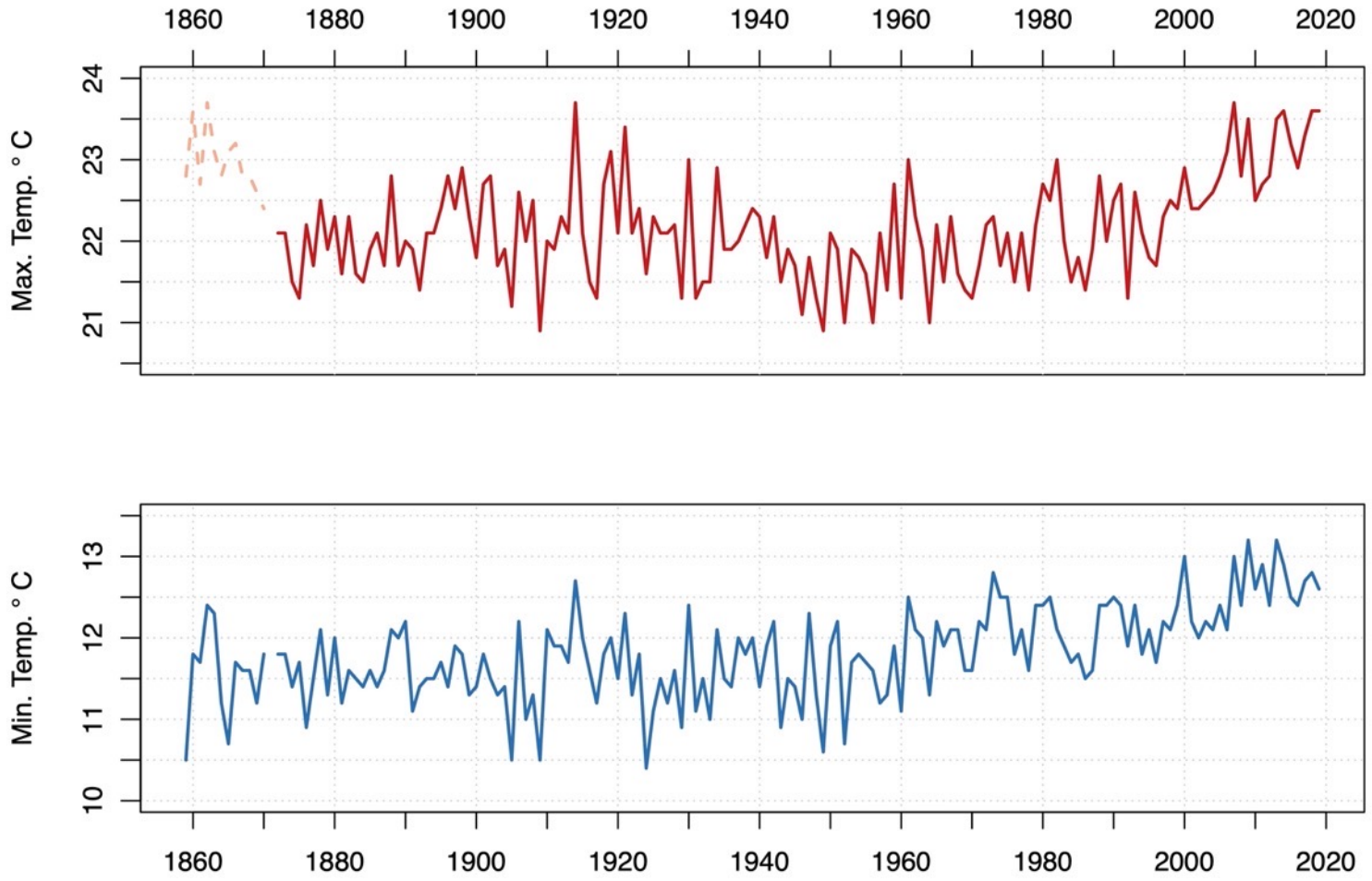


Image: Gergis et al. 2020

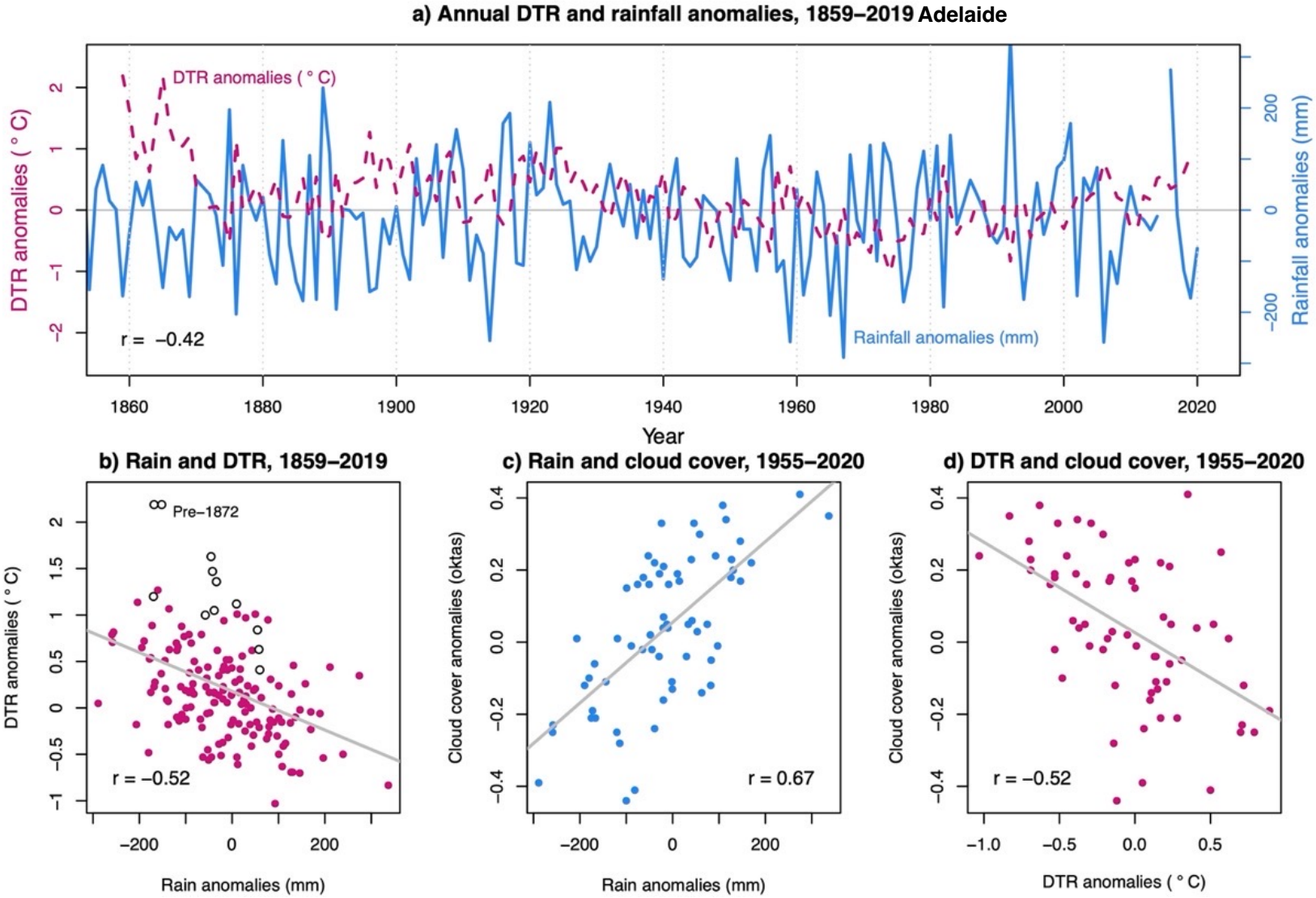
4. Using intervariable relationships



Annual temperatures, Adelaide, 1859–2020

Images: Ashcroft et al. 2021, State Library of South Australia

4. Using intervariable relationships



Images: Ashcroft et al. 2021, State Library of South Australia

5. Comparing with palaeoclimate reconstructions

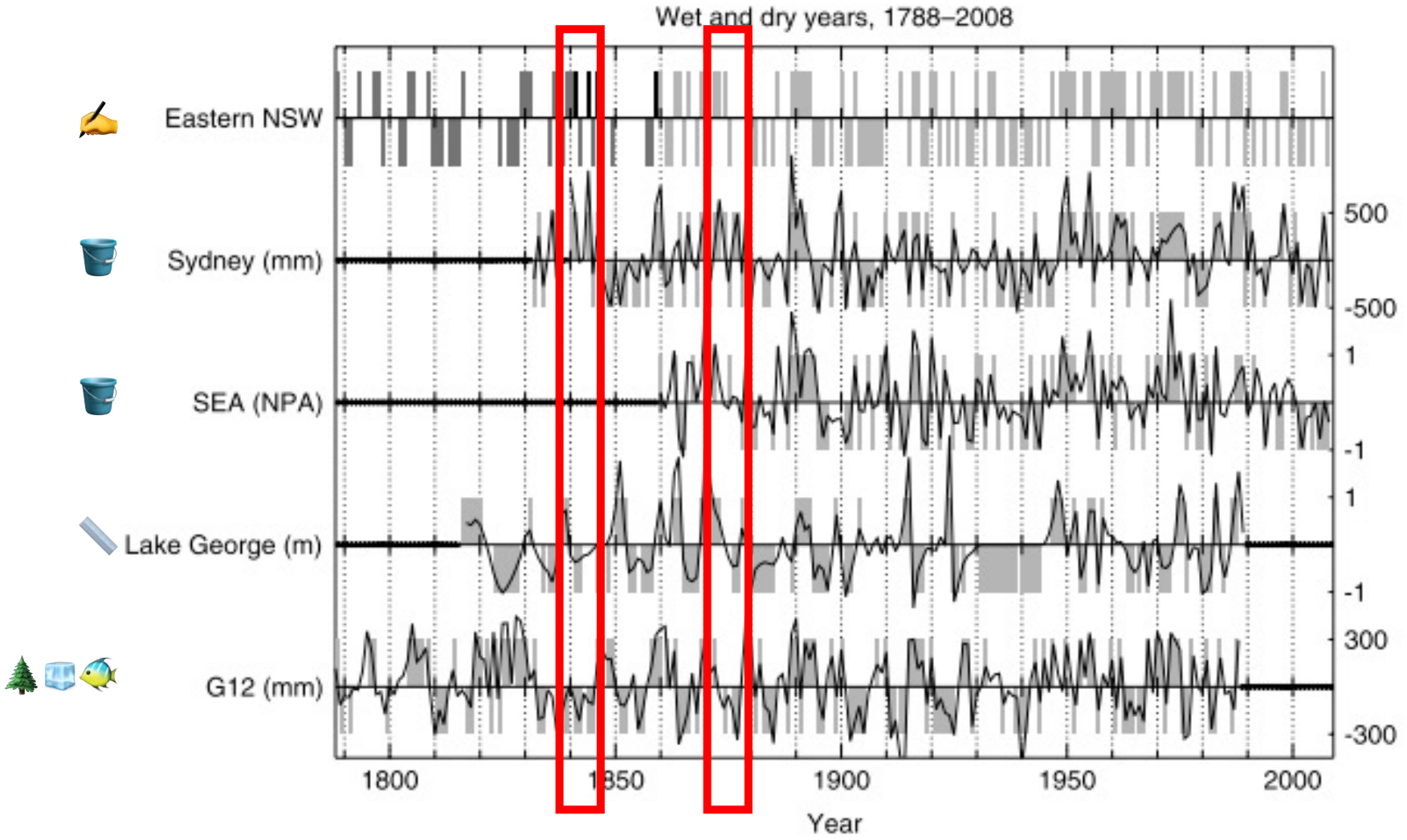
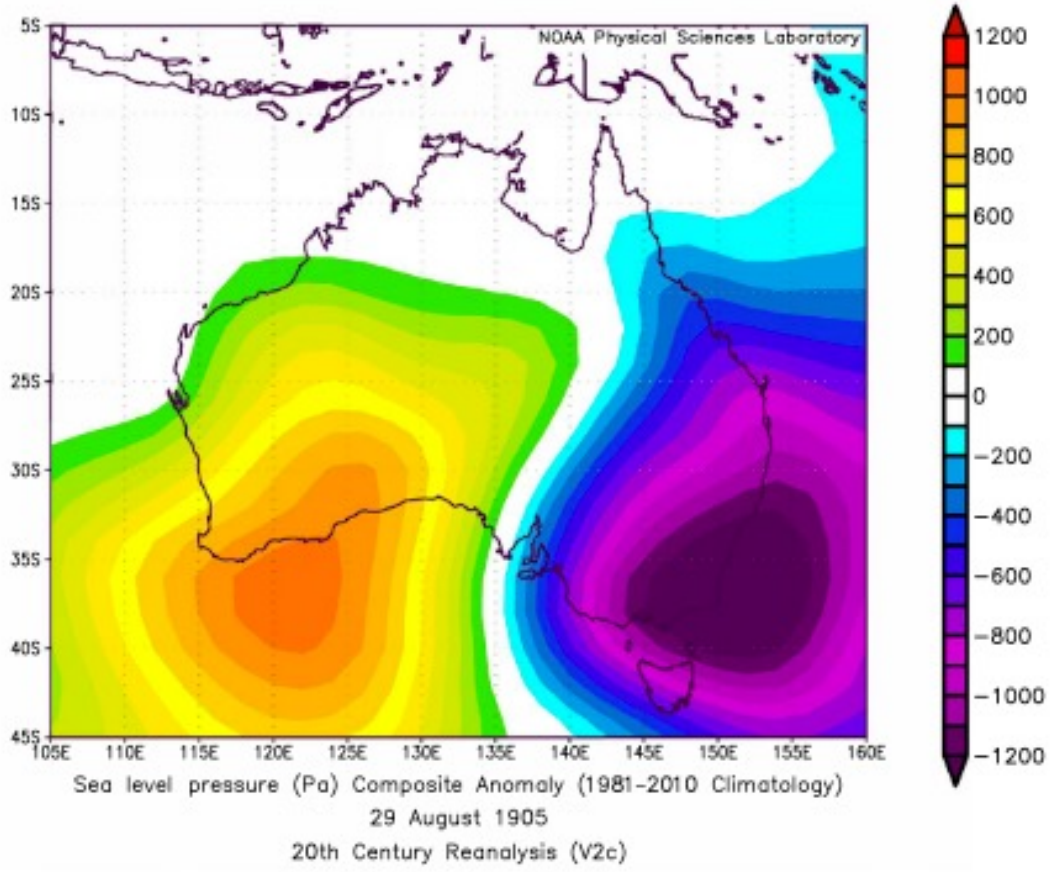


Image: Gergis and Ashcroft, 2012

5. Using reanalysis

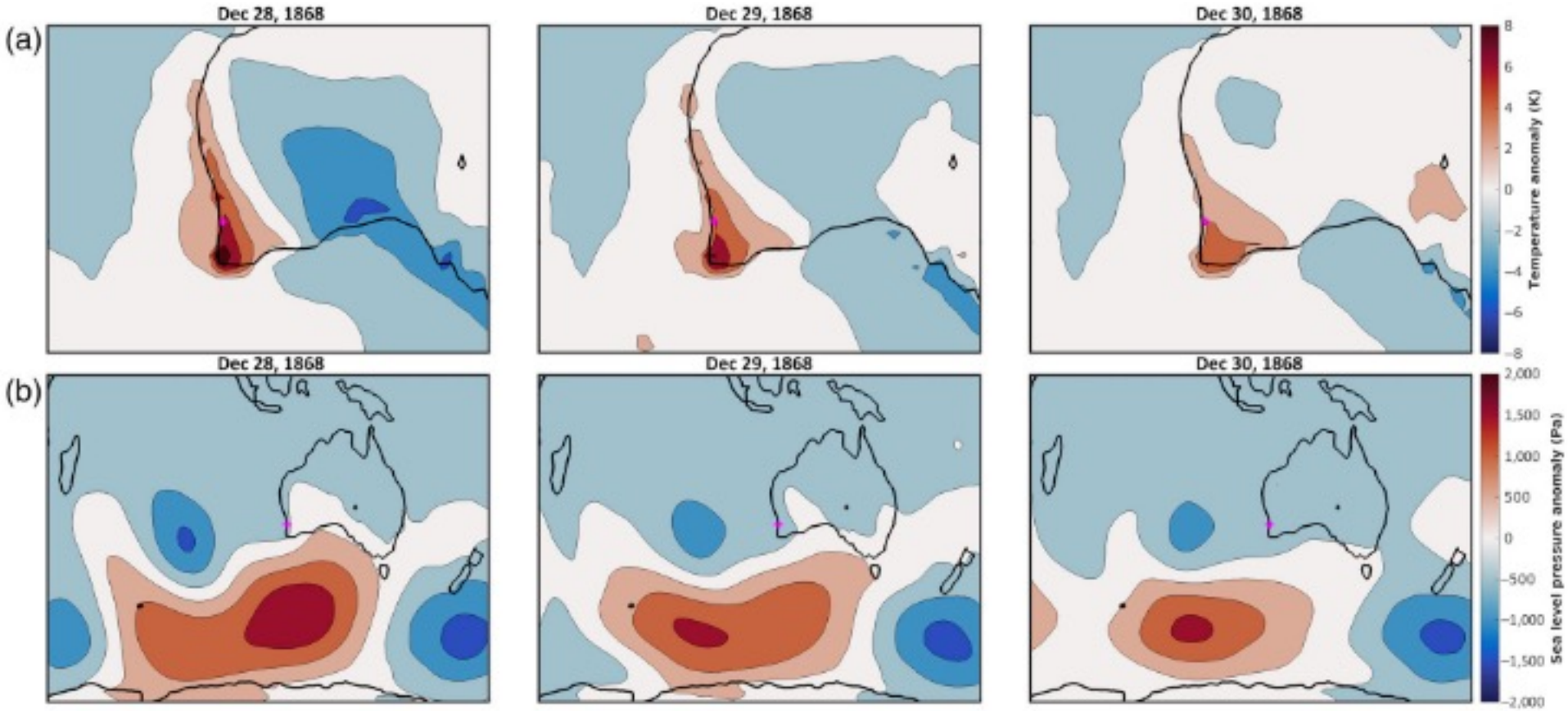
Reports of snowfall around Adelaide, 29 August 1905. Red pins = settled snow



Ensemble mean MSLP anomaly from 28 August 1905

Image: Gergis et al. 2020

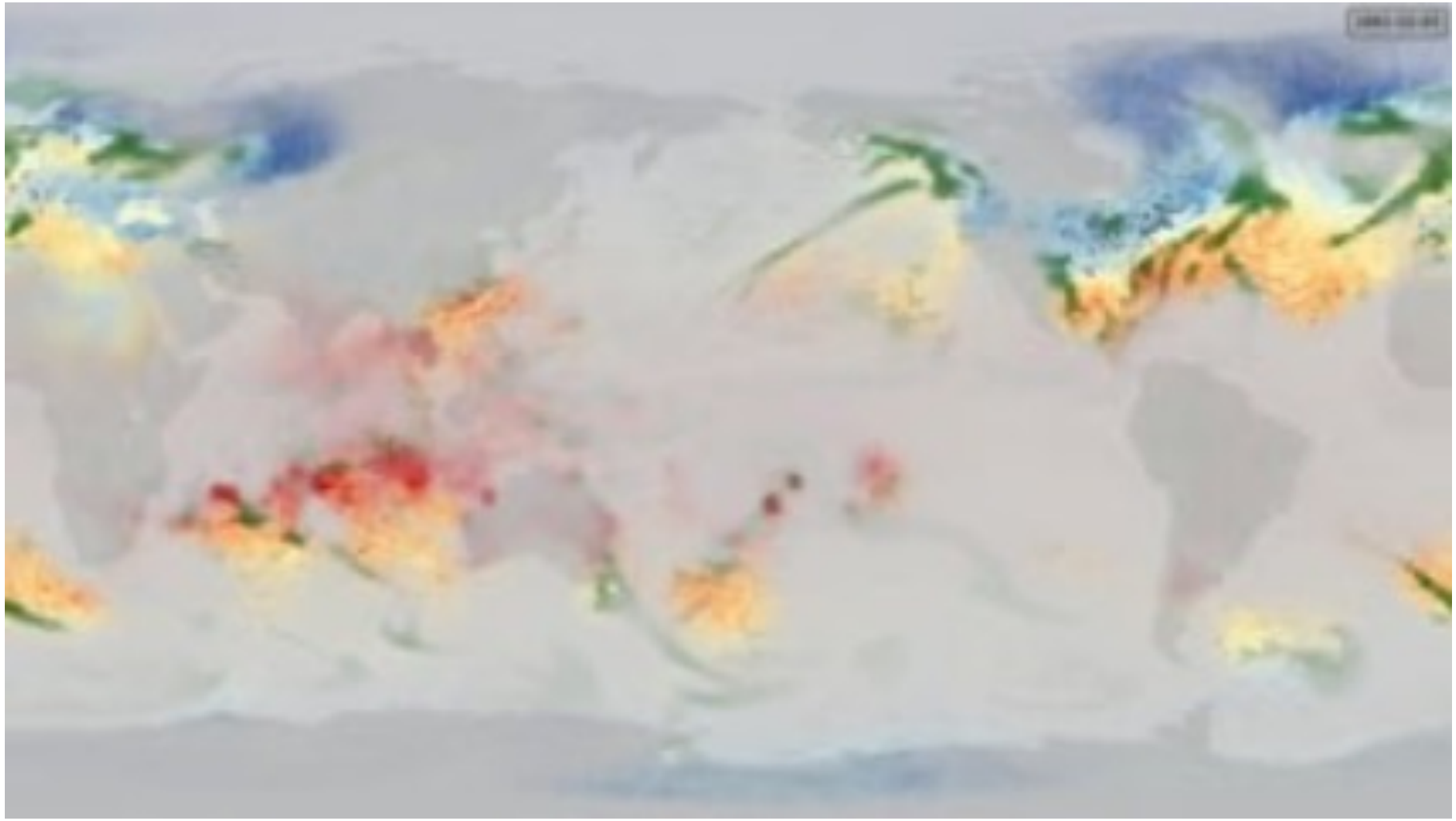
5. Using reanalysis



2m temperature anomalies and associated MSLP anomalies in 20CRV3 ensemble mean during an identified heatwave, 28–30 December 1868.

Image: Gergis et al. 2021

Using reanalysis?



Source: Philip Brohan, https://vimeo.com/455072133

Questions

linden.ashcroft@unimelb.edu.au

What do you see are the most significant advances for the field of reanalysis in 5-10 years?

- *recovery of 'new' data*
- *increased inclusion of ensembles, with associated uncertainty estimates*

What do you see are the most significant barriers to progress in the field of reanalysis?

Which collaborations are currently working and which collaborations need to be fostered?

- *increased international coordination of data rescue activities will lead to increased awareness of recovered datasets*
- *still some way to go in connecting data rescue activities to some databases*

What are the critical requirements for consistent Earth system reanalysis?

What observational datasets are required to support these requirements?

- more old ones!

What modeling components are mature enough to enable reanalysis for your specific science question or application?

How is uncertainty quantified for your application? Are there significant barriers for quantifying uncertainty in your field?