

Sparse data rescue

Future US Earth System Reanalysis Pre-Workshop Webinar Series

8 March 2022

Image: State Library of QLD



Sparse data rescue

Future US Earth System Reanalysis Pre-Workshop Webinar Series

8 March 2022

Image: Darren England/AAP



- 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

$\begin{array}{c c c c c c c c c c c c c c c c c c c $	Emerican Warter. Here bet Oldy Orace I Sty Mady Raan
Latitude South. Altitude above Sea Level Peet. Longitude East Name	Emerica Wester. The bit Oldy Coace I Sty Mosty Ram
$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	Emarks on Weather. the bit addy Chace I by Mady Rain
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	Emarks on Weather. This bit Oldy Orace I Ily Midy Rain
$ \begin{array}{c} 1 & 50 \ 0.28 \ 70 & 29 \ 916 \ 29 \ 976 \ 72 \ 976 \ 918 \ 74 \ 57 \ 65 \ 74 \ 57 \ 65 \ 76 \ 78 \ 78 \ 78 \ 78 \ 78 \ 78 \ 78$	the ht ddy Orace I sty Mady Rain
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	Mesty Rain
$ \begin{array}{c} s & so 150 & bq 20 + bq 71 & 30 + 25 & 30 + qg \\ 6 & 37 + q4 & 51 & 57 & 61 & 57 & 77 & 73 + 4 & 57 & 67 & 57 & 61 & 57 & 57 & 77 & 73 + 4 & 57 & 67 & 57 & 67 & 57 & 77 & 73 + 4 & 57 & 67 & 57 & 67 & 57 & 77 & 73 + 4 & 46 & 46 & 46 & 47 & 47 & 48 & 48 & 48 & 46 & 47 & 47 & 48 & 48 & 48 & 46 & 47 & 47 & 48 & 48 & 48 & 48 & 48 & 49 & 48 & 48$	the all the
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	Ownerd
$\begin{array}{c ccccccccccccccccccccccccccccccccccc$	this he ledy
$\begin{array}{ c c c c c c c c c c c c c c c c c c c$	2
$\begin{array}{c} 9 & 36 \cdot 324 & 42 \cdot 36 \cdot 44 & 45 \cdot 64 & 7 \cdot 51 & 66 & 7 \cdot 51 & 67 \cdot 52 & 67 \cdot 51 & 67 \cdot 51 & 67 \cdot 51 & 67 \cdot 52 & 67 \cdot 51 & 67 \cdot 5$	An 1 apt
$ \begin{array}{ c c c c c c c c c c c c c c c c c c c$	the & Char
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	· · · · dar
13 50510 74 30 227 30 226 81 30 134 83 61 12 72 60 69 133 56 69 69 64 66 65 D Stor 26 3 0 0 14 20 154 78 30 171 30 110 88 29 999 29 999 97 65 \$1 97 76 64 58 145 143 54 72 71 68 65 64 62 80 32 2 0 0	
12 80134 18 30171 3110 88 24 499 33999 47 05 11 41 76 64 58 143 143 54 72 71 68 63 64 Get 11 32 2 0 8	
som and start an established the mast in second and the mark and the marked	
Mess 30228 67 30.28 78 30.114 76 56 66 62 62 68 132 46 66 67 67 65 66 144 4 10 15	
15 30100 81 30 008 30 000 87 29 972 96 63 79 96 76 60 53 145 54 70 73 69 66 65 8 M 30 2 0 0	This dear
10 50 138 19 50 138 30 201 48 71 50 16 93 15 19 76 62 56 142 58 78 72 70 66 65 50 40 - 29 2 0 90 17 50 271 72 30 218 30 201 75 30 148 30218 79 33 167 58 75 52 78 132 18 48 71 71 70 66 65 8 310 - 20 5 0 0	
18 30-204 72 30-136 30-014 85 29 gel 77 55 66 53 67 53 68 130 48 48 70 71 70 66 65 8 16 24 5 0 0	
10 30.018 76 29.95 29.967 87 29.87 1 68 79 78 64 63 74 96 77 62 56 75 75 70 66 65 62 10 30 2 50 50 20 20 20 20 20 20 20 20 20 20 20 20 20	
21 29.970 78 29.187 30 534 76 29.900 93 63 78 30 61 65 145 59 74 74 71 66 65 N Met 28 4 75 100	Sully
Ser 216-626540260 653211-455372209807 (25632.627) 5491114457 996 381 507 507 490 462.655 187 22 125326 Marine 20 689 77 50.008 20 460 81 29 992 89 62 75 78 59 62 142 54 72 72 70 66 65 3 18 45	
22 30.156 73 30.086 30.152 76 30.055 30.086 82 57 69 69 53 60 135 53 72 72 71 \$0.66 W 1 25 6 35 75	Salley
23 30 07 0 71 30.005 29 97 078 29.087 76 36 66 66 56 63 53 60 131 49 49 71 72 71 68 66 R SE 25 3 0 75	Fine & Clean
25 29928 71 29 908 29 990 75 39 114 29 14 88 64 76 60 10 62 55 142 142 61 71 72 71 68 66 8 173 22 6 100 105 25 29928 71 29 928 29 996 72 29 928 79 57 68 63 56 56 122 514 69 70 70 68 66 111 124 13 6 75 100	Cherces the dia
28 30.104 70.30 0 4 30.884 70 30.022 75 62 67 66 61 78 135 61 69 70 70 64 66 Mr 11 10 6 100 10	Concert flooring
28 30 0 FF 71 30 023 30 060 78 39 977 76 60 68 69 58 75 124 57 69 69 66 62 12 110 13 14 10 10	this but cloby
News R10442 580299 979 210 210 22329 30 547 416480 165 400 453 92 38 8 400 483 492476 462 48 126 35 48 62	
Mars 30.063 71 29497 30.029 74.29 964 18 59 68 66 55 65 132 55 40 40 70 68 66 5 69 86	
20 55018 74 29 11 20 12 50 19 70 29 71 29 15 12 59 70 59 71 58 69 141 141 53 53 71 71 71 68 66 19 510 20 14 50 100 30 20,790 11 20,725 70 12, 70 20,720 20 725 10 13 72 10 73 10 70 10 10 10 10 10 10 10 10 10 10 10 10 10	the Ring
31 111 1 1 1 20 3 10 10 10 10 10 10 10 10 10 10 10 10 10	Ide a dig la wij
Sem \$9.808 14 39 10 0.9 194 150 59 440 165 122.143 143 120 169 276 116 142-142 141 156 152 101 141 9 150 200	
13 100 Harris 19 40 19 19 19 19 19 19 19 19 19 19 19 19 19	
Mars 30 11 12 30 11 30 11 30 10 20 30 2975 80 589 689 970 460 67 56 67 133 162 525 39 69 69 68 66 66 68 4 48 39	
Bornander S. B. Contraction for the contraction of	
by	Neit
Itolax Ever Nax. W.,	Lora
Temperatures (as per table). Non. W	Observer.
12 2000 Do for pervision 23-41 Do design pervision 101 Consultant days. 9 II-12, 15-16-15-17 (17. 8	
Tarthe Brance Mational Archiver	of Austral



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















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



1860: Total # stations = 118









Comparison of observations gained and lost between 4.0.0.0 and 5.0.0.0

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



Sparse - variable



Basic Climatological Station Metadata Current status

Station:	н		Location:	EVERSL	EIGH	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

		DAILY DAT	A HOLDINGS		0%			
RVATION TYPE	FIRST MONTH	LAST MONTH	COMPLETEN (% estimate)	IESS	SINGLE DAYS MISSED		FULL MONTHS MISSED	
FALL	JUN 1877	JUL 1922	6	96	í.	N/A		
	1 9 0 0		1 9 5 0				2 0 0 0	



January	A graph	1 M	Y	S/2	The last	Februa	y.	3	1, 1	
Date	28	29	30	31		1-	2		9 a.m.	
Time	9 a.m.	9 a.m		51	9 a m.	9 a.m.	9 a.m.	9 a.m.	Reading.	
INSTRUMENT.	Reading.	Reading.	Bondin	9 a.m.	Reading.	Reading.	Reading.	Reading.		
Attacked Thermonytee				Reading		FF	57	59	64	
Renameter	23	76	73	69	January.	.418	.639	.612	.574	-
Dry Bulb	-540	1588	, 516	.455	Quari q	57.2	63	59.3	66.3	- 1
Wet Bulb	10.	19:1	7,4	67.8	John 16	.50.7	56	55.8	60	
Maximum	02.	00	aug	837	P. A.u. 27.5	82.8	71.1	70.7	70	2-
Minimum	160.	63	65.3	65.6	zaye	47.9	44	52.2	22.2	
Rain Weight			-	" Gital	duir. 7. 38.1	corr	-	0	0	
Measure	3	0	0	0 9h	ax. 87.0	Gm 11	0			
Evaporation				971	in. 59.5	qu				V
CAmount	1.1	6/1/	MIL	A	14 73.2	8	3	3	NE	
Direction of Wind	W	NW	NW	IV W 4	a 05.1	C				
Velocity	,	2	3	2 #	um. 54	4	3	2	1	122
Cloud	-	1	2	10 F	ain 1.31 i	. 10	8	10	5	D.E.
Cioua	Fine.	Thunder	strong wind	Cloudy Thursday in	Winds W. 11	Cloudy high 2	Cloudy	Clondy like zais	Fine	123
-		in aft. all	cloud from	alt. with high	E.6	hind and		une aur	more by	25
			nw	Emid all	100				ling in	12.9
		1		a night with	Theen temp				the have	2.5
				Just fall of	73.3					196
				emphalmet						1.4
Remarks										1
										100.0
										10.00
										2.17
										a the
				+						1
										100 March
				-						and the second se



How can we trust sparse rescued data?

TAKING THE SHADE TEMPERATURE.



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)





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





3. Comparing with qualitative data

MELBOURNE







3. Comparing with qualitative data

Heatwaves in Adelaide





4. Using intervariable relationships





Annual temperatures, Adelaide, 1859–2020

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

4. Using intervariable relationships



<image><image>

5. Comparing with palaeoclimate reconstructions

THE UNIVERSITY OF MELBOURNE





5. Using reanalysis





Reports of snowfall around Adelaide, 29 August 1905. Red pins = settled snow Ensemble mean MSLP anomaly from 28 August 1905



5. Using reanalysis



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



Using reanalysis?





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 ₂₃ in your field?