Determination of Step Check Quality Control Thresholds on Air Temperature Data at South Tangerang Climatological Station

M Halida¹,² and SA Pramono²

¹ Tangerang Selatan Climatological Station, Jl. Raya Kodam Bintaro No. 82
Tangerang Selatan 15221 Indonesia
² BMKG Database Centre, Jl. Angkasa I No. 2 Jakarta Pusat 10610 Indonesia

*mutiara.klimat@gmail.com

Abstract. All data, including air temperature data, must be verified by conducting quality control using the step check method. Step check quality control is carried out by looking at the difference of a parameter in a certain period compared to the threshold value that was already determined. Therefore before carrying out step check quality control, it is necessary to determine the ceiling and floor boundaries of the difference in air temperature data every hour. The data used in this study are hourly air temperature data and hourly present weather data from weather observations at the South Tangerang Climatological Station during 2016 - 2020. In determining the threshold for air temperature step check quality control, the air temperature data is paired with weather condition data to obtain a threshold value according to rain and no rain conditions. The threshold conducted in this study is based on a check for unusual climatological values, where the limits for an unusual and impossible jump in hourly air temperature changes are determined based on a certain percentage of the data distribution. This study uses percentile analysis to determine the threshold, where 5% in the lower and upper part of the data distribution are used as the threshold. The results show various thresholds every hour. The increase in temperature dominates the changes of hourly air temperature in no-rain conditions. The highest threshold for temperature increase occurs at 00.00 – 01.00 UTC at 3.2°C and continues to decrease over time. The highest threshold for temperature decrease occurs at 09.00 UTC - 10.00 UTC at 2.2°C. In rain conditions, the increase in temperature can still occur. However, the decrease in temperature mainly occurs. The highest threshold for temperature decrease during rainy conditions is 1.8°C at 01.00 - 02.00 UTC, while the highest threshold for the temperature decrease is 5.8°C at 06.00 UTC – 07.00 UTC. With these results, observers can first carry out quality control with the Step Check method before filling in the data into the system database. Thus, any suspect data either from reading errors or tool errors can be minimized and finally produce a valid dataset.
1. Introduction
Weather observation is one of the essential parts of climatological studies in general. Research using weather data results in broad and in-depth climatological information. However, researchers often encounter a problem in conducting their research, one of which is the invalidity of weather data [1]. Weather observations in Indonesia are carried out every hour, with various weather elements being observed, including air temperature. Air temperature is one of the most critical weather elements because the impact of changes in air temperature can be felt directly by the human senses. Air temperature observations are carried out at various altitudes depending on their needs. However, air temperature observations at an altitude below 2 (two) are the most representative of the climatic conditions of a region related to the issue of global warming [2].

Warmest and coldest air temperature is influenced by the circulation of the atmosphere in the air and the intensity of solar radiation. More in-depth analysis related to atmospheric circulation and the intensity of solar radiation is constructive in making a hypothesis of extreme temperatures in a region [3].

Quality control on data is essential to obtain a valid dataset. By carrying out quality control, data with suspicious values and potential errors can be identified and flagged immediately, so it can be considered before carrying out further data processing. Quality control can be done in real-time at the station level and carried out directly by a weather observer immediately after observing. One method of real-time quality control is the step check method. In the step check method, the air temperature data at a certain period is compared with the air temperature data in the previous record. Suppose there is a change in air temperature that exceeds the specified temperature change threshold. In that case, the data needs to be flagged and validated. The calculation to obtain a change threshold value used as a reference in carrying out quality control using the step check method is necessary. The temperature change threshold varies from place to place, depending on the climatic characteristics of the region [4-10].

This study aims to determine the threshold for hourly temperature changes at the South Tangerang Climatological Station, which can be used as a reference in conducting early-stage quality control of air temperature observation data before the data is entered into the database system. Thus any suspicious data either from reading errors or tool errors can be minimized and produce valid air temperature data that can be processed into further information.

2. Data and Methodology
The data used to determine the ceiling and floor boundaries or the threshold of the hourly air temperature changes is hourly air temperature data in Tangerang Selatan Climatological Station from 2016 to 2020. The air temperatures are observed for 13 hours every day, starting from 00.00 UTC to 12.00 UTC, resulting in 23,751 data. Since air temperature is closely related to weather conditions (e.g. rain and no rain condition), the weather condition from present weather data in synoptic observation is included, then paired with hourly air temperature data to obtain two sets of hourly air temperature changes in the rain and no rain conditions.

Step check quality control is carried out by comparing the changes of two data in a specific time range. In this study, we compare the changes of air temperature data every hour, for example, air temperature data at 02.00 UTC being subtracted with air temperature data at 01.00 UTC, and so on. The formula for this calculation is shown in equation 1.

$$\Delta T = T_n - T_{n-1}$$

Where $\Delta T$ is the changes in air temperature every hour, $T$ is temperature, and $n$ is period.

A positive value indicates that the air temperature has increased, and a negative value indicates that the air temperature has decreased. From the calculation above, 12 air temperature differences data were obtained in 1 day resulting in 21,924 (100%) data in total to be used in this study. Before calculating the difference per hour, the completeness data is checked, resulting in 368 (2%) blank air temperature
data. These blank data cannot be used to calculate the difference in air temperature data every hour, so it is excluded to analyze this study further.

After calculating the difference of hourly air temperature data both in the rain and no rain conditions, the next step is to determine the threshold for Step Check Quality Control. The threshold conducted in this study is based on a check for unusual climatological values (probable error), where the limits for an unusual and impossible jump in hourly air temperature changes are determined based on a certain percentage of the data distribution [11]. This study uses percentile analysis to determine the threshold, where 5% in the lower and upper part of the data distribution are used as the threshold. Therefore percentile-5th is used as the floor boundary while percentile-95th is used as the ceiling boundary [12-13]. If there is a change in hourly air temperature that exceeds these thresholds, the data must be flagged, and further inspections are needed to see if there are errors in readings or errors in the instrument.

3. Results and Discussions

The pairing process of air temperature data with weather condition data is carried out by adjusting weather condition data to air temperature data. This pairing process must be done because the air temperature data is not 100% available (table 1).

| Table 1. Data availability check. |
|-----------------------------------|
| Availability | Air Temperature Data | Present Weather Data | Paired Data |
|----------------|----------------------|----------------------|-------------|
| Available      | 98 %                 | 100 %                | 98 %        |
| Unavailable    | 2 %                  | 0 %                  | 2 %         |

By separating air temperature data based on rain and no rain weather conditions, 19.067 hours of air temperature data with no rain (88%) and 2,489 hours of air temperature data with rain (12%) were obtained (table 2).

| Table 2. Data availability in rain and no rain conditions. |
|-----------------------------------------------|
| Weather Conditions | Air Temperature Data (in hours) | Air Temperature Data (in percent) |
| No Rain          | 19.067                          | 88%                               |
| Rain             | 2,489                           | 12%                               |

3.1 No Rain Conditions

The highest increase of hourly air temperature data in no rain conditions is 6.4°C, while the highest decrease is 6.0°C. From table 3, extreme analysis using percentile-95th and percentile-5th for all data shows the highest increase for hourly air temperature is 2.2°C. In comparison, the highest decrease is 1.4°C, which means that increasing and decreasing outside this value are considered as a rarely occurring event.

| Table 3. Percentile threshold of hourly air temperature changes in no rain conditions. |
|-----------------------------------------------|
| No Rain Conditions | Hourly Air Temperature Changes |
|-------------------|--------------------------------|
| Percentile-5th    | > 2.2 °C                      |
| Normal            | -1.4 – 2.2 °C                 |
| Percentile-95th   | < -1.4 °C                     |
The frequency of changes in air temperature in no rain conditions is dominated by an increase in air temperature of 57.0% or 10,863 hours, while changes in air temperature in the form of a decrease are 37.9% or 7,224 hours. The condition of the air temperature did not change is 5.1% or 980 hours (Table 4).

Table 4. Frequency of changes in air temperature in no rain conditions.

| Range ( °C) | Frequency | %    |
|------------|-----------|------|
| < -6.0     | 1         |      |
| -5.1 - -6.0| 4         |      |
| -4.1 - -5.0| 17        |      |
| -3.1 - -4.0| 46        | 37.9%|
| -2.1 - -3.0| 160       |      |
| -1.1 - -2.0| 1400      |      |
| -0.1 - -1.0| 5596      |      |
| 0          | 980       | 5.1% |
| 0.1 - 1.0  | 5464      |      |
| 1.1 - 2.0  | 4117      |      |
| 2.1 - 3.0  | 1114      |      |
| 3.1 - 4.0  | 155       |      |
| 4.1 - 5.0  | 11        | 57.0%|
| 5.1 - 6.0  | 0         |      |
| > 6.0      | 2         |      |

The changes in air temperature in detail for every hour are also conducted to represent the natural conditions because air temperature fluctuations are also greatly influenced by solar radiation. From Table 5, it can be seen that changes in air temperature at 00.00 - 01.00 UTC have the highest increase. An increase in air temperature of 3.1°C at this hour is still considered an everyday event because the extreme limit is 3.2°C.

Table 5. Percentile threshold and mean of hourly air temperature changes in no rain condition.

| Hours (UTC)   | Mean of Hourly Air Temperature Changes ( °C) | Percentile-5th ( °C) | Percentile-95th ( °C) |
|--------------|---------------------------------------------|-----------------------|-----------------------|
| 00.00 – 01.00| 1.7                                         | 0.4                   | 3.2                   |
| 01.00 – 02.00| 1.8                                         | 0.6                   | 3.0                   |
| 02.00 – 03.00| 1.3                                         | 0.4                   | 2.2                   |
| 03.00 – 04.00| 1.2                                         | 0.0                   | 2.2                   |
| 04.00 – 05.00| 0.7                                         | -0.4                  | 1.6                   |
| 05.00 – 06.00| 0.3                                         | -0.6                  | 1.3                   |
| 06.00 – 07.00| 0.1                                         | -1.2                  | 1.2                   |
| 07.00 – 08.00| -0.3                                        | -1.6                  | 1.0                   |
| 08.00 – 09.00| -0.5                                        | -1.6                  | 0.4                   |
| 09.00 – 10.00| -1.0                                        | -2.2                  | 0.2                   |
| 10.00 – 11.00| -0.7                                        | -1.6                  | 0.0                   |
| 11.00 – 12.00| -0.7                                        | -1.6                  | 0.0                   |

The following analysis sees the hourly air temperature changes in general in boxplot and dot plot histogram (figure 1 - 2). From this graph, we can identify outliers and see the range of the changes in hourly air temperature. The lowest range of air temperature changes occurs at 04.00 - 05.00 UTC with the highest decrease of 2.8°C and the highest increase of 3.8°C.
3.2 Rain Conditions
The result shows that changes in hourly air temperature are dominated by the drop of air temperature, with the highest decrease is 11.2°C, and the highest increase is 4.2°C. In calculation using all data, the percentile-5th has a high value of decrease of 4.6°C. Percentile-95th calculation shows that the increase of hourly air temperature below 0.9°C is still considered a common event (table 6).

| Table 6. Percentile threshold of hourly air temperature changes in rain conditions. |
|---------------------------------|---------------------------------|
| No Rain Conditions              | Hourly Air Temperature Changes  |
| Percentile-5th                  | < -4.6 °C                       |
| Normal                          | -4.5 – 0.9 °C                   |
| Percentile-95th                 | > 1.0 °C                        |
In rainy weather conditions, changes in air temperature are dominated by a decrease in air temperature of 64.6% or 1,608 hours. The air temperature changes that have increased are 27.4% or 682 hours, and data that does not experience an increase or decrease in air temperature during rainy conditions is 8% or 199 hours (table 7).

| Table 7. Frequency of changes in air temperature in rain conditions. |
|---------------------------------------------------------------|
| Range (°C) | Frequency | % |
| < -9.0 | 4 | |
| -8.1 - -9.0 | 5 | |
| -7.1 - -8.0 | 16 | |
| -6.1 - -7.0 | 25 | |
| -5.1 - -6.0 | 44 | |
| -4.1 - -5.0 | 85 | 64.6% |
| -3.1 - -4.0 | 128 | |
| -2.1 - -3.0 | 214 | |
| -1.1 - -2.0 | 321 | |
| -0.1 - -1.0 | 766 | |
| 0 | 199 | 8.0% |
| 0.1 - 1.0 | 569 | |
| 1.1 - 2.0 | 95 | |
| 2.1 - 3.0 | 14 | |
| 3.1 - 4.0 | 3 | 27.4% |
| 4.1 - 5.0 | 1 | |
| > 5.0 | 0 | |

Analysis of hourly changes in air temperature shows a more representative and more detailed value in seeing the fluctuation. In rain conditions, the average temperature change at 00.00 UTC – 03.00 UTC shows an increasing move, followed by the decline starting from 04.00 UTC to 12.00 UTC. Percentile-5th shows a decline that varies from 0.6°C to 5.8°C while percentile-95th indicates a positive value between 0.4°C and 1.8°C. Table 8 shows that rainy weather conditions are not absolute in decreasing air temperature.

| Table 8. Percentile threshold and mean of hourly air temperature changes in rain condition. |
|---------------------------------------------------------------|
| Hours (UTC) | Mean of Hourly Air Temperature Changes (°C) | Percentile-5th (°C) | Percentile-95th (°C) |
|---------------------------------------------------------------|
| 00.00 – 01.00 | 0.4 | -0.6 | 1.3 |
| 01.00 – 02.00 | 0.6 | -0.6 | 1.8 |
| 02.00 – 03.00 | 0.3 | -1.4 | 1.6 |
| 03.00 – 04.00 | -0.2 | -3.4 | 1.6 |
| 04.00 – 05.00 | -0.7 | -3.5 | 1.1 |
| 05.00 – 06.00 | -1.2 | -5.2 | 1.2 |
| 06.00 – 07.00 | -1.5 | -5.8 | 1.4 |
| 07.00 – 08.00 | -1.4 | -5.0 | 0.8 |
| 08.00 – 09.00 | -1.3 | -5.1 | 0.8 |
| 09.00 – 10.00 | -1.5 | -5.5 | 0.6 |
| 10.00 – 11.00 | -1.0 | -4.4 | 0.4 |
| 11.00 – 12.00 | -0.9 | -3.4 | 0.4 |
According to the boxplot and dot plot (figure 3 – 4), the broadest range of temperature changes occurs at 06.00 - 07.00 UTC for -9.2°C to 4.2°C, while the narrowest range occurs at 01.00 - 02.00 UTC with a temperature change value between -2.0°C to 3.3°C.

**Figure 3.** Box Plot of Hourly Air Temperature Changes 2016 - 2020 (Rain Conditions).

**Figure 4.** Dot Plot of Hourly Air Temperature Changes 2016 - 2020 (Rain Conditions).

4. **Conclusions**

The threshold for hourly air temperature data varies every hour. In no rain conditions, changes in air temperature are dominated by an increase in temperature, especially at 00.00 UTC to 07.00 UTC, and then dominated by a decrease in temperature from 07.00 UTC to 12.00 UTC. Rain does not decrease the air temperature absolutely, for the increase of air temperature can still occur, but it rarely happens.

With these thresholds, we can carry quality control using the step check method. We suggest a database system that can automatically generate the step check quality control. It can also give a warning
for observers whenever they enter a value that exceeds the threshold. With this warning, the observer can do further inspection on the data before re-entry the data into the database system, thus resulting in a valid dataset.

References

[1] Rollenbeck R, Trachte K and Bendix J 2016 a new class of quality controls for micrometeorological data in complex tropical environments. *J. of Atm and Oceanic Techn.* **33** 169

[2] Bernhardt J, Carleton AM, and Lamagna C 2018 a comparison of daily temperature-averaging methods: spatial variability and recent change for the CONUS *J. of Climate* **31** 979

[3] Dupuis DJ 2017 assessing the role of hourly changes in the occurrence of daily extreme temperature *J of Climate* **30** 83-93

[4] Abbot PF 1986 *Guidelines on the Quality Control of Surface Climatology Data.* WCP-85, WMO/ND-111.

[5] Haryoko U, Nurmayati T, Handiana D and Fitri HT 2014 *Kendali Mutu Data Cuaca dan Iklim,* (Jakarta: Puslitbang BMKG)

[6] Hubbard KG, Goddard S, Sorensen WD, Wells N and Osugi TT 2005 performance of quality assurance procedures for an applied climate information system *J. of Atmospheric and Oceanic Technology* **22** 105-12

[7] Fiebrich CA, Morgan CR, McCombs AG, Hall Jr PK and Renee AM 2010 quality assurance procedures for mesoscale meteorological data. *J. of Atmospheric and Oceanic Technology.* **27** 1565-82

[8] Hubbard KG, Guttman NB and You J, Chen Z 2007 an improved qc process for temperature in the daily cooperative weather observations *J. of Atmospheric and Oceanic Technology.* **24** 206-13

[9] Zahumensky I 2004 guidelines on quality control procedures for data from automatic weather stations

[10] Anne Ru Cheng, Tim Hau Lee, Hsin I Ku and Yi Wen Chen 2016 quality control program for real-time hourly temperature observation in Taiwan *J. of Atmospheric and Oceanic Technology* **33** 953-76

[11] Vejen F, Jacobsson C, Frederiksson U, Moe M, Andresen L, Hellsten E, Rissansen P, Palsdottir T and Arason T 2002. quality control of meteorological observations: automatic methods used in the Nordic countries *Climate Report* Report 8/2002

[12] Bonsal BR, Zhang X, Vincent LA and Hogg WD 2001 characteristics of daily and extreme temperatures over Canada *J. of Climate.* **14** 1959-76

[13] Pramono SA 2020 quality control of air temperature data in Semarang city using step check method *Buletin Meteorologi Klimatologi dan Geofisika* **12** 19-27