Quality-controlled meteorological datasets from SIGMA automatic weather stations in northwest Greenland

Motoshi Nishimura[1], Teruo Aoki[2], Masashi Niwano[2], Sumito Matoba[3], Tomonori Tanikawa[2], Tetsuhide Yamasaki[4], Satoru Yamaguchi[5], Koji Fujita[6]

[1] Shinshu University, [2] Meteorological Research Institute, Japan Meteorological Agency, [3] Hokkaido University, [4] Avangnaq Arctic Project, [5] National Research Institute for Earth Science and Disaster Resilience , [6] Nagoya University

EGU General Assembly 2025
April 29, 2025@Vienna, Austria
AS4.1 Polar Meteorology and Climate and their Links to the Rapidly Changing Cryosphere
[EGU25-13984]

Contact: nishimura.motoshi@shinshu-u.ac.jp

This poster's contents have been published in Earth System Science Data (https://doi.org/10.5194/essd-15-5207-2023)

ABSTRACT |

Objective: To establish an automatic method for Quality Control for in situ meteorological data observed on glaciers and ice sheets.

(i) We propose a Quality Control (QC) method for meteorological data observed on the ice sheet and Qaanaaq Ice Cap in northwestern Greenland.

(ii) The criteria for determining erroneous values are based on meteorological and glaciological evidence ⇒ including Generality and versatility.

(iii) 12-years of temporal variation of air temperature and surface height from 2012 were revealed.

(iv) Datasets (2012–2020) and QC algorithms are opened:

SIGMA-A; http://doi.org/10.17592/001.2022041303, SIGMA-B; http://doi.org/10.17592/001.2022041306, Algorithm: https://github.com/MotoshiNishimura/SIGMA-AWS-QC.git

INTRODUCTION

When conducting weather observations in cold regions, observation data with errors and anomalous values are contained. Those data are caused by environmental factors such as snow and ice on sensors and the effects of low temperatures on sensor sensitivity may be recorded as observed values. The criteria for identifying such anomalous values often require subjective judgments, and therefore, a large human cost is required to construct a data set with high scientific value.

Therefore, this study developed an algorithm for quality control (QC) by removing anomalies using data sets obtained from meteorological observation sites on the ice sheet and ice cap in northwestern Greenland. This semi-automates the masking algorithm of anomalous data and significantly reduces the time required for QC. In addition, we were able to establish an algorithm that can be generalized and applied to other sites.

SITE DESCRIPTION & METHOD

≪AWS site≫

・SIGMA-A site:
On the Ice sheet (N78.052. 0176°, W67.628°: 1490 m a.s.l.)
・SIGMA-B site:
On the Qaanaaq Ice Cap (N77.518°, W69.062°: 944 m a.s.l.)
Period: July 2012 – August 2024 (still under observation)

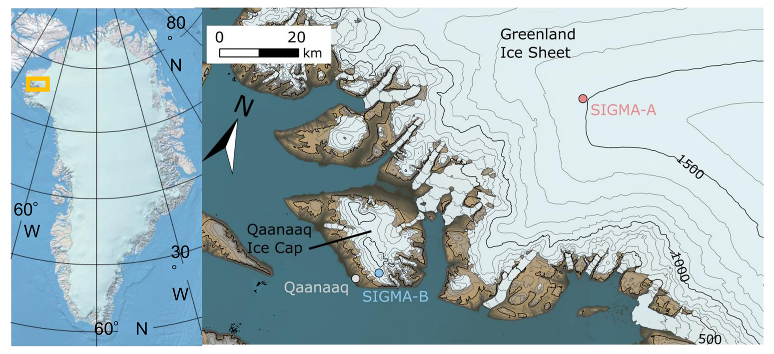

Figure 1. Map showing SIGMA-A and B sites

≪Method≫

QC is divided into two processes: **Initial Control** and **Secondary Control**. Initial Control removes values that violate physical laws and also plays a pre-processing role in improving the accuracy of Secondary Control. Secondary Control sets stricter value ranges than Initial Control and detects abnormal values. The data sets from the SIGMA A and B sites were classified into levels 1.1-1.3, according to the level of data processing.

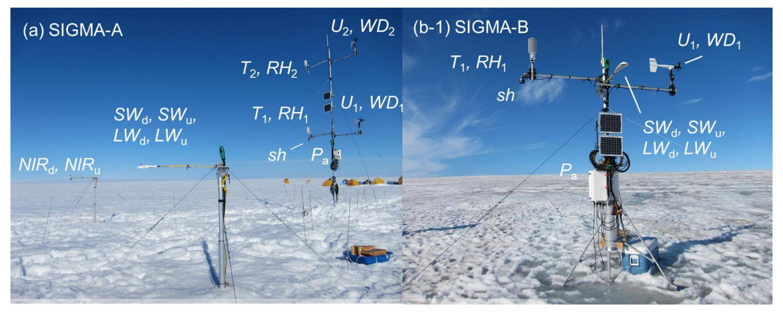

Figure 2. Overview of AWS sites

Initial Control → Secondary Control

Level 1.1 dataset (Raw data) → Level 1.2 dataset → Level 1.3 dataset (QC completed data)

Initial Control (IC):

Range Test: Set the rough range of possible values for each parameter and judge abnormal data.

Manual mask: Eliminate specific abnormal data that frequently recorded due to sensor degradation.

Data correction: Correction of additional information on the main pole of SIGMA-A data masking for AWS maintenance

Secondary Control (SC):

Range test: More stringent range of values than Initial Control is set to determine abnormal data.

Anomaly test: Calculate representative and threshold values of meteorological elements at a given time from other data. Data that deviates significantly from these values are judged as anomalies.

Manual mask: Remaining anomalies are visually checked and masked.

Level 1.1: Data set with masked data for the period when loggers are stopped (actual raw data).

Level 1.2: Data set of Level 1.1 with Initial Control to mask anomalous values that are impossible in terms of environmental conditions and physics. It also includes data pre-processing for Secondary Control, such as data processing for AWS maintenance.

Level 1.3: Data set of Level 1.2 with Secondary Control. This is a data set from which all erroneous values have been removed. The process is based on the variability of the observed values and the possible values of general meteorological physical quantities and the observation environment.

[Reference]

Brock and Arnold (2000): A spreadsheet-based (microsoft excel) point surface energy balance model for glacier and snow melt studies, https://doi.org/10.1002/1096-9837(200006)25:6<649::AID-ESP97>3.0.CO;2-U

Nishimura et al. (2023): Quality-controlled meteorological datasets from SIGMA automatic weather stations in northwest Greenland, 2012–2020, https://doi.org/10.5194/essd-15-5207-2023

Niwano et al. (2024): Field activities at the SIGMA-A site, northwestern Greenland ice sheet, 2023, https://doi.org/10.5331/bgr.24R03

Ohmura (1981): Climate and energy balance on the Arctic tundra, https://doi.org/10.1002/joc.3370020106

RESULT

QC example of downward longwave radiation (LW_d) (Fig. 3e)

The standard atmospheric radiation (LW_{std}) is calculated from the air temperature (Eq. (2)), and if the difference between LW_d and LW_{std} exceeds its median ± SD (×γ), LW_d is masked. The coefficient γ, which defines the standard deviation range, is determined by the degree of erroneous data variability.

$$|LW_{std} - LW_d| > med\_LW_d + \gamma \times SD\_LW_d \rightarrow \text{masking } LW_d \quad (1)$$

$$LW_{std} = \varepsilon^* \sigma T_A{}^4 \quad (2)$$

$\varepsilon^* = (1 + \kappa n)\varepsilon_0$, (Brock and Arnold, 2000)

$\varepsilon_0 = 8.733 \times 10^{-3} \times (T_A + 273.15)^{0.788}$. (Ohmura, 1981)

QC example of surface height (sh) (Fig. 3f)

If the difference between sh and the mean value (sh_{mean}) from any given time up to 72 hours before exceeds the median of the difference + SD (×γ), sh is masked. The coefficient γ that defines the standard deviation range is determined by the degree of erroneous variation in the dataset.

$$|sh - sh_{mean}| > median\_\Delta sh + \gamma \times SD\_sh \rightarrow \text{masking } sh \quad (3)$$

12-years of temporal variation of air temp. and surface height (Fig. 4)

The surface height at SIGMA-A has increased by 9.5 m over the past 12 years, although there have been some missing measurements due to mechanical malfunctions. As a result of the SIGMA-A AWS upgrade, observations have been resumed. The surface height at SIGMA-B showed a decrease of nearly 1.3 m in the summer of 2023.

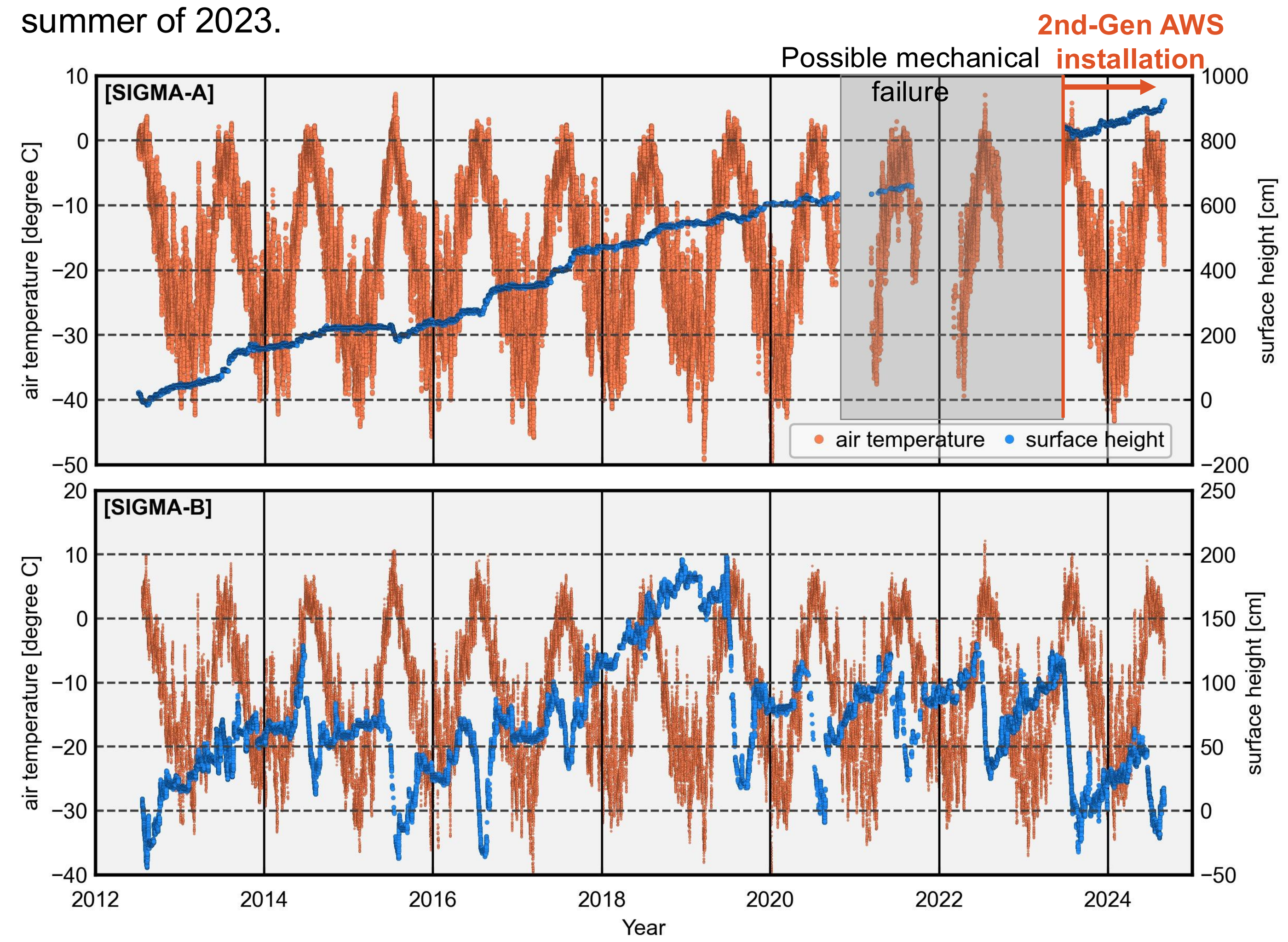

Figure 4. Temporal variation of hourly air temperature and surface height at SIGMA-A and B sites from July 2012 to August 2024. SIGMA-A AWS was updated on June 25, 2023 at the same site (see Niwano et al., 2024 for details). As a result of the upgrade, the recording of observational data was restarted.

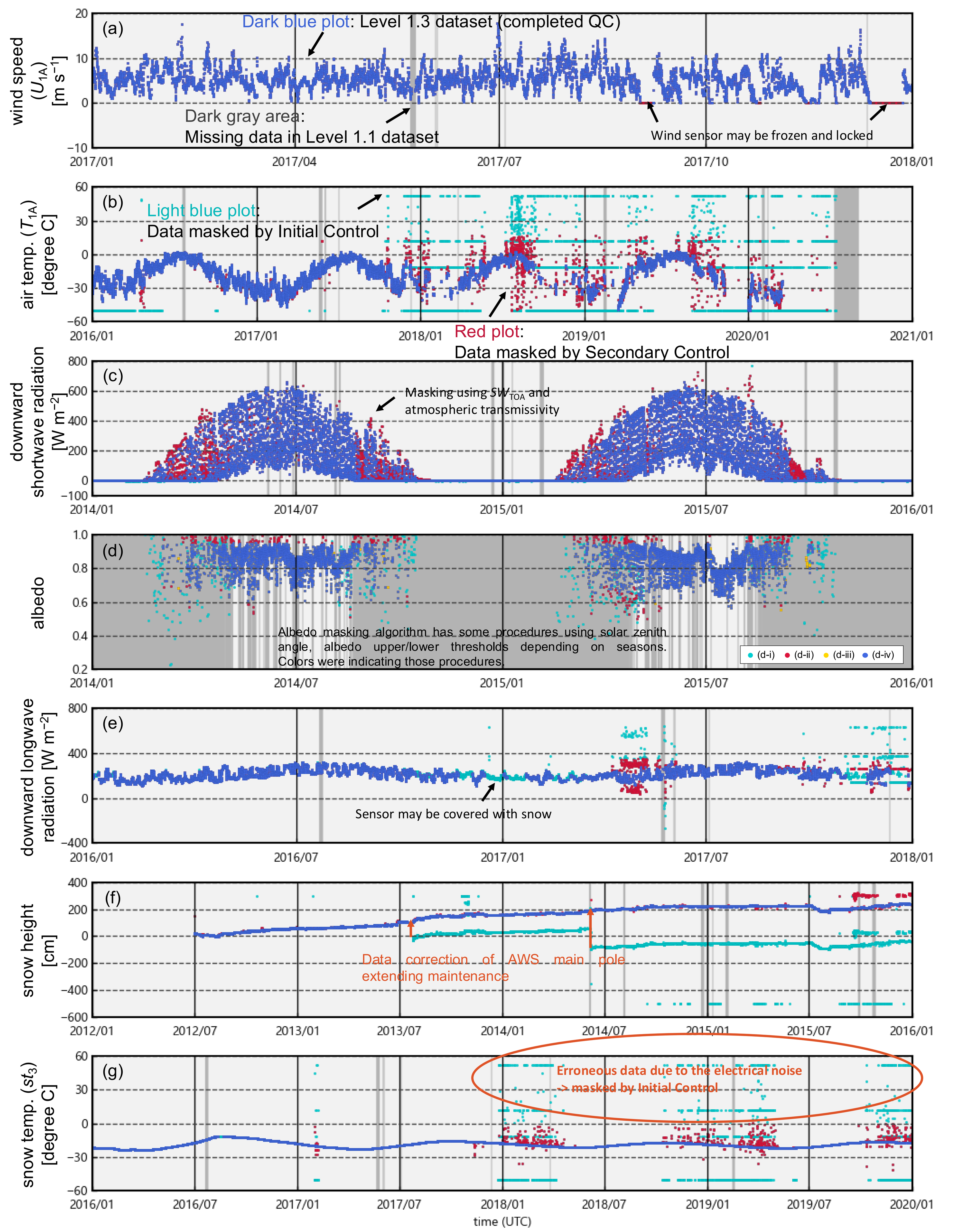

Figure 3. QC effects of (a) wind speed, (b) air temperature, (c) downward shortwave radiation, (d) albedo, (e) downward longwave radiation, (f) snow height, and (g) snow temperature at the SIGMA-A site. Except for (d), the dark gray areas represent the masked areas of Level 1.1 data, the light blue plots represent the data masked by Initial Control, the red plots represent the data masked by Secondary Control, and the dark blue dots represent the Level 1.3 data. Level1.3 data. In (d), the gray shaded area represents masked (-9999) data where the downward shortwave radiation is zero or cannot be calculated for other reasons. The light blue, red, and yellow dots represent the data masked by the three QC operations of Secondary Control, respectively.