

Review of “A rescued dataset of sub-daily meteorological observations for Europe and the southern Mediterranean region, 1877-2012” by Linden Ashcroft et al.

The paper by Ashcroft et al. describes their strategy and approaches to rescue sub-daily meteorological observations within the UERRA project. They address the important issue of large amounts of undigitized data that are in danger of being lost completely. The resulting dataset contains a wide range of relevant variables. These data are of great value for the science community. Besides providing valuable data, the work contributes to foster data rescue and to improve data rescue approaches. Therefore, I recommend acceptance of the paper after several important corrections and clarifications.

Major comments

1. There seem to be some inconsistencies in the dataset. Below I show some examples of what I found:

a. In some files, days without measurements are not shown:

ALG0060355	SkiddacapBougarouni	60355	37.08	6.47	195	so04	28	28	28	28	3.0.0	f130	nc	1931-06-29T07:00
ALG0060355	SkiddacapBougarouni	60355	37.08	6.47	195	so04	22.4	22.4	22.4	22.4	3.0.0	f130	nc	1931-06-30T07:00
ALG0060355	SkiddacapBougarouni	60355	37.08	6.47	195	so04	35	35			3.0.0	f136	nc	1931-07-02T07:00
ALG0060355	SkiddacapBougarouni	60355	37.08	6.47	195	so04	26	26			3.0.0	f136	nc	1931-07-03T07:00
ALG0060355	SkiddacapBougarouni	60355	37.08	6.47	195	so04	25.6	25.6			3.0.0	f136	nc	1931-07-04T07:00
ALG0060355	SkiddacapBougarouni	60355	37.08	6.47	195	so04	25.4	25.4			3.0.0	f136	nc	1931-07-05T07:00
ALG0060355	SkiddacapBougarouni	60355	37.08	6.47	195	so04	25.6	25.6			3.0.0	f136	nc	1931-07-11T07:00
ALG0060355	SkiddacapBougarouni	60355	37.08	6.47	195	so04	32.4	32.4			3.0.0	f136	nc	1931-07-15T07:00
ALG0060355	SkiddacapBougarouni	60355	37.08	6.47	195	so04	23	23			3.0.0	f136	nc	1931-07-16T07:00

In some other cases, the time steps are continuous, but there is no raw data shown:

SLV0014249	NovoMesto	99999	45.8	15.18	220	so65	13	13	13	13	3.0.0	f110	nc	1964-12-29T99:99
SLV0014249	NovoMesto	99999	45.8	15.18	220	so65	9	9	9	9	3.0.0	f110	nc	1964-12-30T99:99
SLV0014249	NovoMesto	99999	45.8	15.18	220	so65	9	9	9	9	3.0.0	f110	nc	1964-12-31T99:99
SLV0014249	NovoMesto	99999	45.8	15.18	220	so65					3.0.0	f113	nc	1965-01-01T99:99
SLV0014249	NovoMesto	99999	45.8	15.18	220	so65					3.0.0	f113	nc	1965-01-02T99:99
SLV0014249	NovoMesto	99999	45.8	15.18	220	so65					3.0.0	f113	nc	1965-01-03T99:99
SLV0014249	NovoMesto	99999	45.8	15.18	220	so65					3.0.0	f113	nc	1965-01-04T99:99
SLV0014249	NovoMesto	99999	45.8	15.18	220	so65					3.0.0	f113	nc	1965-01-05T99:99
SLV0014249	NovoMesto	99999	45.8	15.18	220	so65					3.0.0	f113	nc	1965-01-06T99:99
SLV0014249	NovoMesto	99999	45.8	15.18	220	so65					3.0.0	f113	nc	1965-01-07T99:99
SLV0014249	NovoMesto	99999	45.8	15.18	220	so65					3.0.0	f113	nc	1965-01-08T99:99
SLV0014249	NovoMesto	99999	45.8	15.18	220	so65					3.0.0	f113	nc	1965-01-09T99:99
SLV0014249	NovoMesto	99999	45.8	15.18	220	so65					3.0.0	f113	nc	1965-01-10T99:99
SLV0014249	NovoMesto	99999	45.8	15.18	220	so65					3.0.0	f113	nc	1965-01-11T99:99
SLV0014249	NovoMesto	99999	45.8	15.18	220	so65					3.0.0	f113	nc	1965-01-12T99:99
SLV0014249	NovoMesto	99999	45.8	15.18	220	so65					3.0.0	f113	nc	1965-01-13T99:99
SLV0014249	NovoMesto	99999	45.8	15.18	220	so65					3.0.0	f113	nc	1965-01-14T99:99
SLV0014249	NovoMesto	99999	45.8	15.18	220	so65	14				3.0.0	f113	nc	1965-01-15T99:99
SLV0014249	NovoMesto	99999	45.8	15.18	220	so65					3.0.0	f113	nc	1965-01-16T99:99
SLV0014249	NovoMesto	99999	45.8	15.18	220	so65					3.0.0	f113	nc	1965-01-17T99:99

In the second case, not showing raw data but using the flag f113 is not compatible. I would suggest to always use continuous time steps within each time series and to include a code for “missing value” and for “test not applied”. Continuous time steps and completely filled fields in the table will be much more user friendly.

b. In some files, undefined numbers/texts appear in the column of the raw data:

ALG0060357	Annabacpedegarde	60357	36.97	7.79	161	so04	29.8	29.8	29.8	29.8	3.0.0	f130	nc	1922-08-07T07:00
ALG0060357	Annabacpedegarde	60357	36.97	7.79	161	so04	NO DATA				3.0.0	f111	nc	1922-08-13T07:00
ALG0060357	Annabacpedegarde	60357	36.97	7.79	161	so04	26.2	26.2	26.2	26.2	3.0.0	f130	nc	1922-08-15T07:00
SLV0014249	NovoMesto	99999	45.8	15.18	220	so65	2	2	2	2	3.0.0	f110	nc	1975-03-31T07:00
SLV0014249	NovoMesto	99999	45.8	15.18	220	so65	42-44				3.0.0	f117	nc	1975-08-01T07:00
SLV0014249	NovoMesto	99999	45.8	15.18	220	so65	56-58				3.0.0	f117	nc	1975-08-01T19:00
SLV0014249	NovoMesto	99999	45.8	15.18	220	so65	0	0	0	0	3.0.0	f110	nc	1976-01-01T07:00
EGY0062342	E[Mansura	62342	31.34	31.08	10	so16	23	23	23	23	3.0.0	f130	nc	1953-05-30T18:00
EGY0062342	E[Mansura	62342	31.34	31.08	10	so16	25	25	25	25	3.0.0	f130	nc	1953-05-31T06:00
EGY0062342	E[Mansura	62342	31.34	31.08	10	so16	-100				3.0.0	f117	nc	1953-05-31T12:00
EGY0062342	E[Mansura	62342	31.34	31.08	10	so16	-100				3.0.0	f117	nc	1953-05-31T18:00
EGY0062342	E[Mansura	62342	31.34	31.08	10	so16	-100				3.0.0	f117	nc	1953-06-01T06:00
EGY0062342	E[Mansura	62342	31.34	31.08	10	so16	-100				3.0.0	f117	nc	1953-06-01T12:00

c. For some files, the station ID seems to be corrupt:

ALG10XXXXX	FortNational	99999	36.63	4.2	942	so04	16	16	16	16	3.0.0	f130	nc	1938-10-24T07:00
ALG10XXXXX	FortNational	99999	36.63	4.2	942	so04	16	16	16	16	3.0.0	f130	nc	1938-10-25T07:00
ALG10XXXXX	FortNational	99999	36.63	4.2	942	so04	12.3	12.3	12.3	12.3	3.0.0	f130	nc	1938-10-26T07:00
ALG10XXXXX	FortNational	99999	36.63	4.2	942	so04	10.2	10.2	10.2	10.2	3.0.0	f130	nc	1938-10-27T07:00

d. For wind direction, there appears a “C” and a “0” in some files:

CRO0014236	Zagreb-Gr1c	14236	45.82	15.98	157	so56	ESE	ESE	ESE	112.5	3.0.0	f110	16ps	1949-01-27T14:00
CRO0014236	Zagreb-Gr1c	14236	45.82	15.98	157	so56	C	C	C	0	3.0.0	f110	16ps	1949-01-27T21:00
CRO0014236	Zagreb-Gr1c	14236	45.82	15.98	157	so56	WNW	WNW	WNW	292.5	3.0.0	f110	16ps	1949-01-28T07:00
CRO0014236	Zagreb-Gr1c	14236	45.82	15.98	157	so56	S	S	S	180	3.0.0	f110	16ps	1949-01-28T14:00
CRO0014236	Zagreb-Gr1c	14236	45.82	15.98	157	so56	C	C	C	0	3.0.0	f110	16ps	1949-01-28T21:00
CRO0014236	Zagreb-Gr1c	14236	45.82	15.98	157	so56	C	C	C	0	3.0.0	f110	16ps	1949-01-29T07:00
CRO0014236	Zagreb-Gr1c	14236	45.82	15.98	157	so56	S	S	S	180	3.0.0	f110	16ps	1949-01-29T14:00
CRO0014236	Zagreb-Gr1c	14236	45.82	15.98	157	so56	C	C	C	0	3.0.0	f110	16ps	1949-01-29T21:00
CRO0014236	Zagreb-Gr1c	14236	45.82	15.98	157	so56	C	C	C	0	3.0.0	f110	16ps	1949-01-30T07:00
CRO0014236	Zagreb-Gr1c	14236	45.82	15.98	157	so56	SW	SW	SW	225	3.0.0	f110	16ps	1949-01-30T14:00
ALG0060355	skiddacapBougarouni	60355	37.08	6.47	195	so04	WSW	WSW	WSW	247.5	3.0.0	f130	16ps	1933-03-30T07:00
ALG0060355	skiddacapBougarouni	60355	37.08	6.47	195	so04	0	0	0	0	3.0.0	f130	16ps	1933-03-31T07:00
ALG0060355	skiddacapBougarouni	60355	37.08	6.47	195	so04	0	0	0	0	3.0.0	f130	16ps	1933-04-01T07:00
ALG0060355	skiddacapBougarouni	60355	37.08	6.47	195	so04	E	E	E	90	3.0.0	f130	16ps	1933-04-02T07:00
ALG0060355	skiddacapBougarouni	60355	37.08	6.47	195	so04	0	0	0	0	3.0.0	f130	16ps	1933-04-03T07:00
ALG0060355	skiddacapBougarouni	60355	37.08	6.47	195	so04	0	0	0	0	3.0.0	f130	16ps	1933-04-04T07:00
ALG0060355	skiddacapBougarouni	60355	37.08	6.47	195	so04	0	0	0	0	3.0.0	f130	16ps	1933-04-05T07:00
ALG0060355	skiddacapBougarouni	60355	37.08	6.47	195	so04	W	W	W	270	3.0.0	f130	16ps	1933-04-06T07:00
ALG0060355	skiddacapBougarouni	60355	37.08	6.47	195	so04	W	W	W	270	3.0.0	f130	16ps	1933-04-07T07:00
ALG0060355	skiddacapBougarouni	60355	37.08	6.47	195	so04	W	W	W	270	3.0.0	f130	16ps	1933-04-08T07:00
ALG0060355	skiddacapBougarouni	60355	37.08	6.47	195	so04	0	0	0	0	3.0.0	f130	16ps	1933-04-09T07:00
ALG0060355	skiddacapBougarouni	60355	37.08	6.47	195	so04	0	0	0	0	3.0.0	f130	16ps	1933-04-10T07:00

Are these codes for “no wind”? If yes, there should be a single code only, and it should not result in a “0” in the columns “convertedvalue” (0°=360°). I recommend to include a special code for “no wind”.

- e. Please clean the database from such inconsistencies. I recommend to create and run specific tests for each column in order to detect and correct cases as shown above.
2. For selecting areas and time periods to digitize observations the authors used the MARS. In the text, questions remain regarding these choices:
 - a. In the text, it does not get clear how comprehensive the data collected in the MARS are. Are there other digitized datasets that are not yet included in MARS (and could fill gaps more efficiently than digitizing analog data sources)? Did you search for such datasets?
 - b. Give some examples on how your work affects the MARS. How much did the data completeness increase in certain areas and/or time periods? How much longer near-complete time series are available thanks to your efforts? A simple figure could demonstrate the benefit of the new dataset.
 - c. Furthermore, giving a brief greater context of data availability the investigated regions (e.g. demonstrating that observations from these early time period are quite unique) would highlight the importance of the newly digitized dataset.
 - d. Data users are normally interested in long-term station records. Providing some instructions on how data user can generate the most comprehensive time series possible would be very useful (E.g. where to find further digitized data to combine the newly digitized data with? In the MARS or are there other data sources available?).
 3. Some questions regarding HQC:
 - a. You applied the HQC only on the original time scale of the observations, is that correct? Aggregating the observation to daily data may strongly increase the number of neighboring stations usable for spatial consistency checks.
 - b. Did you include other stations resp. time series for HQC or did you just use the data you digitized? Using as many time series as possible will strongly increase the efficiency of the test. Please define these points clearly in the text and explain your decisions.
 - c. Similarly, you applied the SAQC on the time scale of observations only? Do you think you could have identified more errors if aggregating the data on larger time scales (particularly on daily time scale) and run adapted SAQC tests again?
 - d. Most of the tests in HQC are not spatial consistency test (Table 5) and seem to be repetitions or additions to the SAQC tests. Please define the HQC more clearly and explain better why you choose to use these tests.

Minor comments

P2 line 15: delete “in Germany”.

P4 line 5: Please indicate earlier in the text how you define sub-daily (half-hourly to daily) and/or refer earlier to Table 7 (e.g. here in 2.2).

P5 line 8: “key as you see” approach: How where the digitizers supposed to proceed in ambiguous cases such as unclear handwriting? Was there a code to mark unrecognizable fields?

P5 line 12: Add the WMOs “GUIDELINES ON BEST PRACTICES FOR CLIMATE DATA RESCUE” as a reference.

P5 line 17: What variations are you talking about here? Can you give an example?

P5 line 26: Delete “e.g.” before “Fig. 3”.

P6 line 3: Could you make available these metadata? They could be important particularly for subsequent data homogenization.

P6 line 5: A complete QC procedure is supposed to detect non-systematic errors in the data too in my opinion. A simple solution would be to delete “non-systematic”.

P6 line 18: You imply that automated QC approaches include spatial consistency tests which is not necessarily the case.

P6 line 31: Who are these climatologists? Were they part of UERRA or did you work together with the National Weather Services? Did you do this tests on the all digitized data or at random?

P6 line 33: “... common digitization errors...”

P8 line 3: Please show what “HQC” abbreviates.

P8 line 7-11: It does not seem very logical that low spatial resolution and observations taken at inconsistent times lead to a large number of positive flags (I would expect some thresholds for correlation and the number of neighboring stations to apply the test on a time series). Please explain.

P8 line 9-11: Rephrase the sentence.

P8 line 12-14: I would expect that reducing these thresholds would allow to apply the test on more time series, but rather produce more false positives. Was the problem of using 10 neighbors too low station correlation, which would mean that correlation is not included in the test?

P8 line 14-15: These test are not spatial consistency checks. According to the description of HQC, I would expect spatial consistency tests only.

P8 line 29-31: It remains unclear what tests the final check includes and what the flag of 3 is about. Please explain more clearly.

P8 line 34 – P9 line 2: This paragraph is mostly a repetition.

P9 line 39-40: Was this digitizer part of the UERRA project? Why did you set the data to missing and did not use the appropriate flag for the error?

P10 line 6-9: How can changing frequency of observation skew resulting analysis (aggregating the hourly data should be sufficient)? What correction did you apply?

P10 line 10-13: This paragraph is not clear to me. Are the 6-hourly observations from Slovenia too and/or from other countries?

P10 line 14-20: This paragraph is not very clear to me. The case of the effects of WW2 on the data seems very interesting, but it should be explained a bit more how the specific problems may have occurred.

P10 line 21: Delete duplication of “Quality control result”.

P10 line 23: I think the source resolution and temporal gaps should not lead to more frequent detections by the QC method?

P10 line 25: For clarity, I would call flags always “fl..” and not just by the number (here “36”).

P10 line 28: Same as previous comment.

P10 line 31 and 34: What is described cannot be seen clearly in Fig. 12. See also my suggestions for adaptations of Fig. 12.

P10 line 32-33: Do you want to say that about the same amount of observations was flagged by spatial consistency tests like in other analyses?

P10 line 37-38: It is unclear to what the last sentence refers to.

P11 line 4-5: Meaning of prefix “4” remains rather unclear. Please define more clearly.

P11 line 36: “Triple, or even five-time keying...”: Where do these numbers come from? Reference?

P12 line 4-5: Please reformulate the sentence to make it more clear.

P12 line 10-12: Even though the statement is probably generally true, can you conclude that (in such a strong form) from your work? The fraction of errors detected in your dataset is rather small.

P12 line 25-26: I think the recommendation of five-time keying applies rather to specific examples (such as these observation take at sea) than it can be considered as a general rule?

P12 line 29: Delete “and useful”.

P12 line 34-39: What are the concrete conclusions of these assessments? Does it result in integration or removal of data sources? Correction/prevention of errors e.g. by re-scanning original data sources? Giving more specific instructions of digitizers?

P13 line 1-13: This paragraph is rather unclear and I suggest to restructure it. Suggestion: 1. Our findings in the case of Zagazig do not confirm WMOs recommendation regarding templates. Mention some advantages of using templates too. 2. The case highlights the importance of creating user friendly templates. 3. Recommendation of creating user friendly templates (e.g. include feedback of digitizers when designing templates).

P13 line 22-23: Delete “It is no use... already complete.”.

P13 line 23-25: Suggestion for this sentence: “Conducting QC as soon as data become available means that the digitizers may detect own keying errors, and that you can advise them on how to increase the quality of their work.”.

P13 line 26-31: Did you check for and detect such systematic quality issues? If yes, how was that included in your study respectively why not? If you did not check: Do you think such systematic quality issues are a potential problem in your dataset?

P14 line 6: “Manual checking of values and decisions based on expert knowledge may mean...”

P14 line 11: Replace “The digitized dataset” by “It”.

Table 2: I suggest to always use the same number of decimals within one field.

Table 3: Do not use the same illustration for different tests (Climatic outliers, Big jumps and sharp spikes).

Table 3: I suggest to mention the change in the Duplicate values test only in the text and not in the table.

Table 3: Precipitation totals seems not to be a test but just a daily aggregation. Please specify or remove it from the table.

Table 4: Define more clearly to which test the flags relate to (e.g. does If10 mean “Passed SAQC”?).

Table 5: I think this table could be moved to the supplementary material. Furthermore, only a few tests are spatial consistency test (HQC is described as QC for spatial consistency in the text).

Table 6: The caption of the table is unclear, please reformulate.

Table 6: Observing times are mostly not in accordance with observing times from Table 7. Please explain these differences.

Figure 1: Remove small titles above figures.

Figure 1: Years shown in small titles above figures (1960-2010) not in accordance with time period in legend (1950-2010).

Figure 2: Delete last sentence (same to find in the text) or correct it and provide a direct link to the images.

Figure 6: Is the title needed?

Figure 6: I also suggest to replace "Percentage of tested data %" by "Flagged observations [%]".

Figure 6: Show variable acronyms horizontally.

Figure 7: Correct last sentence in caption.

Figure 7: I suggest to remove longitude and latitude here.

Figure 7: An indication (or additional plot) of the length of the near-complete time series of these stations after filling data gaps within your study would be an interesting asset for data users.

Figure 8: remove "1e6" from plots.

Figure 8 d): Invert scale of hours of the day, separate "daily" more clearly from the rest would be helpful.

Figure 8: I suggest to replace "Data count" by "Observations".

Figure 9: I would remove the first pie-chart from the figure, describing the numbers in the text is sufficient.

Figure 9: It is not clear how you chose the flags shown here (% of each flag of all flagged values? Other flags too small to be represented? Very similar to Fig. 12?). Please clarify or remove the figure.

Figure 10 d): I do not think this figure provides much information, I suggest to remove it.

Figure 10: Remove "1e1" from figures.

Figure 10: I suggest to replace "Percentage of total data %" by "Flagged observations [%]".

Figure 10: Colors are hard to distinguish.

Figure 11: Remove title.

Figure 11: Replace "Percentage of tested data %" by "Flagged observations [%]".

Figure 12: Changing % between "Total" and "Flags" is unclear. I suggest to show the fraction of flagged observation of the total in %.

Figure 12: The order in "Flags" seems a bit random.

Figure 12: The color code is unclear. There are three darkness classes for each color, but in the caption only two classes are explained.

Figure 12: What group is "Passed SAQC but recheck required"? Why would you recheck observations that passed the SAQC test?