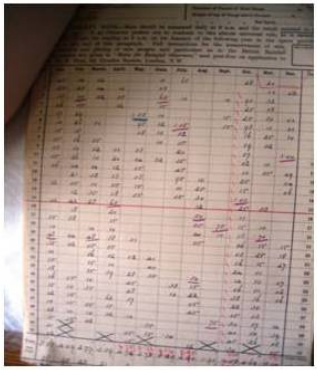
Data rescue in the classroom: research-led teaching to extend historical records

Ciara Ryan[1], Catriona Duffy[1], Ciaran Broderick[1], Peter W. Thorne[1], Mary Curley[2], Séamus Walsh[2], Conor Daly[2], Mairéad Treanor[3,4], Conor Murphy[1]



[1]Irish Climate Analysis and Research Units, Department of Geography, Maynooth University, Maynooth, Co. Kildare, Ireland.

[2]Climatology and Observations Division, Met Éireann, Dublin, Ireland.

[3]Library, Met Éireann, Dublin, Ireland.

[4] The Oireachtas Library, Houses of the Oireachtas.

Correspondence to: Ciara Ryan (ciara.ryan@mu.ie)

ICARUS Irish Climate Analysis and Research Units

Maynooth University National University of Ireland Maynooth

MET éireann

Project Overview

- Work explores:

(i) the potential for integrating data rescue activities into the classroom;

(ii) the ability of students to produce reliable transcriptions and;

(iii) the learning benefits for students.

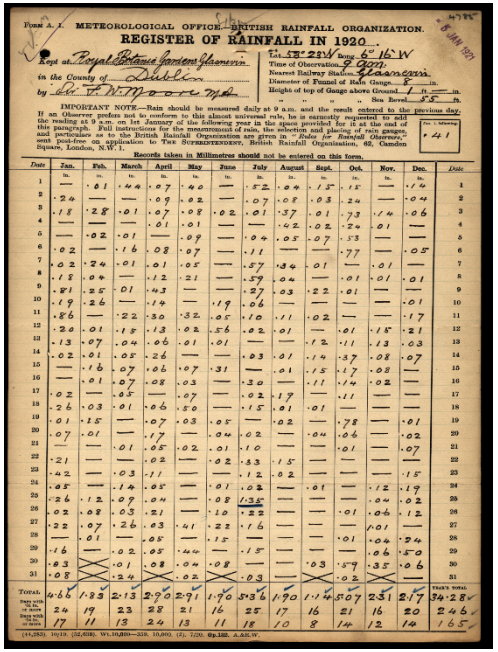
- Transcribe ~1300 annual rainfall sheets relating to 45 stations across Ireland.

- Each student (142 students) given 18 sheets - double keyed for QA.

- Excel template provided by Met Éireann.

- 274 sheets previously transcribed (single keyed) by Met Éireann.

- 3 Steps:
  (1) Receiving the data (Dropbox)
  (2) Transcribing the data (Excel)
  (3) Returning the data (Moodle)


FORM A. I. METEOROLOGICAL OFFICE—BRITISH RAINFALL ORGANIZATION.
REGISTER OF RAINFALL IN 1920.

Kept at Royal Botanic Gardens Glasnevin
in the County of Dublin
by Dr F. W. Moore M.A.

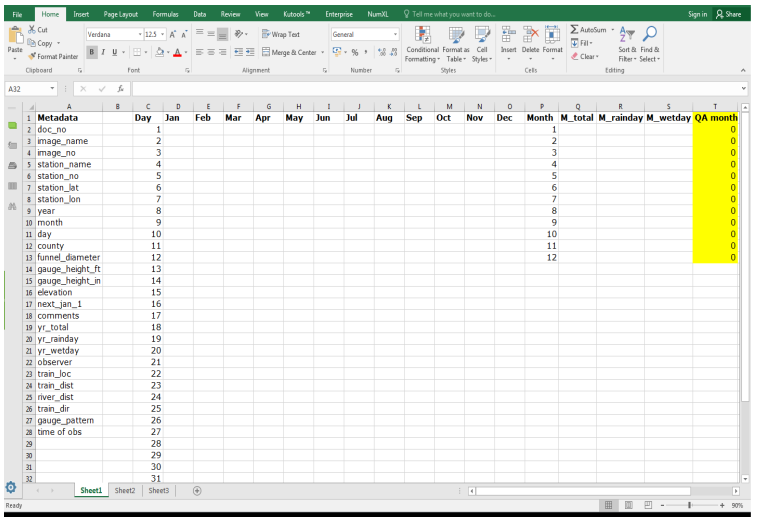
Lat. 53° 23' N Long. 6° 16' W
Time of Observation 9 a.m.
Nearest Railway Station Glasnevin
Diameter of Funnel of Rain Gauge 8 in.
Height of top of Gauge above Ground 1 ft. in.
" " " " Sea Level 55 ft.

IMPORTANT NOTE.—Rain should be measured daily at 9 a.m. and the result entered to the previous day. If an Observer prefers not to conform to this almost universal rule, he is earnestly requested to add the reading at 9 a.m. on 1st January of the following year in the space provided for it at the end of this paragraph. Full instructions for the measurement of rain, the selection and placing of rain gauges, and particulars as to the British Rainfall Organization are given in "Rules for Rainfall Observers," sent post-free on application to THE SUPERINTENDENT, British Rainfall Organization, 62, Camden Square, London, N.W. 1.

Records taken in Millimetres should not be entered on this form.

Date	Jan.	Feb.	March	April	May	June	July	August	Sept.	Oct.	Nov.	Dec.	Date
	in.	in.	in.	in.	in.	in.	in.	in.	in.	in.	in.	in.	
1	—	.01	.44	.07	.40	—	.52	.04	.15	.15	—	.14	1
2	.24	—	—	.09	.02	—	.07	.08	.03	.24	—	.04	2
3	.18	.28	.01	.07	.08	.02	.01	.37	.01	.73	.14	.06	3
4	—	—	—	.01	.01	—	—	.42	.02	.24	.01	—	4
5	—	.02	.01	—	.09	—	.04	.05	.07	.53	—	—	5
6	.02	—	.16	.08	.07	—	.11	—	—	.77	—	.05	6
7	.02	.24	.01	.01	.05	—	.57	.34	.01	—	.01	—	7
8	.18	.04	—	.12	.21	—	.59	.04	—	.01	.01	.01	8
9	.81	.25	.01	.43	—	—	.27	.03	.22	.01	—	—	9
10	.19	.26	—	.14	—	.19	.06	—	—	—	—	.01	10
11	.86	—	.22	.30	.32	.05	.10	.11	.02	—	—	.17	11
12	.20	.01	.15	.13	.02	.56	.02	.01	—	.01	.15	.21	12
13	.13	.07	.04	.06	.01	.01	—	—	.12	.11	.13	.03	13
14	.02	.01	.05	.26	—	—	.03	.01	.14	.37	.08	.07	14
15	—	.16	.07	.06	.07	.31	—	.01	.15	.17	.08	—	15
16	—	.01	.07	.08	.03	—	.30	—	.11	.14	.02	—	16
17	.02	—	.05	—	.07	—	.02	.17	—	.11	—	—	17
18	.26	.03	.01	.06	.50	—	.15	.01	.01	—	—	—	18
19	.01	.15	—	.07	.03	.05	—	.02	—	.78	—	.01	19
20	.07	.01	—	.17	—	.04	.02	—	.04	.06	—	.02	20
21	—	—	.01	.05	.02	.01	.10	—	—	.01	—	.07	21
22	.21	—	—	.02	—	.02	.33	.15	—	—	—	—	22
23	.42	—	.03	.11	—	—	.12	.02	—	—	—	.15	23
24	.05	—	.14	.05	—	.01	.02	—	.01	—	.12	.19	24
25	.26	.12	.09	.04	—	.08	1.35	—	—	—	.04	.02	25
26	.02	.08	.03	.21	—	.10	.22	—	—	.01	.06	.12	26
27	.22	.07	.26	.03	.41	.22	.16	—	—	—	.01	—	27
28	—	.01	—	.05	—	.15	—	—	—	.01	.04	.24	28
29	.16	—	.02	.05	.44	—	.15	—	—	—	.06	.50	29
30	.03	X	.01	.08	.04	.08	—	—	.03	.59	.35	.06	30
31	.08	X	.24	X	.02	X	.03	—	X	.02	X	—	31
TOTAL	4.66	1.83	2.13	2.90	2.91	1.90	5.36	1.90	1.14	5.07	2.31	2.17	TRAN'S TOTAL 34.28
Days with 10 hs. or more	24	19	23	28	21	16	25	17	16	21	16	20	246
Days with 10 hs. or more	17	11	13	24	13	11	18	10	8	14	12	14	165

(44,283). 10/19. (52,638). Wt. 10,000—318. 10,000. (2). 7/20. Gp. 132. A.&E.W.



Project Overview – student supports

- Motivating Students –section of module devoted to the importance of historical climatology.
- Met Éireann gave a talk to convey the scientific, cultural and social importance of the data.
- Step by step written instructions
- Video tutorial
- QA check
- Check-in clinic - FAQ
- Class Discussion Forum

Re: Continuous Assessment Questions and Answers
by SEAN O KANE - Monday, 7 November 2016, 9:16 PM

Hi Ciara,

I'm having trouble making out the observer's name on Image number 6281.

Thanks in advance.

Sean.

Show parent | Edit | Split | Delete | Reply

Re: Continuous Assessment Questions and Answers
by CIARA RYAN - Tuesday, 8 November 2016, 10:00 AM

Hi Sean

The observer was Thomas Jackson.

Ciara

Show parent | Edit | Split | Delete | Reply

Re: Continuous Assessment Questions and Answers
by SUZANNE O CALLAGHAN - Wednesday, 2 November 2016, 7:55 PM

Hi Ciara

In relation to form 5 image 6055 for the monthly total in december I am getting 5.85 but 5.75 was recorded. If you wouldnt mind taking a look at it.

Thank you,

Suzanne.

Show parent | Edit | Split | Delete | Reply

Re: Continuous Assessment Questions and Answers
by CIARA RYAN - Wednesday, 2 November 2016, 11:14 PM

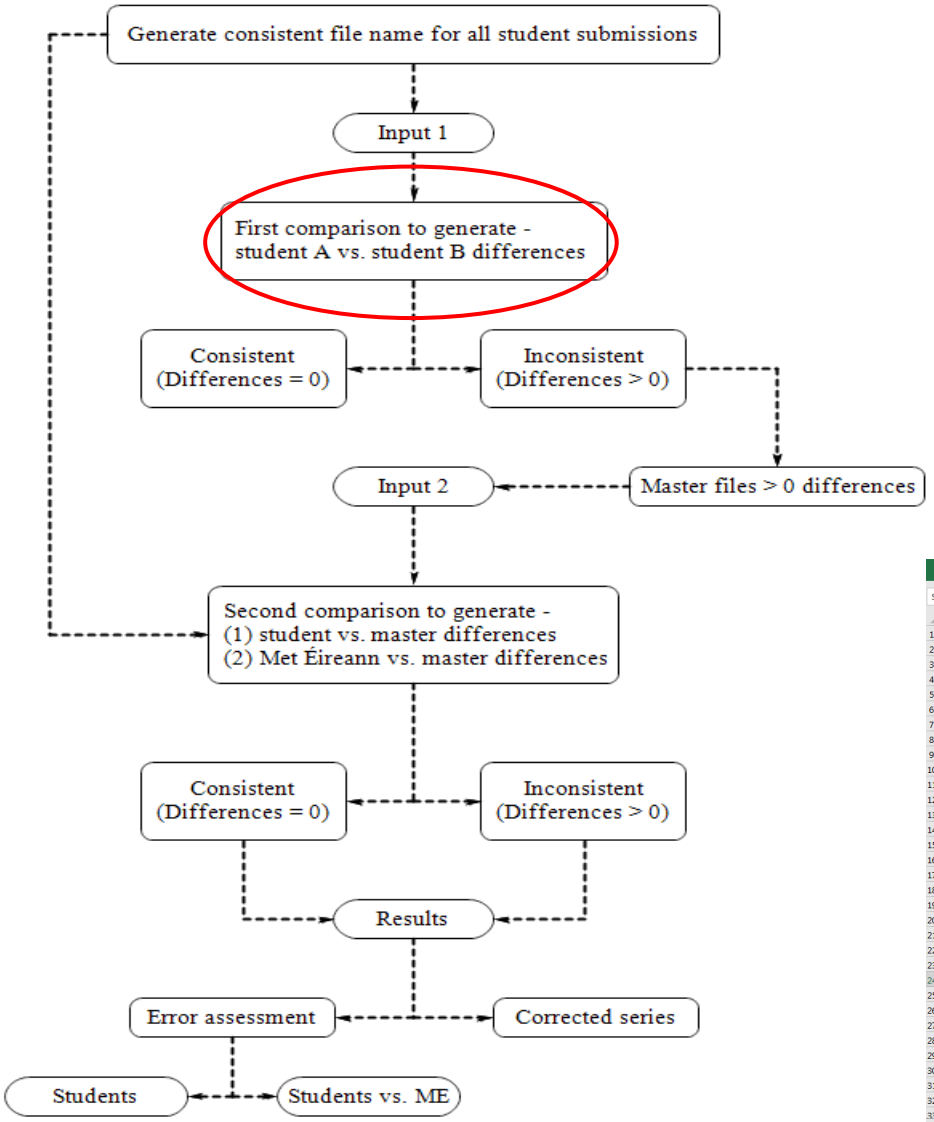
Hi Suzanne,

I'm getting 5.85 too so you can use this as your total for dec and adjust the yearly total accordingly.

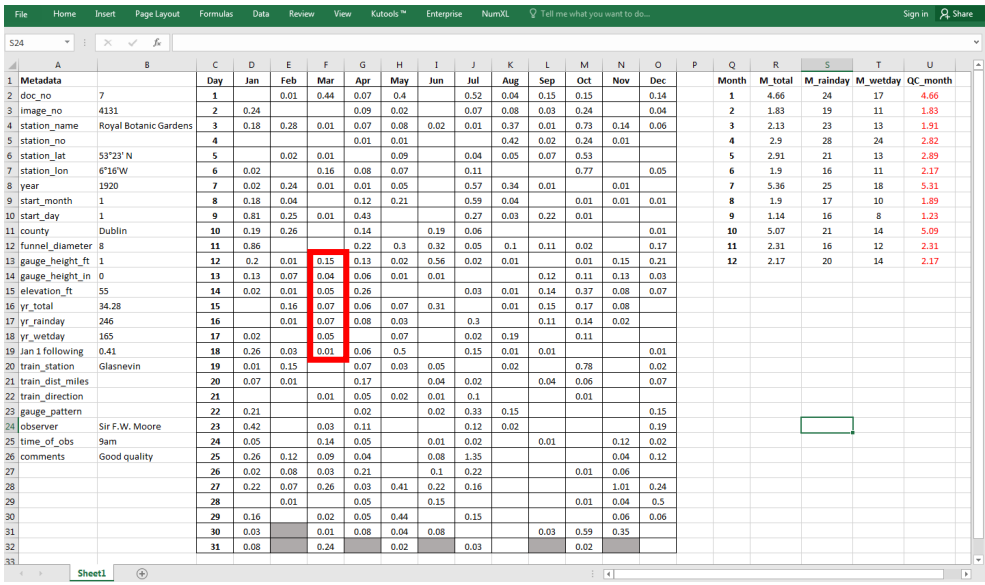
Ciara

Show parent | Edit | Split | Delete | Reply

Evaluating student performance



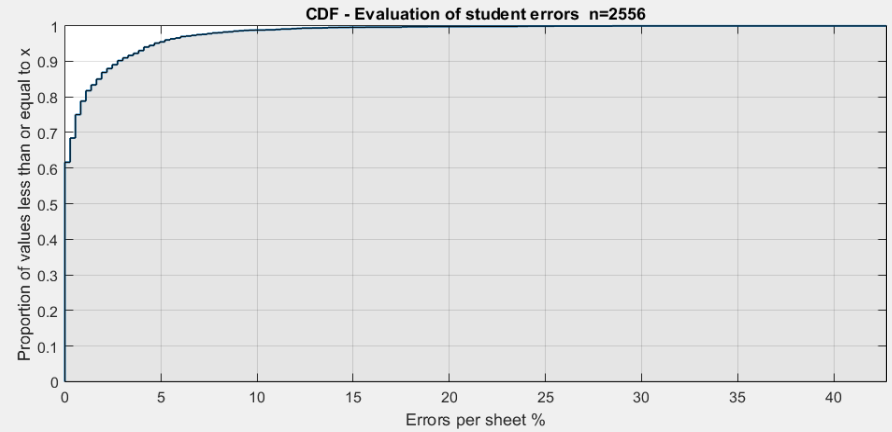
- Compare double keyed sheets.
- Create a master "correct" data file by examining highlighted differences.
- Run the comparison again to evaluate student performance.

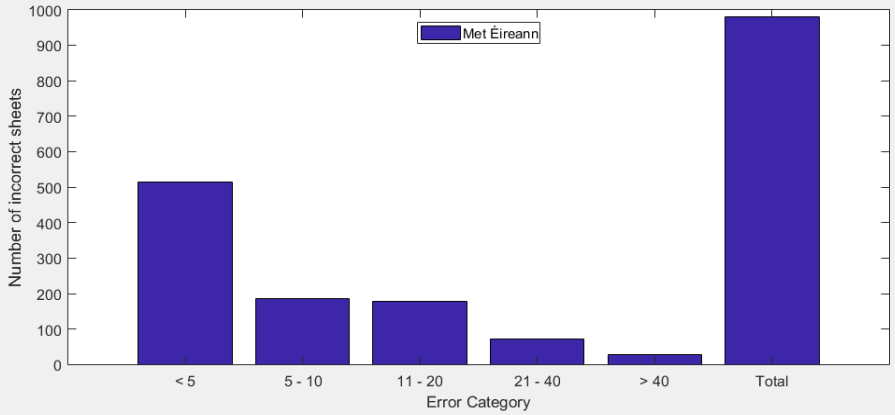


Metadata		Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
doc_no	7	1		0.01	0.44	0.07	0.4		0.52	0.04	0.15	0.15		0.14
image_no	4131	2	0.24			0.09	0.02		0.07	0.08	0.03	0.24		0.04
station_name	Royal Botanic Gardens	3	0.18	0.28	0.01	0.07	0.08	0.02	0.01	0.37	0.01	0.73	0.14	0.06
station_no		4				0.01	0.01			0.42	0.02	0.24	0.01	
station_lat	53°23' N	5		0.02	0.01		0.09		0.04	0.05	0.07	0.53		
station_lon	6°16'W	6	0.02		0.16	0.08	0.07		0.11			0.77		0.05
year	1920	7	0.02	0.24	0.01	0.01	0.05		0.57	0.34	0.01		0.01	
start_month	1	8	0.18	0.04		0.12	0.21		0.59	0.04		0.01	0.01	0.01
start_day	1	9	0.81	0.25	0.01	0.43			0.27	0.03	0.22	0.01		
county	Dublin	10	0.19	0.26		0.14		0.19	0.06					0.01
funnel_diameter	8	11	0.86			0.22	0.3	0.32	0.05	0.1	0.11	0.02		0.17
gauge_height_ft	1	12	0.2	0.01	0.15	0.13	0.02	0.56	0.02	0.01		0.01	0.15	0.21
gauge_height_in	0	13	0.13	0.07	0.04	0.06	0.01	0.01			0.12	0.11	0.13	0.03
elevation_ft	55	14	0.02	0.01	0.05	0.26			0.03	0.01	0.14	0.37	0.08	0.07
yr_total	34.28	15		0.16	0.07	0.06	0.07	0.31		0.01	0.15	0.17	0.08	
yr_rainday	246	16		0.01	0.07	0.08	0.03		0.3		0.11	0.14	0.02	
yr_wetday	165	17	0.02		0.05		0.07		0.02	0.19		0.11		
Jan 1 following	0.41	18	0.26	0.03	0.01	0.06	0.5		0.15	0.01	0.01			0.01
train_station	Glasnevin	19	0.01	0.15		0.07	0.03	0.05		0.02		0.78		0.02
train_dist_miles		20	0.07	0.01		0.17		0.04	0.02		0.04	0.06		0.07
train_direction		21			0.01	0.05	0.02	0.01	0.1			0.01		
gauge_pattern		22	0.21			0.02		0.02	0.33	0.15				0.15
observer	Sir F.W. Moore	23	0.42		0.03	0.11			0.12	0.02				0.19
time_of_obs	9am	24	0.05		0.14	0.05		0.01	0.02		0.01		0.12	0.02
comments	Good quality	25	0.26	0.12	0.09	0.04		0.08	1.35				0.04	0.12
		26	0.02	0.08	0.03	0.21		0.1	0.22			0.01	0.06	
		27	0.22	0.07	0.26	0.03	0.41	0.22	0.16				1.01	0.24
		28		0.01		0.05		0.15				0.01	0.04	0.5
		29	0.16		0.02	0.05	0.44		0.15				0.06	0.06
		30	0.03		0.01	0.08	0.04	0.08			0.03	0.59	0.35	
		31	0.08		0.24		0.02		0.03			0.02		

Month	M_total	M_rainday	M_wetday	QC_month
1	4.66	24	17	4.66
2	1.83	19	11	1.83
3	2.13	23	13	1.91
4	2.9	28	24	2.82
5	2.91	21	13	2.89
6	1.9	16	11	2.17
7	5.36	25	18	5.31
8	1.9	17	10	1.89
9	1.14	16	8	1.23
10	5.07	21	14	5.09
11	2.31	16	12	2.31
12	2.17	20	14	2.17

Evaluating student performance -results





An evaluation of student performance revealed that:

62% of student transcribed sheets had no errors

In 96% of student sheets fewer than 5% of data entries were incorrect
i.e., 96% of sheets had < 20 errors

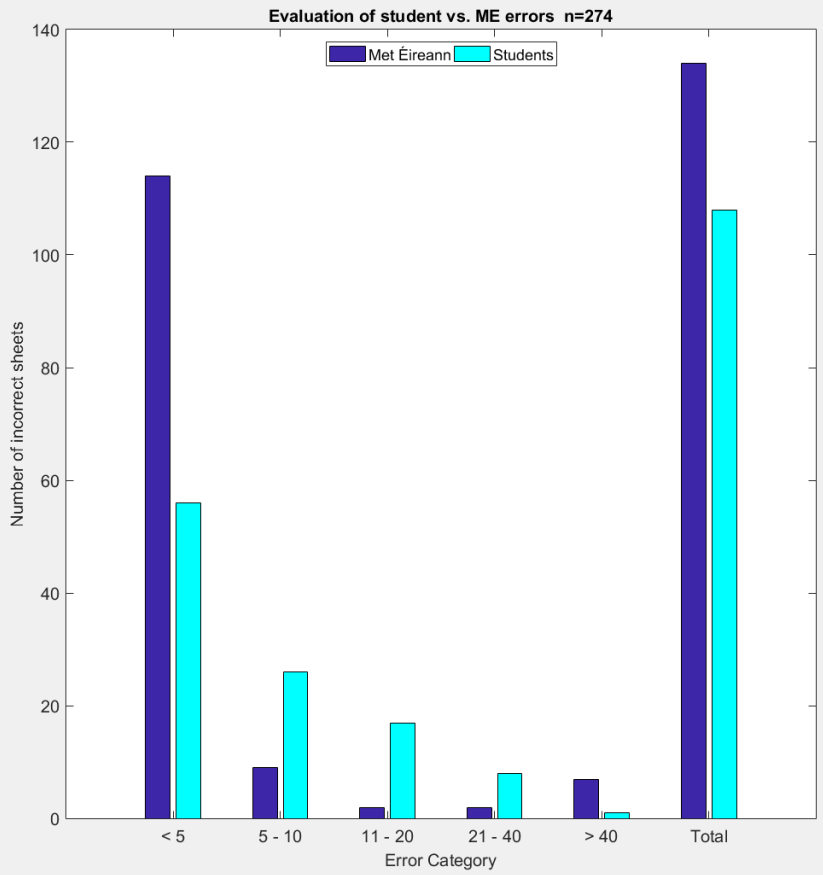
A review of all incorrect sheets reveals that:

52% of the sheets containing errors had fewer than 5 errors, 90% had fewer than 20 errors and only 3% had > 40 errors

Cumulative errors across all 2556 sheets transcribed by students reveals a percentage error of less than 1%

Empirical CDF showing the frequency (%) of student submissions being less than or equal to x, where x is the percentage of errors per sheet (top). Bar graph categorising the total number of incorrect sheets by actual number of errors per sheet (bottom).

Benchmarking student performance

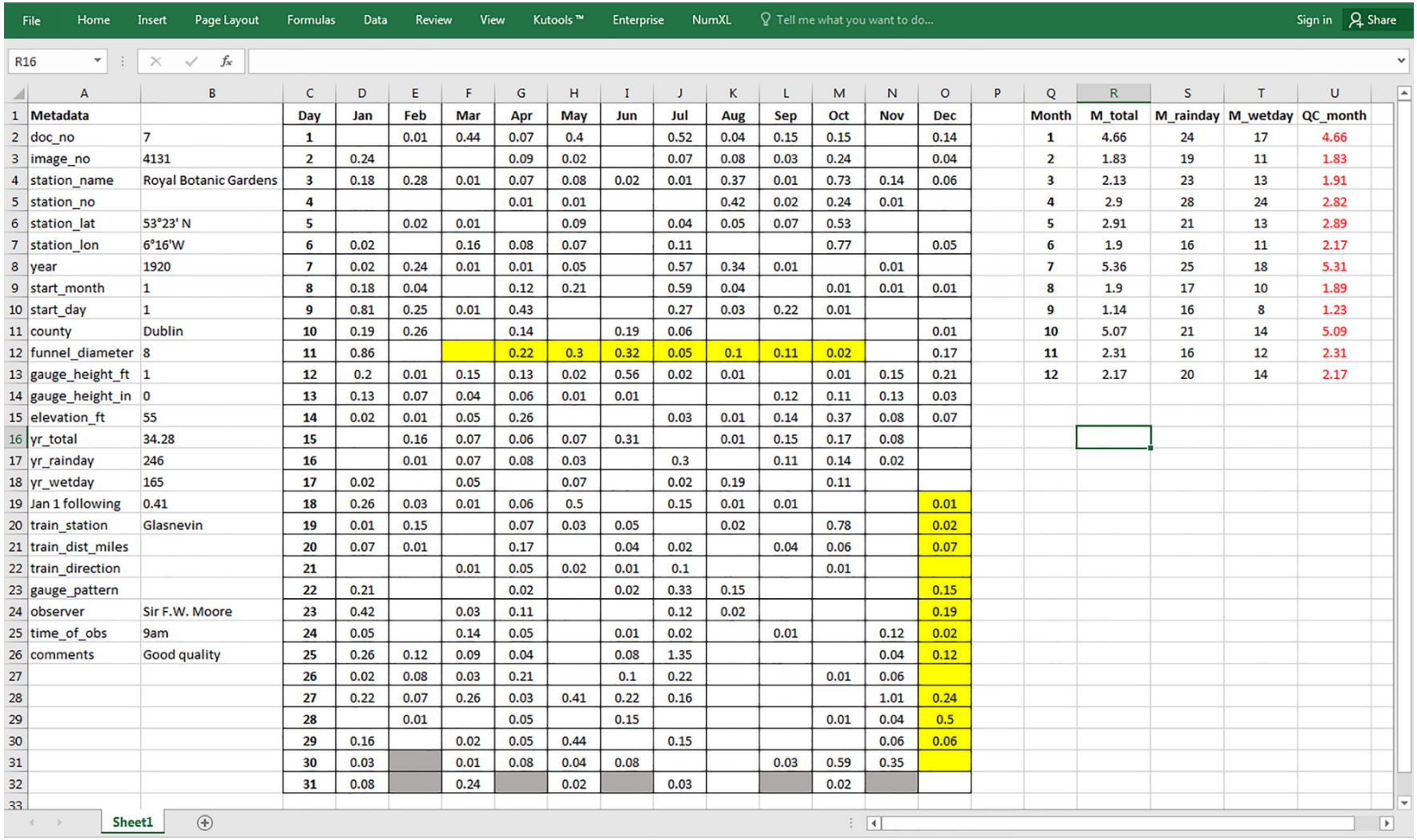


An assessment of the 274 sheets transcribed by both Met Éireann and the students revealed that while the students have a smaller number of incorrect files overall (39% for students compared to 49% for Met Éireann), the majority (85%) of Met Éireann's incorrect files lie in the lowest error propensity category.

Different approaches to the transcription process i.e., row based or column based had an impact on the number of errors produced within individual files.

Benchmarking students: Evaluation of errors for common sheets (n = 274). Proportion of incorrect files by error category.

Comparing double keyed sheets



	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	Metadata		Day	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec		Month	M_total	M_rainday	M_wetday	QC_month
2	doc_no	7	1		0.01	0.44	0.07	0.4		0.52	0.04	0.15	0.15		0.14		1	4.66	24	17	4.66
3	image_no	4131	2	0.24			0.09	0.02		0.07	0.08	0.03	0.24		0.04		2	1.83	19	11	1.83
4	station_name	Royal Botanic Gardens	3	0.18	0.28	0.01	0.07	0.08	0.02	0.01	0.37	0.01	0.73	0.14	0.06		3	2.13	23	13	1.91
5	station_no		4				0.01	0.01			0.42	0.02	0.24	0.01			4	2.9	28	24	2.82
6	station_lat	53°23' N	5		0.02	0.01		0.09		0.04	0.05	0.07	0.53				5	2.91	21	13	2.89
7	station_lon	6°16'W	6	0.02		0.16	0.08	0.07		0.11			0.77		0.05		6	1.9	16	11	2.17
8	year	1920	7	0.02	0.24	0.01	0.01	0.05		0.57	0.34	0.01		0.01			7	5.36	25	18	5.31
9	start_month	1	8	0.18	0.04		0.12	0.21		0.59	0.04		0.01	0.01	0.01		8	1.9	17	10	1.89
10	start_day	1	9	0.81	0.25	0.01	0.43			0.27	0.03	0.22	0.01				9	1.14	16	8	1.23
11	county	Dublin	10	0.19	0.26		0.14		0.19	0.06					0.01		10	5.07	21	14	5.09
12	funnel_diameter	8	11	0.86			0.22	0.3	0.32	0.05	0.1	0.11	0.02		0.17		11	2.31	16	12	2.31
13	gauge_height_ft	1	12	0.2	0.01	0.15	0.13	0.02	0.56	0.02	0.01		0.01	0.15	0.21		12	2.17	20	14	2.17
14	gauge_height_in	0	13	0.13	0.07	0.04	0.06	0.01	0.01			0.12	0.11	0.13	0.03						
15	elevation_ft	55	14	0.02	0.01	0.05	0.26			0.03	0.01	0.14	0.37	0.08	0.07						
16	yr_total	34.28	15		0.16	0.07	0.06	0.07	0.31		0.01	0.15	0.17	0.08							
17	yr_rainday	246	16		0.01	0.07	0.08	0.03		0.3		0.11	0.14	0.02							
18	yr_wetday	165	17	0.02		0.05		0.07		0.02	0.19		0.11								
19	Jan 1 following	0.41	18	0.26	0.03	0.01	0.06	0.5		0.15	0.01	0.01			0.01						
20	train_station	Glasnevin	19	0.01	0.15		0.07	0.03	0.05		0.02		0.78		0.02						
21	train_dist_miles		20	0.07	0.01		0.17		0.04	0.02		0.04	0.06		0.07						
22	train_direction		21			0.01	0.05	0.02	0.01	0.1			0.01								
23	gauge_pattern		22	0.21			0.02		0.02	0.33	0.15				0.15						
24	observer	Sir F.W. Moore	23	0.42		0.03	0.11			0.12	0.02				0.19						
25	time_of_obs	9am	24	0.05		0.14	0.05		0.01	0.02		0.01		0.12	0.02						
26	comments	Good quality	25	0.26	0.12	0.09	0.04		0.08	1.35				0.04	0.12						
27			26	0.02	0.08	0.03	0.21		0.1	0.22			0.01	0.06							
28			27	0.22	0.07	0.26	0.03	0.41	0.22	0.16				1.01	0.24						
29			28		0.01		0.05		0.15				0.01	0.04	0.5						
30			29	0.16		0.02	0.05	0.44		0.15				0.06	0.06						
31			30	0.03		0.01	0.08	0.04	0.08			0.03	0.59	0.35							
32			31	0.08		0.24		0.02		0.03			0.02								
33																					

Sheet1

ICARUS
Irish Climate Analysis and Research Units

Maynooth University
National University of Ireland Maynooth

MET éireann

Evaluating the student experience

Learning outcomes were designed to provide students with:

- A first-hand experience of working with historical climate observations.
- A critical appreciation of the processes involved in data rescue, digitisation and quality assurance procedures that are essential to understanding past climate variability and change.
- First-hand experience of the powerful contribution that citizen science can make to the study of climatology, geography and other disciplines.

Student Feedback:

>90% gained insights into the process of data rescue and an appreciation of the role of historical data in climate research.
>90% could see value in their work and were motivated by the fact that they were contributing to research.
80% stated that they would prefer to participate in assignments like this over other, more traditional, assignments.

Next steps for the project:

- Delineate the margins of the columns and rows in the Excel template to reduce propensity for errors.
- Highlight the benefits of using the row based approach employed by Met Éireann.
- Create a metadata file for the students.
- Enhance the research experience offered to the students.

Next steps for the data:

- Application of comprehensive quality assurance techniques.
- Analysis of the long-term record to assess changes in the characteristics of extreme rainfall events.

Summary

- Over 1300 annual rainfall sheets and associated metadata transcribed by students.
- Performance of the students comparable to the professionals - percentage error ~1%
- Positive feedback from students.
- Experience demonstrates the potential to extend this project to other universities.
- Project resources will be made available via maynoothuniversity.ie/icarus

Correspondence to: Ciara Ryan (ciara.ryan@mu.ie)

ICARUS
Irish Climate Analysis and Research Units

Maynooth University
National University of Ireland Maynooth

MET éireann