



Assessing the Quality of Non-Professional Meteorological Data for Operational Purposes

D. Sládek* and P. Kolář

University of Defence, Brno, Czech Republic

The manuscript was received on 10 October 2022 and was accepted after revision for publication as research paper on 11 December 2023.

Abstract:

Non-professional weather stations are often omitted from the networks of standard/professional stations at various spatial scales. Nevertheless, there are many tasks when such non-professional datasets can serve as the only or the most relevant available source respectively. Its acquisition costs, sufficient quality and capacity together with its moveability represent properties that should be taken into consideration when planning operational usage of various meteorological data. In this paper, we focus on the datasets of air temperatures and relative humidities measured both with professional and non-professional devices at nearly the same location. Four years of almost continual measurements (2016-2019) ensure robust sample of mutual comparison, which we analyze in the paper more in detail in order to assess the potential of non-professional datasets for utilization in aviation meteorology. Particular issues such as value difference patterns, large errors occurrence, temporal signal stability and seasonality are elaborated as well.

Keywords:

air temperature, exploratory analysis, meteorological measurements, relative humidity

1 Introduction

Non-professional weather stations can be a useful source of information for a variety of common tasks. The network of professional stations is often either not dense enough in many areas of the world, or the spatial distribution of stations can be inhomogenous and, moreover, their accuracy for many tasks is therefore rather questionable. Although this topic has been already researched by international scientific groups, e.g. [1, 2], comparison of two simultaneously measuring stations on one location is not commonly applied [3]. Researchers usually assess more than two weather

* Corresponding author: Department of Military Geography and Meteorology, University of Defence, Kounicova 156/65, CZ-662 10 Brno, Czech Republic. Phone: +420 728 80 20 47, E-mail: david.sladek@unob.cz. ORCID 0000-0002-4418-3203.

stations, i.e., when evaluating overall potential and/or limits of crowdsourcing as in the case of some urban areas [4, 5].

Weather monitoring is not only the domain of civilian services and institutions but also of military ones. Such interest was accelerated after the Second World War [6]. From the current point of view, military weather stations should among others often meet the requirement of an enhanced device mobility, which led to the development of so-called field or tactical weather stations (e.g., TACMET [7]). Higher mobility of non-professional stations can serve as a deciding factor when planning military training and operations.

The aim of this paper is (1) to discover the most commonly occurring errors in air temperature and relative humidity measurements of non-professional stations and (2) to identify their source in order to attempt proposing a workable procedure for exploratory analysis. Subsequently, elimination of these errors will take place. An error will be defined as the difference of the measurement insufficient for aviation purposes [8]. Based on the previous studies [3, 4], there is a large portion of shared variability particularly in terms of air temperature reference and tested datasets. However, the remaining unexplained variability has not been analyzed yet. Similarly, relative humidity was not investigated at all either. The significance of this topic lies in the drive to attain optimal data cleaning. The identification of questionable values is a widely discussed subject and serves as the basis for the development of more advanced methodologies, particularly in the field of machine learning [9]. Indeed, the next research goal will be to replace measurements during outages or to automatically detect outliers.

2 Data and Methods

An ideal opportunity for testing is to utilize data from two parallel measurements at the same location and to determine the most common errors in the tested measurement compared to the reference measurement. For this purpose, data from the DAVIS and METEOS6 stations covering the period of 1st JAN 2016-31st DEC 2019 were analyzed and their parameters are shown in Tab. 1. Both stations are located in the military quarters area of Černá Pole in Brno, Czechia. Despite the fact that both stations are able to measure several parameters [3], we focused only on the values of air temperature and relative humidity.

Tab. 1 Basic information about measuring stations [1]

Station	Role	Index	Time step	Exposition	Obstacles	Tested values
METEOS6	reference	_x	10'	330-30	yes	T, RH
DAVIS	tested	_y	5'	330-30	yes	T, RH

While performing the analysis, it must be remembered that conditions in practice will often not be ideal for operational use. Despite this, it will be necessary to investigate the quality of the dataset and measurements to determine its usability independently from initial and background conditions (exposition, obstacles, location, weather conditions...) [10].

The very first step to introduce the problem is an automated description of the basic parameters of both datasets (Tab. 2). The average values of the air temperature and relative humidity differences are relatively low, but both the maximum (5.0 °C for T, 33.9 % for RH respectively) and minimum (-6.4 °C for T, -81.1 % for RH respec-

tively) differences are beyond acceptance for military applications (meteorological support – as 0.5 °C according to [8], 5 % for relative humidity as 75 % percentile of values, as shown in Tab. 2).

At first sight, the very low average error should be appreciated. However, looking at the extremes, the challenge is to map these extremes and discover their causes. The low average error of air temperature values suggests that they will not be of systematic origin but there will be exceptional situations of rare occurrence. In general, relative humidity values seem to have higher variability than air temperature values and as we can see in the Tab. 2, the tested station tends to overestimate relative humidity values.

Tab. 2 Basic statistical description of datasets (x index - reference, y - examined, Δ - difference): standard deviation, 25 and 75 percentiles, 50 percentiles standing for median

Index	T_x [°C]	RH_x [%]	T_y [°C]	RH_y [%]	ΔT [°C]	ΔRH [%]	Abs(ΔT) [°C]	Abs(ΔRH) [%]
Mean	10.4	69.8	10.5	72.7	-0.1	-2.9	0.2	3.5
Std	9.2	18.2	9.1	16.8	0.3	3.0	0.2	2.3
Min	-15.0	0.0	-14.6	18.0	-6.4	-81.1	0.0	0.0
25 %	3.0	56.4	3.2	61.0	-0.2	-4.8	0.1	1.7
50 %	9.8	72.2	9.9	77.0	-0.1	-3.0	0.2	3.2
75 %	17.4	85.0	17.5	87.0	0.0	-1.0	0.3	4.9
Max	35.8	100.0	36.9	99.0	5.0	33.9	6.4	81.1

In order to study the datasets, following questions were set out in relation to the measured dataset overview:

- At what air temperatures and relative humidities do the largest differences occur?
- Does the magnitude of the difference depend on the trend from previous measurements? I.e. when a variable rises/declines rapidly – do the sensors respond the same way?
- Both temperature and relative humidity do not markedly differ on average – can any timely (daily, monthly) pattern be found at the extremes of the difference in measured air temperature and relative humidity values?
- Are there any major differences from previous findings available solely for air temperature datasets during shorter time period [3]?

3 Results

First of all, overall datasets characteristics, number of observations and missing values of air temperature and relative humidity measurements are outlined in the following table (Tab. 3). The proportion of outages is slightly higher for air temperature at both stations. Technical issues or maintenance of the sensors according to regular calibration procedures can cause these.

The course of the differences can be seen in Fig. 1. At a very least, it will give us an overall view of the dataset and help us define the main questions for analysis. We could – although prematurely – infer the effect of seasonality there.

Tab. 3 Total number of observations by reference and tested stations and their outages (reference – METEOS6, tested – DAVIS)

	Max terms	Measured/ Outage T	Measured / Outage RH	Unmeasured [%]	Outage T [%]	Outage RH [%]
METEOS6	210 384	197 917 / 5 607	197 916 / 5 608	3.2	2.8	2.8
DAVIS	420 768	314 567 / 9 500	315 667 / 8 400	22.9	3.0	2.6

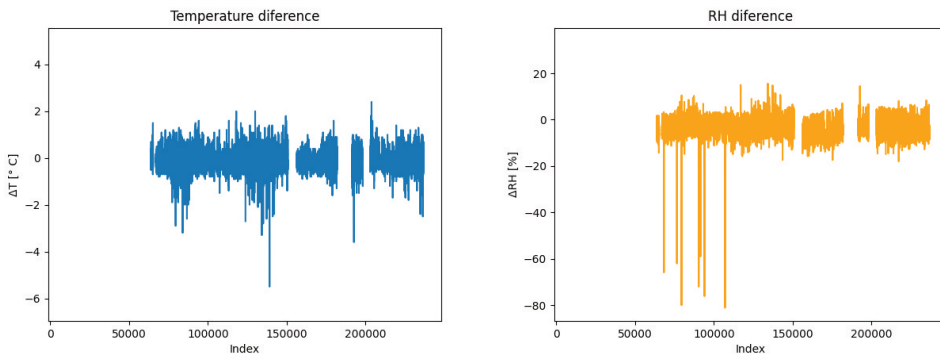


Fig. 1 Measured air temperature difference (left) and relative humidity difference (right) with obvious measurement outages and extreme values; differences are counted as reference – tested values

According to Fig. 1 and Tab. 3, we can confirm that the sensors were experiencing outages, which in itself is a negative indicator of the quality of the measurement. On the other hand, there are also large fluctuations in the differences in measured temperatures, which could seem to indicate some seasonality issues or serious interference from non-systematic factors.

Fig. 2 shows not only the structure of the datasets with typical frequency distribution for air temperature and relative humidity, but also corresponding absolute differences. The absolute error (difference) distribution (last two graphs) shows that the tested one is a very accurate measurement and most of the errors are within an acceptable (concerning 0.5 °C and 5 % difference) range of values.

The distribution indicates that these are more likely to be isolated errors that would need to be diagnosed. Furthermore, when considering the distribution of relative humidity differences with the same number of bins, an apparent inconsistency is observed, potentially attributed to variations in rounding. It is important to note that the measurement step of the test station is 1 % RH, and this factor should be duly considered when analyzing the data. Thus, rounding characteristic distribution was also examined (Fig. 3).

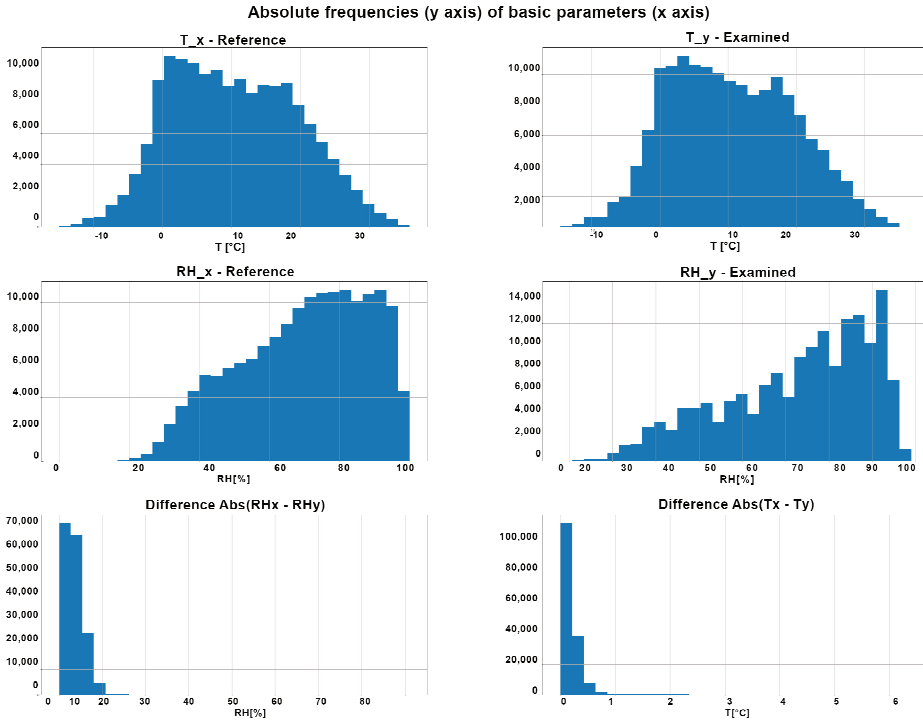


Fig. 2 Basic exploratory analysis of distribution of air temperature (reference – left, examined – right), relative humidity and absolute difference values of temperature and relative humidity

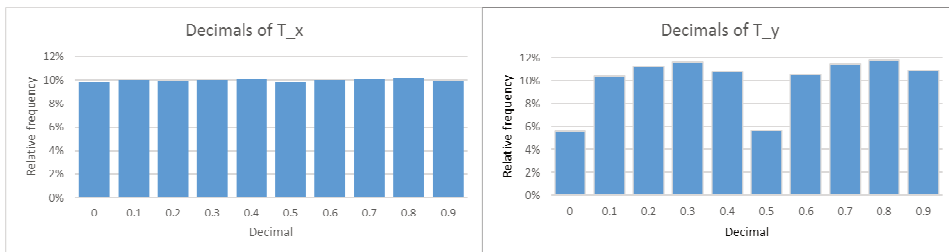


Fig. 3 Distribution of decimals occurrence by measured air temperatures on reference station (left), and tested one (right)

As far as temperature rounding is concerned, the differences at the reference station are not very high. However, for the tested one it is clear that the values 0.0 °C and 0.5 °C are under-represented. Although this finding may not directly affect user requirements, it still can be indicative of a different approach to processing measurement, as it can contribute to the differences in measured values. Interestingly, under-represented values of 0.0 °C and 0.5 °C are not compensated for by increased frequency in neighboring values as expected. The highest frequency fits the 0.3 °C and 0.8 °C decimal values. A similar result for the reference station (Fig. 4) can be derived

from rounding the relative humidity. For operational purposes, rounding has no consequence, but it does indicate some difference between manners of creating two examined datasets.

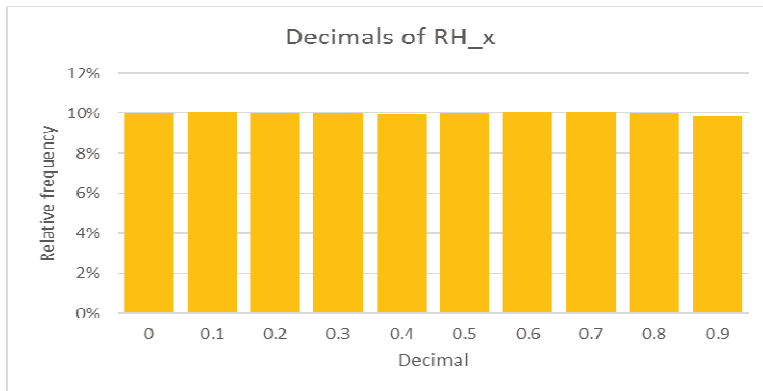


Fig. 4 Distribution of decimals occurrence by measured relative humidities on reference station. Tested station utilizes whole percent units of relative humidity.

The following graphs (Fig. 5) were created to identify the issues and overview the general pattern of measurement errors. These show the overall pattern of temperature and measurement differences from four random fractions of the measured time series.

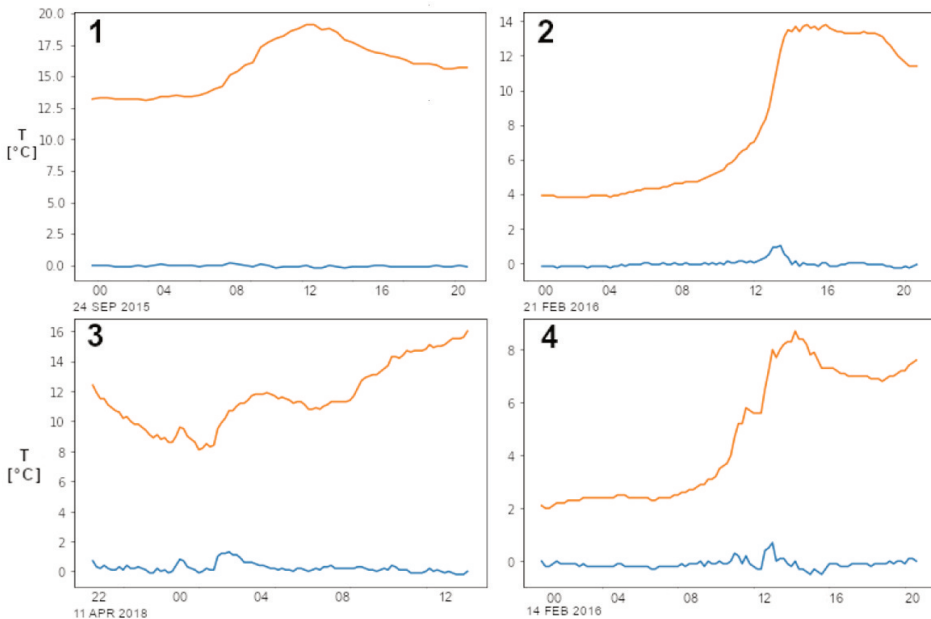


Fig. 5 Random cuts of air temperature profile (orange) and corresponding difference (blue) of measured air temperature in random 20 hours intervals (dates in the lower left corner); differences are counted as reference – tested values

Fig. 5 illustrates that the sensor might react less successfully (afterwards) to situations where the temperature changed suddenly compared to the assumptions from the diurnal cycle. In ordinary situations (such as shown in section 1 in Fig. 5), when the temperature changes continuously, the error oscillates only among very low values. The section 1 shows typical diurnal cycle with one peak of maximum temperature with no unexpected break. Section 2 corresponds to a rapid late winter temperature rise, where the curve is slightly flattened at the time of maximum temperature and thus does not reach the possible predicted maximum temperatures. Thus, the period of sharp air temperature rise is prone to larger differences. The other two sections probably indicate the effect of cloud cover or radiation reaching the two sensors. When clouds suddenly dissipate or form, the daily temperature course is disturbed and consequently, a certain break in the curve is created, which is not only clearly visible, but it also causes a higher error value. Similar effect can be attributed to highly variable pattern of object's shading (buildings, trees etc.) during sunny days mostly in summer. Although the horizontal distance between sensors is in such cases relatively low, it can result in higher differences. In all cases depicted, the difference was still within a very small range. However, its trend shows how the sensor reacts to the change.

In order to make an initial description complete, Pearson's correlation coefficients were calculated for measured elements (Tab. 4).

Tab. 4 Correlation of measured elements (x index – reference, y index– tested)

	T_x	RH_x	T_y	RH_y
T_x	1.00	-0.59	1.00	-0.62
RH_x	-0.59	1.00	-0.57	0.99
T_y	1.00	-0.57	1.00	-0.62
RH_y	-0.62	0.99	-0.62	1.00

It is obvious that the elements are perfectly correlated with each other on the main diagonal. Indicating high accuracy of the dataset, both temperature and relative humidity values show very high correlation coefficients on both sensors with higher values for air temperature cases. It can be claimed that in this basic exploratory analysis, datasets look remarkably similar. Therefore, extreme values of errors will be examined.

3.1 At What Air Temperatures and Relative Humidities Do the Largest Errors Occur?

From the initial distribution plots, we cannot determine exactly where the differences in the measured values stem from. In this context, error categories have been created. Differences of 0.5 °C and above, which are considered significant in aviation meteorology [8], have been defined as high errors.

The difference in measured air temperatures (both the maximum values and their seasonality) could be anticipated. This might lead to the idea that the higher deviations are at high or low temperatures, i.e. in summer or winter.

Comparison of the histograms (Fig. 6) shows that the errors are in general slightly more concentrated at higher measured temperatures (20 °C and more) and disproportion emerges when values dropped under approximately -5 °C. The producer of tested station states that the measurement error increases at temperatures below -7 °C. In particular, the curve quite smoothly follows the distribution of all tempera-

tures, suggesting that the tested sensor operates reliably in the range of the most common air temperature values.

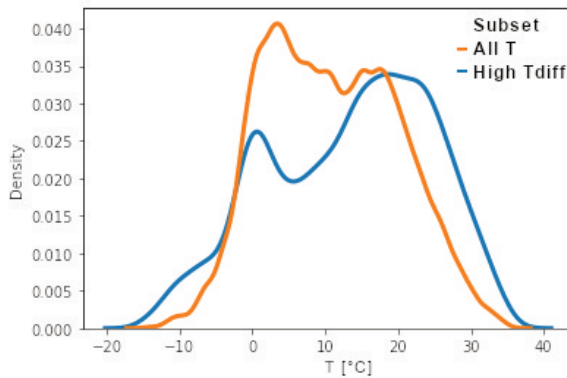


Fig. 6 Histogram comparison of air temperature distribution at high values of absolute T difference (> 0.5 °C, blue) and distribution of all measured reference air temperatures (orange)

Generally, inspecting a difference curve of air temperature, it can be claimed that the distribution follows the distribution of all measured air temperatures with a slight shift towards higher air temperatures. Therefore, let us inspect the correlation matrix of measured values and calculated features dataset in Tab. 5.

Tab. 5 Correlation coefficient of measured values and calculated features

	ΔT	Abs(ΔT)	ΔRH	Abs(ΔRH)
T_x	0.17	0.01	0.03	-0.12
RH_x	-0.05	-0.07	0.53	-0.38
T_y	0.14	0.02	0.04	-0.13
RH_y	0.01	-0.10	0.39	-0.27
ΔT	1.00	-0.31	-0.37	0.21
Abs(ΔT)	-0.31	1.00	0.09	0.13
ΔRH	-0.37	0.09	1.00	-0.78
Abs(ΔRH)	0.21	0.13	-0.78	1.00

It can be concluded from Tab. 5, in the case of temperature measurement error, there is no significant value of Pearson correlation coefficient, which does not help to draw the conclusion suggested by the previous histogram. On the other hand, the tendency of larger differences at higher air temperatures fit for both stations. The higher values of the coefficients in case of the relative humidity differences (more pronounced for reference station METEOS6) and corresponding station values give further scope for visualizing the relationship between the two variables.

Firstly, the distribution of the high errors (absolute difference greater than 5 %) was compared with the distribution of all relative humidity measurements (Fig. 7).

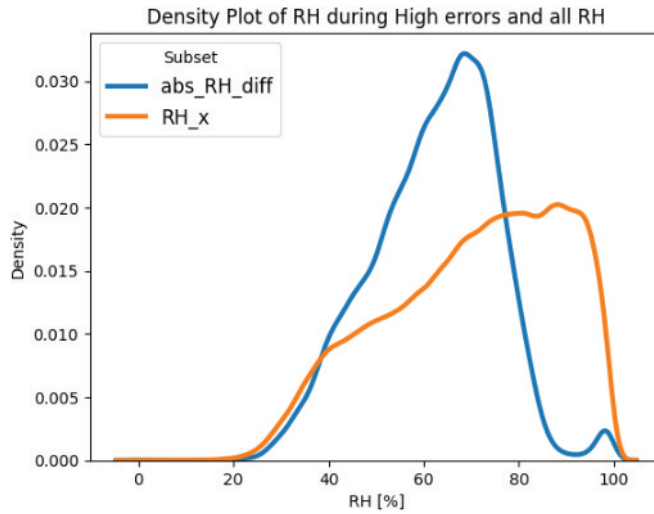


Fig. 7 Histogram comparison of high difference distribution (blue), and all measured values of relative humidity (orange)

However, the value of the correlation coefficient between the measured humidities and the RH measurement difference may be of significant importance. It can be anticipated that there is a possibility that at higher relative humidities (especially at the reference sensor METEOS6) there is a higher measurement error. The graph below (Fig. 8) shows the known value of the correlation coefficient ($r = 0.53$). The measurement difference increases with increasing relative humidity values, while the absolute value decreases slightly ($r = -0.38$). Above the value of approximately 90 %, the values are very similar. It can be concluded that the magnitude of the relative humidity difference is dependent on the measured relative humidity. At higher values, the tested sensor measures generally lower values, and it happens very often. This follows from the fact that the values of absolute difference and ordinary difference are identical. Hence, there are few negative values. Conversely, at lower relative humidities, negative values are more frequent, i.e. the tested sensor tends to measure rather higher values. Such pattern has also resulted in lower overall variability of relative humidity measurements at tested DAVIS station (see Tab. 2).

3.2 Does the Magnitude of the Error Depend on the Trend from Previous Measurements?

It can be assumed that high error values can occur after sudden changes in temperature or relative humidity. Examples of such situations are atmospheric front passing or sudden cloud formation or dissolution, etc. These usually led to significant changes of the normal daily radiation pattern.

In order to compare the measurement errors and their dependence on previous values, the values of the changes in the last 1-5 measurements (10-50 minutes) were calculated. The correlation coefficients with the individual changes (autocorrelation) in the last 1-9 measurements are shown in Fig. 9.

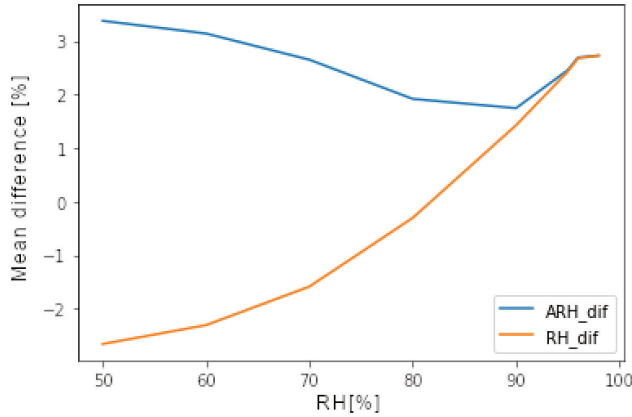


Fig. 8 Dependence of simple (orange) and absolute (blue) relative humidity differences on relative humidity measured by a reference sensor

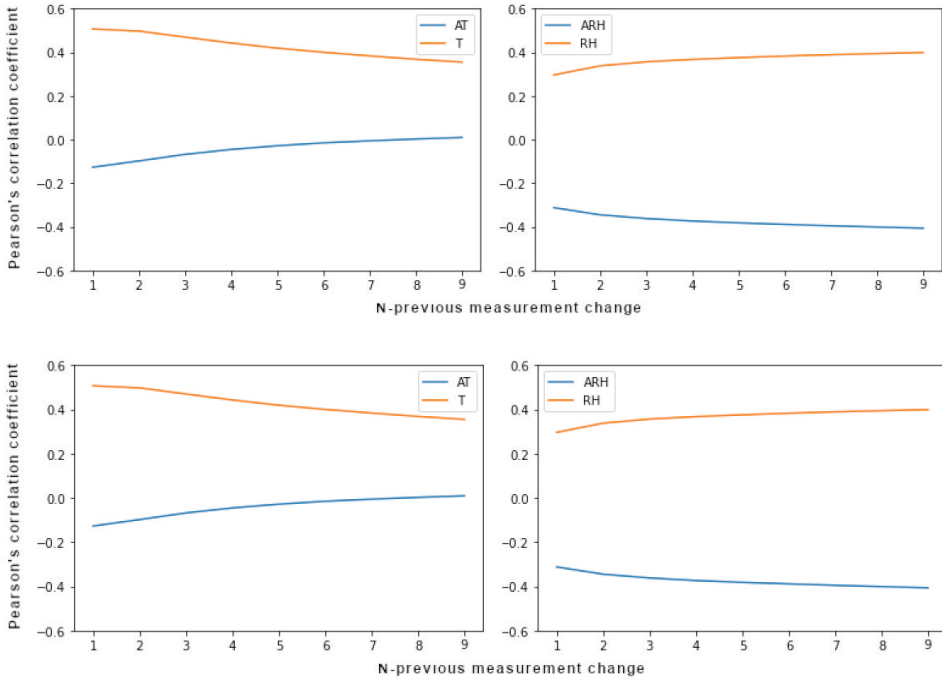


Fig. 9 Autocorrelation of air temperature (left) and relative humidity (right) change in N recent measurements and corresponding differences (absolute difference – blue, simple difference – orange)

Figure 9 reveals that in the case of air temperature, the difference in the first previous measured value (correlation 0.51) determines the value of the error the most. Thereafter, for the following values the correlation coefficient drops lower. For the absolute value (in blue on the graph), the value of the coefficient is insignificant for

this dataset. In case of relative humidity (right), this is more a case of an earlier measurement, but here with lower correlation values (app. 0.30) without obvious trend.

In order to find potential connection between changes in last hour, rolling mean of air temperature was calculated and compared with the absolute difference (Fig. 10).

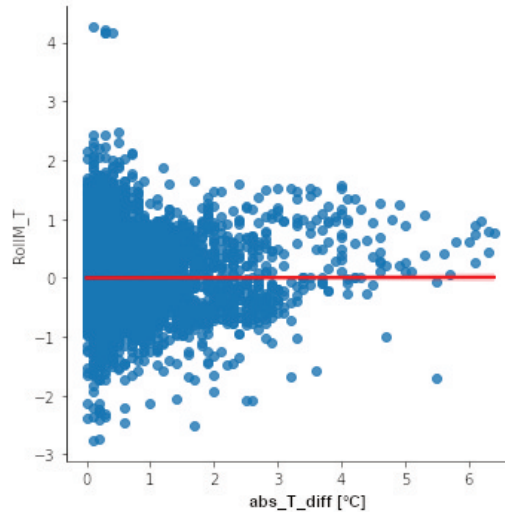


Fig. 10 Rolling mean of five previous measurements dependence on absolute temperature difference (trend shown in red)

According to the values shown, it can be said that in general, higher values of error (the higher the error, the further to the right) are concentrated above the mean at higher values of the rolling mean. However, this finding cannot be used in an absolute way because of a few values that could be confidently called outliers. In fact, there have been cases where air temperature jumped by 4 °C but the error in the measurement was minimal (upper left corner of the Fig. 10).

Since no indication has yet been found that a clear relationship can be found between the magnitude of the error and the directly measured values, let us look at the relationship of the error on the graph. It could show seasonality or some other pattern of error propagation. The difference does increase in bumps around 3-5 (Fig. 11), but no pattern can be found. One must also consider the fact that the waveform is strongly influenced by a small number of measured errors in the higher range (value of absolute difference 0.7 °C is within the 0.98 quantile).

As there were only unclear signs of connections between differences and other values, other major feature of time series should be inspected. Analysis of time characteristics of an error can lead to discovery of seasonal patterns in measurements differences.

3.3 Can There Be Found Any Timely (Daily, Monthly) Pattern at the Extremes of the Difference in Measured Air Temperatures and Relative Humidities?

From the previous analysis it is evident that there is some accumulation of biases in both air temperature and relative humidity, but only around a certain interval of values. However, it has not been established on what this value depends, thus it is

appropriate to try to investigate the time course of the differences in the measured quantities.

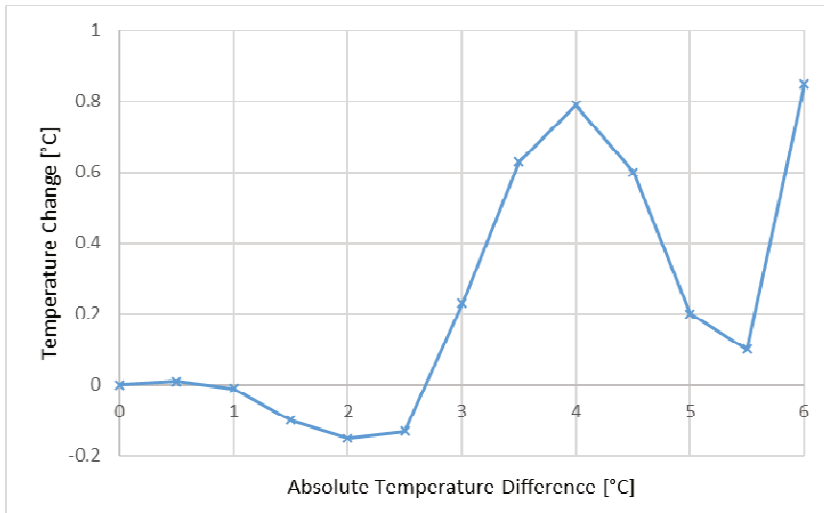


Fig. 11 Dependence of air temperature change in the last hour and absolute difference (right on x axis the greatest value of difference)

Since the main cycles of air temperature are diurnal and annual, we shall concentrate on these.

The diurnal cycle of the air temperature error (Fig. 12) shows a clear two-peak course, which is concentrated in the period (1) around sunrise in the mornings and (2) during maximum air temperatures in the afternoons. Unfortunately, this overall graph smooths the patterns of individual months, which are presented further.

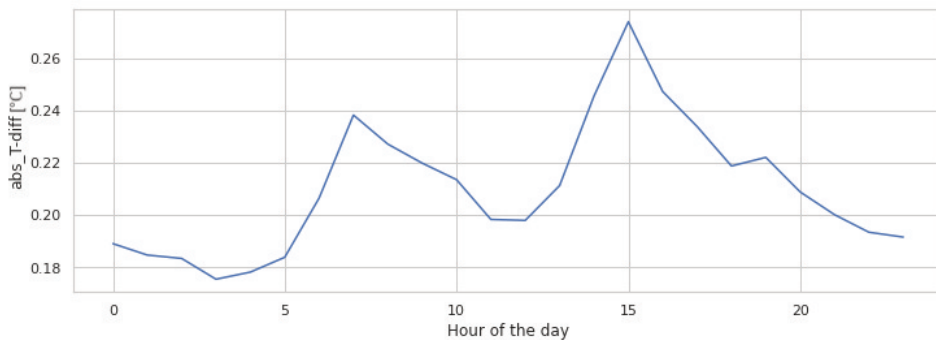


Fig. 12 Mean diurnal (UTC time) cycle of the absolute difference of air temperature

Monthly comparison of diurnal cycle (Fig. 13) of the difference magnitude confirm the above-mentioned with few exceptions. The two-peak course is obviously prominent by two months – May and June, followed by July. By other months, the bias cycle is not that remarkable. Moreover, individual months differ in timing of the largest daily differences. This confirms hypothesis that main bias can be caused by

different conditions initiated by location of sensors, although mutual horizontal distance is only a few meters.

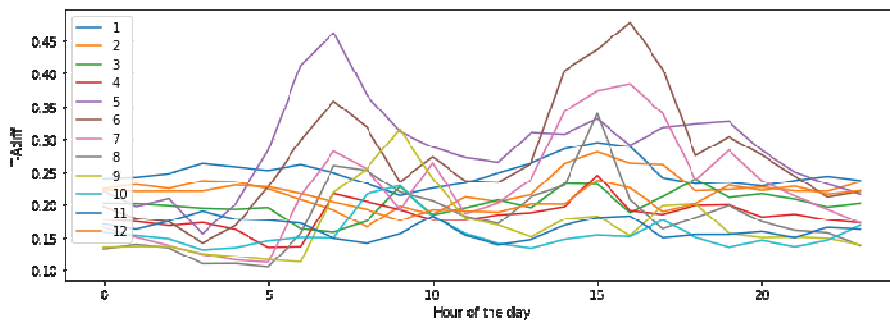


Fig. 13 Diurnal cycle of absolute air temperature differences by month

Since a simple difference of air temperature did not present any odds by months, it is not displayed. Next figure (Fig. 14) shows diurnal cycle of absolute relative humidity differences. It can be stated that the magnitude of error is generally greater in winter, but the course of value is very similar by all of the months. Thus, the statement that the month of the year has a negligible effect on relative humidity appears to be correct. The course of simple difference is basically reversed chart of the magnitude, with the lowest differences about 10 h as low as -6% RH and highest in 0-5 h, as high as 0% in September. As simple difference is usually lower, it also confirms the previously concluded observation that the tested sensor measures rather higher relative humidity values.

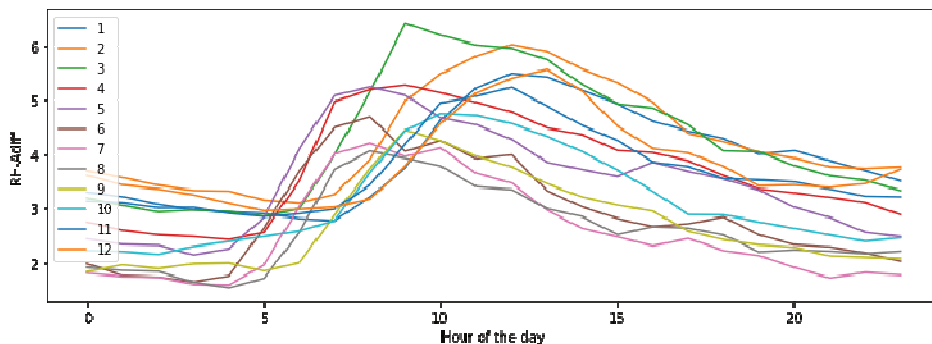


Fig. 14 Diurnal cycle of absolute relative humidity differences by month

4 Discussion and Conclusion

The aim of the research was to find out whether a non-professional station, whose measurements are continuously recorded, can be used for densification of the station network or for operational use for various tasks.

The results showed that the measurements can meet even the strict standards required in aviation meteorology. The trend analyses of previous measurements and daily runs performed lead to major research outputs:

- the tested weather station itself measures relatively reliably,

- the dataset is very accurate and significant errors occur only in approximately 0.1-0.2 % of the measurements,
- in case of large inaccuracies, the ambient values can be manually checked, as high error values are often isolated within the time series (as shown in Fig. 15). Algorithms and methods to find outliers could be designed later specifically for the DAVIS instrument. In addition, possible outages in measurements should be noticed,

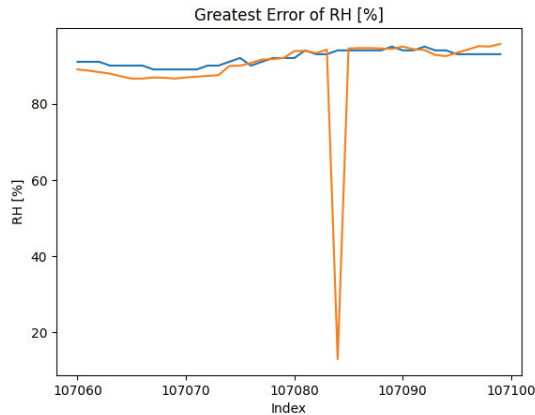


Fig. 15 Time series of humidity measurements with greatest error

- the location of the station seems to have the greatest influence, due to which temporally systematic errors in temperature measurements have been registered,
- in the case of relative humidity, the sensor measures reliably even at high values, close to saturation, which promises a possible application in meteorology,
- unfortunately, in the case of temperature measurements, we register a higher error increase around 0 °C, which can be a concern for many applications where this value is critical,
- for practical use, a suitable radiation shield must be used and the standard guidelines (e.g. WMO [10]) for the placement of weather stations must be followed, and it can be used for routine measurements, taking into account the limits given by the manufacturer.

Regarding previous study covering shorter time and slightly different meteorological elements [3] we agree on the overestimation of measured air temperatures at the tested station DAVIS. Nevertheless, in our study, the overall variability in air temperature was slightly higher ($r = 1.00$) compared to the previous study ($r = 0.95$). This difference can be attributed to the shorter time period considered, which led to increased variability in spring radiation at the specific station location.

The presented study will serve as a basis for further research in the area. The aim is to map the possibilities of creating an algorithm for automatic outlier detection in measurements, to check the quality of measured data in general, the suitability of measurements for use in machine learning methods, etc.

Acknowledgement

This work was supported by the Project for the Development of the Organization “DZRO Military Autonomous and Robotic Assets”.

References

- [1] BELL, S., D. CORNFORD and L. BASTIN. The State of Automated Amateur Weather Observations. *Weather*, 2013, **68**(2), pp. 36-41. DOI 10.1002/wea.1980.
- [2] WOLTERS, D. and T. BRANDSMA. Estimating the Urban Heat Island in Residential Areas in the Netherlands Using Observations by Weather Amateurs. *Journal of Applied Meteorology and Climatology*, 2012, **51**(4), pp. 711-721. DOI 10.1175/JAMC-D-11-0135.1.
- [3] DEJMAL, K., F. HUDEC, P. KOLAR and J. NOVOTNY. Evaluation of Measurement Quality of Selected Elements on the Meteorological Stations Meteos6 and Davis Vantage Pro 2 in the Military Quarters Area of Černá Pole. In: *2017 International Conference on Military Technologies (ICMT)*. Brno: IEEE, 2017, 2017, pp. 318-324. DOI 10.1109/MILTECHS.2017.7988777.
- [4] DEJMAL, K., P. KOLAR, J. NOVOTNY and A. ROUBALOVA. The Potential of Utilizing Air Temperature Datasets from Non-Professional Meteorological Stations in Brno and Surrounding Area. *Sensors*, 2019, **19**(19), 4172. DOI 10.3390/s19194172.
- [5] CHAPMAN, L., C. BELL and S. BELL. Can the Crowdsourcing Data Paradigm Take Atmospheric Science to a New Level? A Case Study of the Urban Heat Island of London Quantified Using Netatmo Weather Stations. *International Journal of Climatology*. 2017, **37**(9), pp. 3597-3605. DOI 10.1002/joc.4940.
- [6] PINCUS, R. ‘To Prostitute the Elements’: Weather Control and Weaponisation by US Department of Defense. *War and Society*, 2017, **36**(1), pp. 64-80. DOI 10.1080/07292473.2017.1295539.
- [7] KOKKO, H. Versatile Automated Weather Observation for Demanding Tactical Military [online]. *VAISALA News*, 2001, **155**, pp. 14-17 [viewed 2022-10-02]. Available from: https://www.vaisala.com/sites/default/files/documents/VN155_Demanding_Tactical_Military_Needs.PDF
- [8] *Meteorological Service for International Air Navigation. Annex 3 to the Convention on International Civil Aviation* [online]. Montreal: International Civil Aviation Organization, 2010. Available from: <https://www.icao.int/airnavigation/IMP/Documents/Annex%203%20-%2075.pdf>
- [9] YAMAMOTO, K., T. TOGAMI, N. YAMAGUCHI and S. NINOMIYA. Machine Learning-Based Calibration of Low-Cost Air Temperature Sensors Using Environmental Data. *Sensors*, 2017, **17**(6), 1290. DOI 10.3390/s17061290.
- [10] *Guide to Instruments and Methods of Observation (WMO-No. 8). Volume I: Measurement of Meteorological Variables* [online]. Geneva: WMO, 2021. [viewed 2022-03-09]. ISBN 978-92-63-10008-5. Available from: https://library.wmo.int/viewer/41650/download?file=8_I_2021_en.pdf&type=pdf&navigator=1)