Today's Lab, Making Appendix 2: Data Rationalization

Transforming raw WxProj data into the final dataset….

DATE	TIME	TEMPERATURE			HUMIDITY (Note wick state & instrument type in e & RH column)				PRECIPITATION (Note when precipitation is from melted snow)				SKY Condition & CLOUD								WIND				PRESSURE			COMMENTS
Entered from earliest to latest date & time		Max Temp. (°C)	Min Temp. (°C)	Present Air Temp. (°C)	Wet bulb Temp. (°C)	F = wick is frozen	Dry bulb Temp. (°C)	e (hPa) & RH A=Assmann B=Bacharach W=Weksler	Snow Depth (mm) Ruler measurement	Snow Water Equivalent SWE (mm)	Rain gauge (mm) T = Trace	M = from melted rain gauge	SKY: CLR FEW SCT BKN OVC OBSCD MISG	Amount (8ths) low	Amount (8ths) mid	Amount (8ths) high	Cloud type: Use 2-letter cloud abbreviations; double dashes are for layers you cannot see. low	Cloud type mid	Cloud type high	Visual Observation: Direction Bearing (8-point compass)	Visual Observation: Beaufort number	UNBC Weather Station: Direction Azimuth (degrees °) from north	UNBC Weather Station: Speed (m/s)	Barometer (mm Hg)	Barometer Temp. (°C)	Corrected Pressure (hPa)	
Feb 12	7:48	-2.0	-8.5	-7.0	-6.3	F	-6.2	A	--		0	--	BKN	6	1	--	St	As	--	SW	1	239.6	1.96	700.8	20.0		Fog lower, in bowl (not in sky view) Frost on bench
Feb 12	12:30	-2.0	-8.0	-2.0	-3.1	F	-2.1	A	0		0	--	OVC	0	0	8	--	--	Cs	E	2	101	0.4	700.8	21.0		Clouds Changed
Feb 12	3:30	0.0	-3.0	-1.0	-1.5	F	-1.4	A	--		--	--	SCT	--	2	2	--	As	Cl	SW	1	248	1.4	700.4	22		
Feb 13	8:05	-0.5	-13.0	-12.5	-10.4	F	-10.2	A	--		0	--	CLR	--	--	--	--	--	--	SW	0	229	0.4	703.4	21		Frost on bench, Stratus clouds on W. Fan
Feb 13	12:30	-8.0	-13.0	-8.0	-8.0	F	-7.8	A	0		0	--	BKN	6	--	--	St	--	--	NE	2	81	1.6	703.5	21		low clouds are changing
Feb 13	15:50	-6.0	-8.0	-6.0	-7.6	F	-6.4	A	--		--	--	Few	1	--	--	Sc	--	--	NE	1	348	0.14	702.6	21		
Feb 13	8:04	-8.0	-15.0	-13.0	-10.9	F	-10.8	A	--		0	--	OBSCD	--	--	--	--	--	--	N	1	16	1.9	702.8	21		Frost on bench, Shelter, railings etc, Fog!
Feb-14	12:21	-12.0	-13.0	-11.0	-9.5	F	-9.5	A	--		0	--	OBSCD	--	--	--	--	--	--	NE	2	26	0.6	702.5	21.0		Fog; can't see anything
Feb 14	3:45	-5.0	-11.0	-5.0	-5.8	F	-5.3	A	--		--	--	CLR	--	--	--	--	--	--	NE	1	348	1.4	702.6	21.0		Frosty
Feb 15	7:35	-4.5	-13.5	-13.0	-11.4	F	-10.6	A	--		0	--	CLR	--	--	--	--	--	--	N	2	358	2.5	703.1	21.0		Frost on bench Fog & stratus clouds in bowl
Feb 15	12:36	-6.0	-13.0	-6.0	-7.2	F	-5.1	A	0		0	--	CLR	0	0	0	--	--	--	N	2	62	0.6	704.2	21.0		Clear sky & variable low speed winds.
Feb 15	3:45	-2.0	-6.0	-2.0	-4.6	F	-2.2	A	0		0	--	CLR	0	0	0	--	--	--	NE	1	52	0.9	705	21		Lovely day isn't

Each student calculates their own e (hPa) & RH (%) values from recorded observations.

Each student calculates their own SWE values from recorded observations.

Each student calculates their own pressures from recorded observations.

DO NOT REMOVE – LEAVE DATA SHEET WITH CLIPBOARD AT ALL TIMES

Today's Agenda:

Returned: **Lab 5, Lab 6, WxProj Data Collection Submission** (electronic data, diary & data collection marks). **Return your ziplock bag** for re-use, if you don't want it.

Handouts: Lab 6 Reflection **(blue sheet)**
2 photocopies of your original completed roof-top datasheet
1 blank final datasheet

Required: Read WxProj Handbook, Report Requirements (pp. 9 -14) & Slides for this lab.

Notice:

1) **Appendix 1 & 2 are part of your WxProj Report**; must be properly incorporated as appendices in your WxProj report's paper copy.
   - use a consistent appendix format in your report, but must fit requirements for each appendix
   - See WxProj Handbook pp. 2, 9-13; lab & lecture slides; your instructor when needed

2) **Produce Appendix 2 – done in today's lab.** *See next slide for more.*

3) **Produce Appendix 1** (your marked WxProj Diary) **on your own time. Keep it!**
   Format your marked weather diary as Appendix 1 ← ready for including in your report
   - use it as marked with comments; do not revise /redo, it is not graded again

Due ~2 day after lab (regular dropbox deadline): **Lab 6**, 2nd submission; **Appendix 2**

Data Rationalization & creating Appendix 2

What is data rationalization? The process of transforming raw data into a final dataset (involves data quality control; calculating final values), and documenting this process. Readers must understand your data & its evolution. [For your report, this is Appendix 2.]

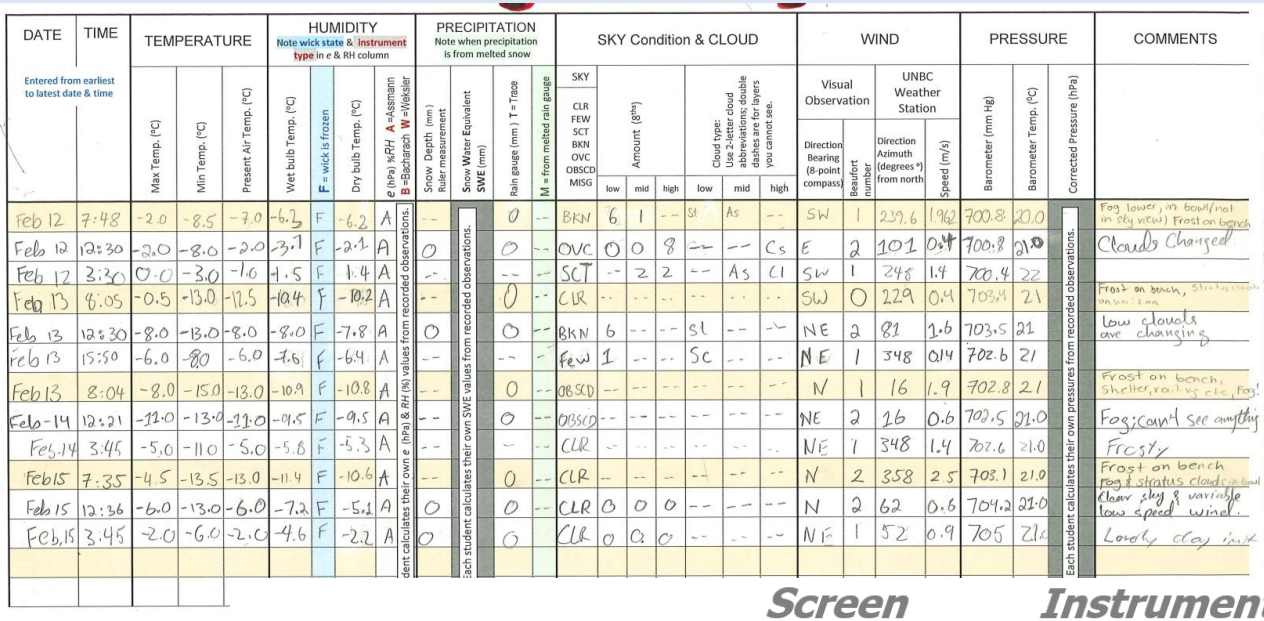
- Clearly document and show your data rationalization process. Documentation must succinctly justify your quality control decisions and be detailed enough for readers to follow each data issue, your considerations, its impact, and your resolution decision.
- Calculations clearly show details (values, units, constants, conversions, show how your final results are obtained)
- Data rationalization (concept and requirements):
  - → Presents original (raw) data & clearly identifies which data points are affected by specific data issues (or types of issues); done using footnotes, colours, etc. (see in-lab examples).
  - → Lists & explains each issue (or issue type); its impacts /limitations (for using that data); states considerations & what was done to address /rectify that issue (or each type of problem).
  - → Calculates computed values from raw measurements. Provides a clear, complete example calculation for each calculation type.
  - → Presents your final dataset as a new data table but shows the evolution of data points with issues by relating back to raw data that were identified as having issues and your list of data issues and explanations (points 1 and 2 above).

Data Rationalization:

Transforming raw data into its final data set.

Not just copying….

Before any data analysis, you must ensure and communicate that your data has value and is the highest quality it can be.



DATE | TIME — Entered from earliest to latest date & time

TEMPERATURE: Max Temp. (°C) | Min Temp. (°C) | Present Air Temp. (°C)

HUMIDITY — Note wick state & instrument type in e & RH column: Wet bulb Temp. (°C) | F = wick is frozen | Dry bulb Temp. (°C) | e (hPa) %RH A=Assmann B=Bacharach W=Weksler

PRECIPITATION — Note when precipitation is from melted snow: Snow Depth (mm) Ruler measurement | Snow Water Equivalent SWE (mm) | Rain gauge (mm) T = Trace | M = from melted rain gauge

SKY Condition & CLOUD: SKY (CLR FEW SCT BKN OVC OBSCD MISG) | Amount (8ths) low mid high | Cloud type: Use 2-letter cloud abbreviations; double dashes for layers you cannot see. low mid high

WIND: Visual Observation — Direction Bearing (8-point compass) | Beaufort number; UNBC Weather Station — Direction Azimuth (degrees °) from north | Speed (m/s)

PRESSURE: Barometer (mm Hg) | Barometer Temp. (°C) | Corrected Pressure (hPa) — Each student calculates their own pressures from recorded observations.

COMMENTS

DATE	TIME	Max Temp. (°C)	Min Temp. (°C)	Present Air Temp. (°C)	Wet bulb Temp. (°C)	F = wick is frozen	Dry bulb Temp. (°C)	e (hPa)	%RH A=Assmann B=Bacharach W=Weksler	Snow Depth (mm) Ruler measurement	Snow Water Equivalent SWE (mm)	Rain gauge (mm) T = Trace	M = from melted rain gauge	SKY	Amount (8ths) low	Amount (8ths) mid	Amount (8ths) high	Cloud type low	Cloud type mid	Cloud type high	Visual Observation: Direction Bearing (8-point compass)	Visual Observation: Beaufort number	UNBC Weather Station: Direction Azimuth (degrees °) from north	UNBC Weather Station: Speed (m/s)	Barometer (mm Hg)	Barometer Temp. (°C)	Corrected Pressure (hPa)	COMMENTS
Feb 12	7:48	-2.0	-8.5	-7.0	-6.7	F	-6.2		A	--		0	--	BKN	6	1	--	St	As	--	SW	1	239.6	19.2	700.8	20.0		Fog lower, in bowl (not in sky view) Frost on bench
Feb 12	12:30	-2.0	-8.0	-2.0	-3.1	F	-2.1		A	0		0	--	OVC	0	0	8	--	--	Cs	E	2	101	0.4	700.8	21.0		Clouds Changed
Feb 12	3:30	0.0	-3.0	-1.0	-1.5	F	-1.4		A	--		--	--	SCT	--	2	2	--	As	Cl	SW	1	248	1.4	700.4	22		
Feb 13	4:05	-0.5	-13.0	-12.5	-10.4	F	-10.2		A	--		0	--	CLR	--	--	--	--	--	--	SW	0	229	0.4	703.4	21		Frost on bench, Stratus on horizon
Feb 13	12:30	-8.0	-13.0	-8.0	-8.0	F	-7.8		A	0		0	--	BKN	6	--	--	Sl	--	--	NE	2	81	1.6	703.5	21		low clouds are changing
Feb 13	15:50	-6.0	-8.0	-6.0	-7.6	F	-6.4		A	--		--	--	Few	1	--	--	Sc	--	--	NE	1	348	0.4	702.6	21		
Feb 13	8:04	-8.0	-15.0	-13.0	-10.9	F	-10.8		A	--		0	--	OBSCD	--	--	--	--	--	--	N	1	16	1.9	702.8	21		Frost on bench, Shelter, railing etc. Fog!
Feb-14	12:21	-11.0	-13.0	-11.0	-9.5	F	-9.5		A	--		0	--	OBSCD	--	--	--	--	--	--	NE	2	26	0.6	702.5	21.0		Fog; can't see anything
Feb.14	3:45	-5.0	-11.0	-5.0	-5.8	F	-5.3		A	--		--	--	CLR	--	--	--	--	--	--	NE	1	348	1.4	702.6	21.0		Frosty
Feb15	7:35	-4.5	-13.5	-13.0	-11.4	F	-10.6		A	--		0	--	CLR	--	--	--	--	--	--	N	2	358	2.5	703.1	21.0		Frost on bench fog & stratus cloud on horizon
Feb 15	12:36	-6.0	-13.0	-6.0	-7.2	F	-5.1		A	0		0	--	CLR	0	0	0	--	--	--	N	2	62	0.6	704.2	21.0		Clear sky & variable low speed wind.
Feb,15	3:45	-2.0	-6.0	-2.0	-4.6	F	-2.2		A	0		0		CLR	0	0	0	--	--	--	NE	1	52	0.9	705	21.0		Lovely day init

Screen **Instrument**

DO NOT REM

DATE	TIME	Max Temp. (°C)	Min Temp. (°C)	Present Air Temp. (°C)	Wet bulb Temp. (°C)	F = wick is frozen	Dry bulb Temp. (°C)	e (hPa)	RH (%) A = Assmann B = Bacharach W = Weksler	Snow Depth (mm) Ruler measurement	Snow Water Equivalent SWE (mm)	Rain gauge (mm) T = Trace	M = from melted rain gauge	SKY	Amount (8ths) low	Amount (8ths) mid	Amount (8ths) high	Cloud type low	Cloud type mid	Cloud type high	Visual Observation: Direction Bearing (8-point compass)	Visual Observation: Beaufort number	UNBC Weather Station: Direction Azimuth (degrees °) from north	UNBC Weather Station: Speed (m/s)	Barometer (mm Hg)	Barometer Temp. (°C)	Corrected Pressure (hPa)	COMMENTS
Entered from earliest to latest date & time																												
2024 Data																												
Mon Feb 12						F			A				--															
Mon Feb 12						F			A				--															
Mon Feb 12						F			A				--															
Tue Feb 13						F			A				--															
Tue Feb 13						F			A				--															
Tue Feb 13						F			A				--															
Wed Feb 14						F			A				--															
Wed Feb 14						F			A				--															
Wed Feb 14						F			A				--															
Thu Feb 15						F			A				--															
Thu Feb 15						F			A				--															

Appendix 2 Required Format & Elements, in this order:

1. Title (/title page) & statement of purpose: A sentence or so, that describes the appendix & its purpose (like a caption). See your textbook /Course Manual for examples.
2. Final rationalized dataset with data issues indicated as footnotes &/colour-coded cells; these match the data problems identified in your raw (original) data & point #3 (below).
3. Your list of annotated data issues, their explanation, impacts & resolution. Best typed. A separate page (or pages) that tracks your data issues, explains them, indicates their impact on your report (especially important if they limit that data's use), and shows how you dealt with each issue (or type of issue).

   Your list must show how your raw data issues are related to their resolution in your final data set, and explain your data's evolution from its original raw form to the final data set used in your report.
4. Sample calculations: Show how your raw measurements becomes their final data values. Required: One sample calculation for each type of calculated value.
5. Raw (original) dataset with data problems indicated as footnotes &/colour-coded cells that match your #3 (above - list of data issues) and your final dataset indicated cells

Remember: Appendices have page numbers (e.g. 2-1, 2-2, etc. as it easily tracks them separately)

Find & flag your data errors…

Consider each…

What do you do with them?

What exactly is an error?

Objective vs Subjective Errors?

Data rationalization – types of errors

Objective errors - data impossibilities

- Max-Min-Present air temperature correspondence.
  - Max less than (<) Min;
  - Min greater than (>) Max;
  - Present air temperatures not equal to or between max & min
- Wet bulb temperatures > dry bulb temperatures
- Sky condition & cloud amounts don't agree or aren't possible
- Missing data? depends on why is it not there…..

Subjective errors – data you suspect but can't prove wrong

- Suspect data: Odd values /data inconsistencies but you don't have objective evidence to show /verify the value is incorrect, e.g. temperatures that seem too low /too high, or don't follow anticipated patterns
- Cloud types that seem unlikely given the sky /weather conditions
- Discrepancies between anemometer & visual wind observations (those not caused by differences in data gathering methods).
- Others? Inconsistencies, odd or anomalous values…

What do you do with problem data?

Must be addressed!

Can’t throw data away!

Data rationalization – Solutions

What do you do with problem data?

Repair /recover?

Replace?

Leave as is?

Identify issues & appropriate solutions:

1. **Understand your data errors** Speaking with those who collected your data or instructors can help. What is the issue's impact /severity? How will you use the problem value(s) or observation(s)? What does the issue limit?

2. Seek **objective** over **subjective** solutions. Solutions supported by independent evidence are best. Identify multiple options; use a hierarchical approach to select the best solution considering objectivity & impact.

3. Consider your **solution's impact on using your data**

4. **Be transparent:** Track and clearly, but succinctly explain your process, thinking and choices (explain the issue; its resolution option(s); indicate the impact /consequences of your best choice, state your decision)

Subjectively changing data is never appropriate!

An undocumented memory of data is not objective. It's the poorest solution for problem data (no evidence, not transparent). Are there other, better options?

Replace?

WxProj methodology is a case study that has multiple data sources & data replication! **Recovering data by replacing it may be possible** (though not always). Consider all your data options. What are the consequences in each case?

Replacing data MUST be documented.
Replacements must improve your data issues!

- Look for best-matched values from a replicate dataset. Consider each reading's observation time & instrument compatibility
- Explicitly identify each data issue. State the problem; succinctly explain the error, your considerations & their implications; indicate the affected data point(s)
- Track & explain your solution. State your resolution. Report the source of each replacement data value, use footnotes to link this information to your original /raw data issues and your list of data problems with explanations.

Sometimes there are no good solutions….

Leave as is?

Yes, but you still must:

- **identify and track data issues**
- **understand and consider their limitations on your data** as some issues can limit interpretation and analysis that can affect your understanding of the weather story (i.e. what occurred & why it happened)
- **When doing nothing but recognising the issue and its impacts is your best solution**, indicate nothing else was done to address the issue, your reasoning, and its consequences for your data /report.]

Seek instructor assistance with these issues

Data rationalization – Calculating final values

Compute final values & Example Calculations

(as done in labs)

- ***e*, Relative Humidity** (track e*, constants, T or T_w values)
- **Precipitation** (convert snowfall measurements to water equivalents)
- **Pressure** (report as a corrected station pressure in hPa)

Use Excel to speed calculating, but track e*, constants, correction factor values. Appropriately record computed data on your final datasheet.

Create Appendix 2 (a required report component)

- Put Appendix 2 elements together; in the required order
- Add appropriate titles /subtitles, statements of purpose, page #'s
- Revisit Appendix 2 requirements ensure nothing is missed…..

So, what is an Appendix?

- See your textbook or Course Manual for examples
- Appendices supplement a report without interrupting the report's focus or flow.
- Like figures, each appendix must be descriptively titled & have a statement of purpose (like a caption) to make the purpose / use of the appendix clear to a reader
- When appendices have multiple components, each component must have a sub-title / statement of purpose
- Like figures, every appendix must be referred to in the body of your report
- Appendices have page numbers (e.g. either carry on from report; or use a system of 1-1, 1-2; 2-1, 2-2…. or A-1, A-2, A-3….etc.)

Recap: What is Appendix 2?

- Presents the final and original raw dataset used in your report
- Transparently shows the evolution of your original raw data (it's quality control measures and calculations) to its final form.
- Lists your specific data issues, their impact, & their resolution
  Presents your final & raw (original) data, identifies any data issues, and shows /explains data problem resolutions. Appendix 2 makes it possible to understand the evolution /progression of your data.
- Shows data calculation details

Appendices 1 & 2 are required report elements.

- Appendix 1: your returned & graded weather diary in its original form
- Appendix 2: your rationalized dataset and its evolution (from raw to final with clear documentation). Follows the required format. For data tables, hand-written is most time efficient

Recap: Appendix 2 Elements (in this order)

- **Title & statement of purpose** e.g. "Appendix 2…. & a descriptive title", a brief description of Appendix 2's purpose.
- **Final, rationalized dataset** with coloured footnotes /symbols that match those in your list of errors and raw data. Clearly show your data's evolution.
- **Your list of identified data issues**, their explanation, impacts, and resolution. Your list should match your raw dataset footnotes /symbols and show how you resolved each issue in your final dataset.
- **Sample calculations** showing how you computed final data set values.
- **Original raw rooftop dataset** with footnotes or symbols that match those in your list of errors, and final dataset.

Remember: Indicate linked changes; use the same footnotes /symbols /colour-coding on your original /raw dataset, listing of errors, and final dataset. Legends are useful for explaining colour coding and symbols. Appendices have page numbers.

Data rationalization – Identify & track

- **Must clearly show how you obtain your final dataset!**
  Use highlighting to draw attention to affected cells on both your raw and final datasets.

  Colours make footnotes easier to see!

How? **Here's an easy footnoted way to do this…**

- Highlight cells with issues
- Explicitly identify errors. Can you group types of errors? Annotate error cells (groups of cells) with a footnote number.
- Create a footnote list that explains each data issue (or types of issue), your considerations, & their impact on your report. Indicate how each issue was addressed /resolved.
- **Use:**
  - the same footnotes (e.g. 1, 2, 3) in the cells with issues in both your raw & final data
  - list each footnote, explain the data issue, its impact, and its resolution. Using sub-notes (e.g. 1a, 1b) can be useful to indicate multiple errors of the same type

This week:

Due in your dropbox by your lab's assignment deadline:

- **Appendix 2** & **Lab 6** (2nd submission)

Next week:

1) Stacked time series graphing (worth 1%; marked in lab)

   Use your marked Appendix 2 data to create a draft stacked time series graph. Must show your draft graph by the end of the lab & bring it for use in your instructor interview.

2) Setting up the WxProj Instructor meeting schedule for the following week – all interviews are during your lab times!