Permafrost long term observatories of the Alfred-Wegener-Institute

Description of the data quality flagging routines

2021-05-25

Lange, Stephan; Grünberg, Inge; Bornemann, Niko; Lehr, Christian; Cable, William; Julia, Boike

Alfred Wegener Institute

1. Introduction

This document gives an overview of the data quality flagging routines used in the data-publications of the permafrost long term observatories of the Alfred-Wegener-Institute.

The data format is standardized as:

- equal time steps in UTC,
- comma separated values (csv),
- decimal point and
- missing values marked as NA.

2. Quality Flag Overview

Each data value is assigned one quality flag number. A flag 0 marks data that passed all quality checks, the other flags between 1 and 8 indicate different data quality issues (Table 1).
The quality flags are assigned to the data according to the order from flag 1 to flag 8. This means if a data value is flagged with, for example 2, and is flagged again with a larger flag, for example 5, than the higher ranked flag (the lower flag number, 2 in this case) is maintained.

The flagging is partly rule based and automatic, and partly done manually.

- Flag 2 and 3 are set manually.
- Flag 6 and 7 are partly set manually and partly rule based.
- The other flags are rule based and set automatically.

Table 1: Overview on the 9 data quality flags.

Flag	Meaning	Description
0	Good data	All quality tests passed
1	No data	Missing value
2	System error	System failure led to corrupted data e.g., when the power supply broke down, sensors were removed from their proper location, sensors broke or the data logger recorded error codes
3	Maintenance	Values influenced by the installation, calibration or cleaning of sensors, or the programming of the data logger; information from field protocols of engineers
4	Physical and sensor limits	Values outside the physically possible or likely limits and values outside the sensor limits
5	Gradient	Values unlikely because of prolonged constant periods or high/low spikes; tested within each single series
6	Plausibility	Values unlikely in comparison with other parameters or for a given time of year; partly rule based (db_filter_II.R) or flagged manually by engineers
7	Decreased accuracy	Values with decreased sensor accuracy e.g., identified when thawing / freezing soil does not have a temperature of 0 °C (zero-curtain)
8	Snow covered and water table under frozen conditions	Unreliable air temperature values because of snow covered sensors and unreliable water table values because of frozen conditions

The measurement frequencies depend on the sensors used. With the sensor internal software, the data is aggregated such that the temporal resolution of the datasets varies between 15 min and 1 hour. Please note that that the temporal resolution of the dataset has been taken into account when parameterizing flag 4 and 5. Thus, it would be valid under a different temporal resolution.

3. Quality Flag Examples

Flag 1: No Data

- All timesteps without data is set to NA.
- A subset of data is set to NA when the snow height (Dsn) reaches up to the sensor height.
- All data ≤ -99999 or > 100000 is set to NA because no parameters are measured that have valid values outside this range.

Flag 2: System Error

For each variable, system errors at the sensors are flagged manually in the filter files for each station and year (example Table 2).

Table 2: An example of a filter file from the Bayelva station in the year 2019.

from	to	dataset	variable	filtername	min	max	flag	date	by
24.09.2015 20:30	28.09.2015 22:30	BaMet2009	Dsn	minmax manual	0.03	0.14	6	08.04.2016 11:47	NB
23.06.2015 20:30	24.06.2015 12:30	BaMet2009	Dsn	minmax manual	0.18	0.22	6	08.04.2016 12:03	NB
25.04.2015 09:00	25.04.2015 12:30	BaMet2009	SwIn	minmax manual	76.99	303.57	6	03.05.2016 14:50	NB
31.12.2014 00:00	30.12.2015 23:30	BaMet2009	Ts_203_53	syserror	-50	50	2	06.05.2016 00:00	NB
31.12.2014 00:00	30.12.2015 23:30	BaMet2009	Ts_203_93	syserror	-50	50	2	06.05.2016 00:00	NB
31.12.2014 00:00	30.12.2015 23:30	0	Ts_203_143	syserror	-50	50	2	06.05.2016 00:00	NB
…	…	…	…	…	…	…	…	…	…

Additionally, there is rule-based flagging for flag 2:

For example, at the Bayelva site in 2000 and 2001, a systematic error in the PT100 temperature sensors occurred when air temperature at 200 cm height (Tair_200) exceeded -3 °C over two annual complete freeze thaw cycles. Erratic low values confirm that most measurements of all sensors are affected, including the PT100 air temperatures, but only when the temperature (most likely the data logger temperature) exceeded a threshold of roughly -3 °C. This concerns the PT100 sensors of the soil profiles 252 and 203 and of the air temperatures 20, 35, 48, 100.
It was determined this was due to lose screws connecting the sensors to the data logger between 2000-04-26 and 2001-11-12. For security, one day is added as buffer to the "loose screws period" and the values from all affected sensors are flagged between 2000-04-25 and 2001-11-13.

Flag 3: Maintenance

For each variable the maintenance periods are defined manually in the maintenance files for each station and year.

Table 3: An example of a maintenance file for the Bayelva station in 2019.

from	to	dataset	variable	flag	date	by
01.09.2015 10:00	01.09.2015 12:00	BaHole2009	Tair_50	3	03.02.2016 14:09	NB
01.09.2015 10:00	01.09.2015 12:00	BaHole2009	Ts_0	3	03.02.2016 14:09	NB
01.09.2015 10:00	01.09.2015 12:00	BaHole2009	Ts_50	3	03.02.2016 14:09	NB
01.09.2015 10:00	01.09.2015 12:00	BaHole2009	Ts_100	3	03.02.2016 14:09	NB
...	...	...	...	...	...	...

Flag 4: Physical and sensor limits

The physical and sensor limits of the observed variables are defined for each station separately. The limits are based on physical considerations (e.g. relative humidity should be in a range of 0-100%) or the manufactures measurement limits of the sensors (e.g. Air temperature below –50 °C and above 40°C), both adapted to the arctic conditions and special conditions resulting from the type of setup (marked with * in table 4 and 5).

Table 4: Physical and sensor limits from the Bayelva station.

variable	category	columnname	unit	min	max
Temperature Air	Tair	Tair_(height in cm)	°C	-50	40
Precipitation	prec	prec	mm	0	30
Relative humidity	RH	RH_(height in cm)	%	0	110
Snow depth*	Dsn	Dsn	m	-0.1	1.4
Radiation net all	RadNet	RadNet	W/m^2	-300	1000
Radiation short wave in	SwIn	SwIn	W/m^2	-10	1000
Radiation short wave out	SwOut	SwOut	W/m^2	-10	1000
Radiation long wave in	LwIn	LwIn	W/m^2	90	620
Radiation long wave out	LwOut	LwOut	W/m^2	140	620
Wind speed	wind_v	wind_v_(height in cm)	m/s	0	30
Wind direction	wind_deg	wind_deg_(height in cm)	Degree	0	360
Wind direction standard deviation	wind_sddeg	windsddeg_(height in cm)	Degree	0	360
Temperature Soil	Ts	Ts_(depth in cm)	°C	-30	30
Volumetric water content	vwc	vwc_(depth in cm)	%	0	1
Electric conductivity	cond	cond_(depth in cm)	S/m	0	0.1
Dielectricity	E2	E2_(depth in cm)		2.5	45
Dielectricity snow	E2_sn	E2_sn_(depth in cm)		0.95	40
Soil heat flux	G	G	W/m^2	-130	130

* for the current setup

Table 5: Physical and sensor limits from the Samoylov station.

variable	category	columnname	unit	min	max
Temperature Air	Tair	Tair_(height in cm)	°C	-56	40
Precipitation	prec	prec	mm	0	30
Relative humidity	RH	RH_(height in cm)	%	0	110
Snow depth*	Dsn	Dsn	m	-0.1	1
Radiation net all	RadNet	RadNet	W/m^2	-300	1000
Radiation short wave in	SwIn	SwIn	W/m^2	-10	1000
Radiation short wave out	SwOut	SwOut	W/m^2	-10	1000
Radiation long wave in	LwIn	LwIn	W/m^2	100	620
Radiation long wave out	LwOut	LwOut	W/m^2	140	620
Wind speed	wind_v	wind_v_(height in cm)	m/s	0	30
Wind direction	wind_deg	wind_deg_(height in cm)	Degree	0	360
Wind direction standard deviation	wind_sddeg	wind_sddeg_(height in cm)	Degree	0	360
Temperature Soil	Ts	Ts_(depth in cm)	°C	-50	30
Temperature ground surface	Tgs	Tgs	°C	-65	40
Temperature water (swimming)	Tw	Tw_(depth in cm)	°C	-65	40
Volumetric water content	vwc	vwc_(depth in cm)	%	0	1
Electric conductivity	cond	cond_(depth in cm)	S/m	0	0.1
Electric conductivity	Cond	Cond_(depth in cm)	S/m	0	30
Dielectricity	E2	E2_(depth in cm)		1.75	100
Dielectricity snow	E2_sn	E2_sn_(depth in cm)		0.5	90
Soil heat flux	G	G	W/m^2	-130	130
Radiation shortwave net	SwNet	SwNet	W/m^2	-10	1000
Radiation longwave net	LwNet	LwNet	W/m^2	-300	300
Radiation Albedo	Albedo	Albedo		0	1
Distance temperature correction	distcor	distcor	m	0.01	2.5
Water table*	WT	WT_(above sensor base)	cm	-0.3	0.85
Water level*	WL	WL_(water level in relation to surface or ground)	cm	-0.45	0.6

* for the current setup

Flag 5: Gradient

Flag 5 is designed to flag “erratic behavior” and unrealistic periods with exactly the same values in the series. “Erratic behavior” here means the erratic up and down of single data points around an overall trend (Figure 1).

This kind of behavior was observed mainly from soil temperature “Ts” sensors during the first years of the monitoring. After 2009, new sensors were installed and soil temperature was not checked any more with the flag 5 routine, because with increasing depths the soil temperature series become very smooth which resulted in too many data points being flagged for exhibiting constant values. Volumetric water content “vwc” was similarly excluded from the flag 5 routine because soil temperature is used in its calculation.

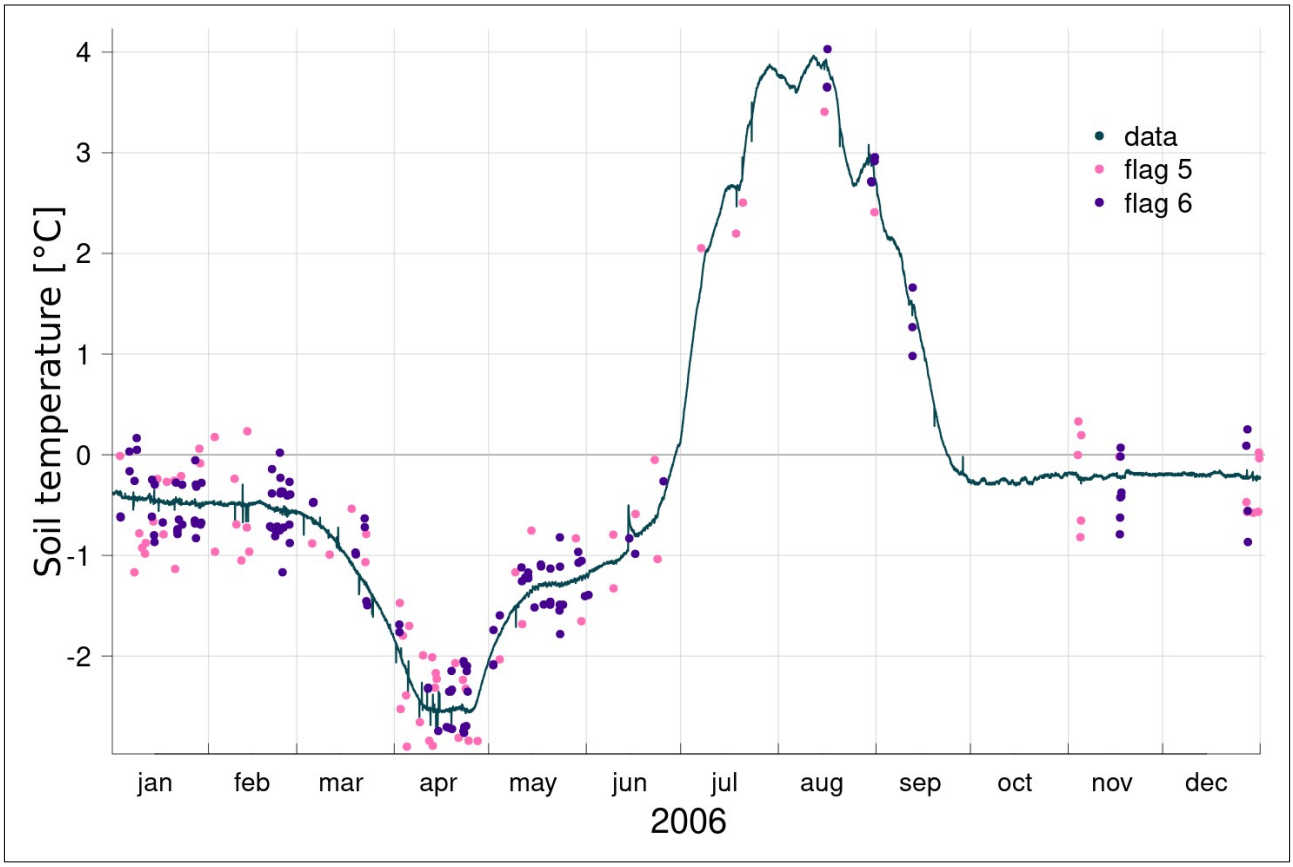

Figure 1: Example for the kind of erratic behavior that is flagged with flag 5 (soil temperature in 99 cm depth at Bayelva, Spitsbergen in year 2006).

Each series is checked for prolonged constant periods and high / low spikes. This is done automatically by a sequence of six steps (a, …, f). The checks are adjusted for categories of the observed variables (Table 6) with six parameters (pp1, ..., pp6):

pp1: maximum absolute difference between a spike and the two adjacent values (before and after) in the measurement unit
-> used in step a)
-> the bigger the number, the larger spikes are allowed to be

pp2: minimum difference between a spike and the two adjacent values used to identify spikes
-> used in step b)
-> the bigger the number, the less spikes are identified (most of these spikes are not flagged in the end, just if they are in an otherwise constant period)

pp3: maximum allowed difference between two values that defines a constant period, in which single spikes are unlikely
-> used in step b)
-> the bigger the number, the more variability is allowed in constant periods and the more spikes are identified

pp4: number of closest non-NA entries for computing the quantiles in a moving window which is centered at the value to be checked
-> used in step c) - e)
-> the bigger the number, the more values are used for the quantiles

pp5: scaling factor that determines the maximum distance a value is allowed to exhibit from the median. Maximum positive deviation: median + pp5 * (95th percentile - median). Maximum negative deviation: median - pp5 * (median - 5th percentile).
-> used in step c) - e)
-> the bigger the number, the less spikes are identified

pp6: number of time steps in a row with exactly constant values which indicate a suspicious period; values of 100 or more indicate no removal of constant periods
-> used in step f)
-> the bigger the number, the fewer constant periods are identified

The checks are performed separately for single years and single categories of variables.
The checks are performed only if there are, for the selected year and variable, more non-Na's than 2 * pp4 + 1.

The parameters of the checks for the different categories of variables are defined in Table 6. Most parameters are defined for all datasets of the permafrost stations. Some parameters are defined separately for distinct datasets.

Table 6: Parameters of the checks for the different categories of variables.

variable	category	unit	dataset	pp1	pp2	pp3	pp4	pp5	pp6
Temperature Air	Tair	Degree C	all	5	1	0.1	3	3	24
Precipitation	prec	mm	all	100	100	0	1	100	100
Relative humidity	RH	%	all	18	10	0.1	8	3	12
Snow depth	Dsn	m	all	0.15	0.02	0.01	2	3	12
Radiation net all	RadNet	W/m^2	all	200	40	1	6	3	12
Radiation short wave in	SwIn	W/m^2	all	500	40	1	3	3	100
Radiation short wave out	SwOut	W/m^2	all	500	40	1	3	3	100
Radiation long wave in	LwIn	W/m^2	all	100	40	1	8	3	12
Radiation long wave out	LwOut	W/m^2	all	100	40	1	8	3	12
Wind speed	wind_v	m/s	all	8	5	0.1	4	3	12
Wind direction	wind_deg	Degree	all	360	360	0	6	100	12
Wind direction standard deviation	wind_sddeg	Degree	all	360	360	0	6	100	12
Electric conductivity	cond	S/m	all	0	0	0	48	3	12
Dielectricity	E2		all	2	0.25	0.01	48	3	6
Dielectricity snow	E2_sn		all	2	0.1	0.01	48	3	12
Soil heat flux	G	W/m^2	all	16	1	0.1	10	3	48
Radiation shortwave net	SwNet	W/m^2	all	500	40	1	3	3	100
Radiation longwave net	LwNet	W/m^2	all	100	40	1	8	3	12
Radiation Albedo	Albedo		all	0.1	0.03	0.01	6	3	12
Distance temperature correction	distcor	m	all	0.15	0.03	0.01	4	3	12
Counts Gamma radiation K	CountsK		all	2000	40	1	3	3	100
Counts Gamma radiation TL	CountsTL		all	500	40	1	3	3	100
Snow water equivalent	SWE	mm	all	500	40	1	3	3	100
Operating temperature crystal	Tcryst	Degree C	all	4	0.2	0.1	8	3	48
Temperature Soil	Ts	Degree C	all	4	0.2	0.1	8	3	48
Temperature water	Tw	Degree C	all	4	0.2	0.1	8	3	48
Distance raw	distraw	m	all	0.15	0.03	0.01	4	3	12
Water table	WT	Degree C	all	5	1	0.1	3	3	12

variable	category	unit	dataset	pp1	pp2	pp3	pp4	pp5	pp6
Water table change	WTch	Degree C	all	5	1	0.1	3	3	12
Water level	WL	Degree C	all	5	1	0.1	3	3	12
Discharge	Q	l/s	all	100	40	1	8	3	12
Dielectricity	E2		SaSoil2002	3	0.25	0.01	48	5	6
Soil heat flux	G	W/m^2	SaSoil2002	50	1	0.1	10	3	48
Temperature ground surface	Tgs	Degree C	SaSoil2002	5	1	0.1	4	3	24
Electric conductivity	cond	S/m	SaSoil2002	0	0	0	8	3	12
Electric conductivity	cond	S/m	SaSoil2012	0.3	0.2	0.1	12	3	12
Dielectricity snow	E2_sn		SaSoil2012	5	1	0.1	3	3	24
Dielectricity	E2		SaSoil1998	13	1	0.01	48	5	6
Soil heat flux	G	W/m^2	SaSoil1998	50	1	0.1	10	3	48
Dielectricity snow	E2_sn		SaSoil1998	13	1	0.1	3	3	24
Snow depth	Dsn	m	SaMet1998	5	1	0.1	3	3	24

The data is flagged according to the following steps:

a) Flag spikes (local maxima or minima of one or two values) which are more than pp1 different from both neighbouring values.

- This is performed excluding NAs because some strange minima / maxima are situated next to NAs. Thus, “neighbouring values” is restricted here to non-NAs.
- Two value peaks are defined such that
  i. the lag-2 difference of the values adjacent to the peak is larger than the threshold pp1, that is the sum of the difference to the neighbouring value ((x+1) - x) and the difference of the neighbouring value and the second neighbour ((x+2) - (x+1)), AND
  ii. the lag-1 difference of the values adjacent to the peak is larger than the threshold pp1

b) Flag one value-spikes bigger than pp2 in constant periods defined by pp3.
The following conditions have to be fulfilled:

i) spikes > pp2, and

the spike is situated:

ii) 2 time steps after the absolute values of the lag-4 difference < pp3, and
iii) both:
2 time steps after the absolute values of the lag-1 difference < pp3 AND
1 time step earlier the absolute values of the lag-1 difference < pp3

Condition ii) ensures that 2 timesteps after the spike the values are again back on almost the same level as 2 timesteps before the peak (the lag-4 difference 2 timesteps prior the peak is smaller than pp3). This serves as indication for a relatively “constant” 4 timesteps period with the spike in the middle. Condition iii) ensures that 2 timesteps before the spike as well as 1 time step after the spike, the lag-1 difference is smaller than pp3.

c) Flag spikes which are

i) smaller than the median - pp5 * (median - 5th percentile) or
ii) larger than the median + pp5 * (95th percentile - median)

where the median and quantiles are calculated for the pp4 non-NA values before and after the spike.
In the operational window defined by pp4 values of the 0.05 (0.95) quantile relatively close to the median are replaced by standardized values median - pp1 * 0.02 (median + pp1 * 0.02).
Step c) is performed THREE times in a row as the quantiles are affected by the spikes.

d) Same conditions as step c) with the additional condition that all flagged values are a

i) one, or
ii) two

value "spike" in the sense of non-NA values directly surrounded by NA values.

e) Repeat step c) ONCE with cleaned data.

f) Flag periods with at least pp6 (e.g., 10) consecutive constant non-NA values.

Flag 6: Plausibility

The plausibility flag is set automatically based on rules or manually.

Rule based flag 6 flagging:

Cases from the Bayelva site:

a) All positive or negative spikes larger than 0.07 W/m^2 and smaller than 1 W/m^2 in ground heat flux before 2000-04-18 00:00. In that period there were some erratic values exceeding those thresholds.
b) Albedo values for which “SwIn” is NA or smaller than 5 W/m^2 because values below this threshold are not reliable due to the precision of the sensor. And albedo the ratio of reflected/incoming radiation and ratio calculations are tricky with values close to 0.
c) Precipitation values of:
    i) > 0 mm if Tair_200 < -2 °C because liquid precipitation is not likely if the air temperature is < -2 °C.
    ii) > 1 mm if wind_v_200 > 12.5 m/s because a wind velocity above 12.5 m/s hourly average does not allow precise precipitation measurements.

Cases from the Samoylov Site:

d) All values of “SwIn”, “SwOut”, “SwNet”, “LwIn”, “LwOut”, “LwNet”, “Albedo” and “NetRad” if SwNet (=SwIn – SwOut) < -10 W/m^2 .
   Negative SwNet occurs if more short-wave radiation is going out than coming in, for example because of snow, dirt or birds on the incoming radiation sensor(s). Consequently, the values of all radiation measurements and calculated parameters are suspect.
e) All values of relative humidity in height 50 and 200 cm (“RH_50”, “RH_200”) if the air temperature of that height (“Tair_a_50”, “Tair_a_200”) was flagged with flag 4 (physical and sensor limits). Because both parameters are measured with the same device. Consequently, if the physical limits of air temperature are exceeded, the relative humidity values are not reliable any more.

The manual flagging is based on the experience of the staff. It consists of visual checks of:

i) the course of the respective series within one year,
ii) the course of all available years of that variable, and
iii) where applicable the comparison of the checked series to the series of other variables (for example: liquid precipitation is unlikely if the temperatures are below -2°C) or the photos of the surveillance camera of the respective station.

The manual plausibility flags are stored in the filter files for each station and year (for example Table 2).

Flag 7: Decreased accuracy

Flag 7 is only applied to soil temperature sensors. Soil temperatures are flagged automatically to exhibit decreased sensor accuracy if all of the following three conditions are fulfilled:

i) the absolute mean soil temperature of the zero-curtain period is greater than 0.5 °C (zero-curtain: the period of time during which a nearly constant temperature, very close to the freezing point, exists during annual freezing of the active layer)

ii) the standard deviation of the soil temperature during the zero-curtain period is smaller than 0.25 °C, because the zero-curtain period should exhibit little variation in temperature

iii) the 0.9 percentile of the soil temperature from the period of August 1st to the 30th is larger than 0, meaning the soil was thawed.

In Table 7 and Table 8 the first and last calendar day of the zero-curtain periods are given for each year.
Additionally, periods of decreased sensor accuracy can be flagged manually for all variables in the filter files for each station and year (for example Table 2).

Table 7: Zero curtain periods at Bayelva for the years 1998-2020. For each block: bold: years; grey: first calendar day; white regular: last calendar day of the zero-curtain period.

							1998	**1999**	**2000**
							257	270	287
							266	280	305
2001	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
271	277	277	263	276	281	273	279	267	266
293	287	291	279	283	294	288	293	275	277
2011	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**
300	281	275	271	284	313	295	295	295	295
303	291	287	289	300	317	305	305	305	305

Table 8: Zero curtain periods at Samolyov for the years 1998-2020. For each block: bold: years; grey: first calendar day; white regular: last calendar day of the zero-curtain period.

							1998	**1999**	**2000**
							266	266	266
							269	269	269
2001	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**
274	270	273	270	273	270	273	NA	270	270
275	282	279	282	279	282	279	NA	282	282
2011	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**
270	273	270	270	270	270	270	270	270	270
282	279	282	282	282	282	282	282	282	282

Flag 8: Snow covered and water table under frozen conditions

For flag 8 there are two rules.

1) Sensor snow covered (only for air temperature):
For all available sensor heights, the standard deviations of the air temperature series in a 4-day moving window are calculated. This results in series of air temperature standard deviations as a measure of the air temperature variability in the different heights (Bayelva: 0.04 m, 0.20 m, 1.00 m, 2.00 m; Samoylov: 0.50 m, 2.00 m).

For each timestep of those series:
i) the maximum air temperature standard deviation of all available sensor heights is determined, and
ii) a ratio is calculated from the air temperature standard deviations from the different heights and the maximum air temperature standard deviation.

This results in a series of air temperature variability for the different heights relative to the maximum air temperature variability from each timestep.

These values are classified into two classes.

A. All readings with standard deviation >= 80 % OR a temperature > 0.1 °C
B. All readings with standard deviation < 80 % AND (temperature ≤ 0.1 °C OR NA value)

Class B defines the conditions sensors should fulfil in case they are covered by snow. Namely, a relatively low temperature variability in the 4-day window and a temperature ≤ 0.1 °C.

For each timestep the maximum snow cover height is determined. Sensors up to this height are considered likely to be covered by snow. Snowfall and snowmelt events are identified as changes in the maximum snow cover height.

To give a more robust estimate of snow cover height, all snow heights between snowfall events and snowmelt events of which

i) the snowmelt event takes place less than 4 days after the snowfall event, and
ii) of which the new snow height is equal or below to the snow height of the snowfall event,

are replaced with the snow height prior to that snowfall event.

All air temperature values from sensors up to the height of this filtered series of snow cover height are flagged with 8.

2) Water table under frozen conditions:
This concerns only the sensor measuring water table ("WT") at the meteorological station on Samoylov. The selected year is divided into 2 parts on day 200 of the year (certainly in the unfrozen period). Thus, the first part starts in winter ("winter-spring-summer period"), the second part starts in summer ("summer-autumn-winter period").

Because the variable of interest is the liquid water table, and not the frozen water:

- all data points in the first part with soil temperature at 6 cm depth (Ts_6) < 0.4 °C, and
- all data points in the second part with soil temperature at 6 cm depth (Ts_6) < 0.1 °C, which are not flagged with another flag, other than 0, are flagged with 8.

It is assumed that coming from the colder "winter-spring-summer period", warmer temperatures are needed to change the phase of the water from solid to liquid, and that coming from the warmer "summer-autumn-winter period", colder temperatures are needed to change the phase of the water from liquid to solid. This is accounted for with the different temperature thresholds for the winter and summer influenced period. Those thresholds were fit manually.

4. Calculated variables

1) The volumetric water content (vwc) of the soil at different depths at the Bayelva station is recalculated from soil temperature (Ts) and dielectric permittivity (E2) from the respective depths in the processing of the data from LV0 to LV1. It inherits the smallest non-zero flag of Ts and E2.
In the calculation, all soil temperature and dielectric permittivity values with flags between 1 and 6 are set to NA. This has the effect that the corresponding vwc values also become NA. Thus, those vwc values are NA but flagged with flag numbers between 1 and 6, not only flag 1 for “no data”.

2) From 2009 to 2018 the net radiation in the meteorological data sets from the Bayelva station is calculated from incoming shortwave radiation (SwIn), reflected shortwave radiation (SwOut), incoming longwave radiation (LwIn), and outgoing longwave radiation (LwOut) according to:

$$RadNet = (SwIn + LwIn) - (SwOut + LwOut)$$

rounded to 3 digits and flagged with the smallest non-zero flag of the variables used in its calculation (SwIn, SwOut, LwIn and LwOut).
In the years before net radiation was calculated on the logger and the flagging occurred only for the net radiation values themselves.

5. Final processing step

In the level 1 data published in Pangaea all values with flag 2 or flag 4 are set to NA.

6. Dataset published

Soil and air data from the permafrost stations at Bayelva, Spitsbergen, Norway from 1998 to 2019 and at Samoylov, Siberia, Russia from 2002 to 2019. The data is openly accessible at the Pangaea archive at https://doi.org/10.1594/PANGAEA.880120 and https://doi.org/10.1594/PANGAEA.905236 and the related data-publications.

For details on the dataset please see Boike at al. (2018) and Boike at al. (2019).

All related files and scripts are open on https://gitlab.awi.de/sparcs/lto/time-series-preprocessing accessible.

7. References

Boike, J., Juszak, I., Lange, S., Chadburn, S., Burke, E., Overduin, P. P., Roth, K., Ippisch, O., Bornemann, N., Stern, L., Gouttevin, I., Hauber, E., and Westermann, S.: A 20-year record (1998–2017) of permafrost, active layer and meteorological conditions at a high Arctic permafrost research site (Bayelva, Spitsbergen), Earth Syst. Sci. Data, 10, 355–390, https://doi.org/10.5194/essd-10-355-2018, 2018.

Boike, J., Nitzbon, J., Anders, K., Grigoriev, M., Bolshiyanov, D., Langer, M., Lange, S., Bornemann, N., Morgenstern, A., Schreiber, P., Wille, C., Chadburn, S., Gouttevin, I., Burke, E., and Kutzbach, L.: A 16-year record (2002–2017) of permafrost, active-layer, and meteorological conditions at the Samoylov Island Arctic permafrost research site, Lena River delta, northern Siberia: an opportunity to validate remote-sensing data and land surface, snow, and permafrost models, Earth Syst. Sci. Data, 11, 261–299, https://doi.org/10.5194/essd-11-261-2019, 2019.