https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

Open Access Earth System Science Data Discussions

A comprehensive analysis of three meteorological datasets for the East Siberian continuous permafrost zone: long-term changes in air temperature, snow depth and precipitation.

Robert Sysoliatin[1,2], Zhi Wen[1], Qiang Gao[1], Kun Xiang[1], Yana Tikhonravova[1,2]

1 - Key Laboratory of Cryospheric Science and Frozen Soil Engineering, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Science, Lanzhou, 730000, China

2 – Melnikov Permafrost Institute, Yakutsk, 677010, Russia

Correspondence to: Zhi Wen (wenzhi@lzb.ac.cn)

Abstract

Climate and its evolution strongly influence numerous natural processes, and in permafrost areas this impact outweighs that of any other factors. The main publicly available databases consist of data only for 167 meteorological stations within Republic of Sakha (3,08 ml km^2), Magadan Obslat (0,46 ml km^2) and Chukotka (0,72 ml km^2), which makes the meteorological network in that region - almost entirely underlain by continuous permafrost - one of the sparsest on the planet, at approximately 1 station per 25000 km^2. The longest series dates back to 1834 at the first official meteorological station in Yakutsk-24959, while the shortest to 1987 in Anabar-21608), however, data availability is highly inconsistent, and more than half of meteorological stations contain gaps of up to 40% in their timeseries. Despite the abundance of reanalysis products, here we aim to assess the current meteorological network, the available observation data, and to merge three different datasets into a unified database containing the following parameters: mean, maximum, and minimum daily air temperature, daily precipitation totals and snow cover thickness. The resulting database spans the period from 1 January 1966 to 30 September 2025, with daily temporal resolution, and includes source references as well as the number of measurements recorded for each day as a quality indicator parameter. We carried out automated cleaning for obvious mistakes and typos as well as deleting duplicates. Chosen parameters are particularly important for a wide range of researchers, as well as for society and policymakers. This dataset has been used for filling gaps in air temperature time series using machine learning approaches. For convenient access, we have developed an online dashboard that presents the main meteorological parameters at daily, decadal, monthly, and annual scales. The database presented and described in this article is available for download at https://doi.org/10.5281/zenodo.19883706 and at https://robertseesaw.shinyapps.io/meteo-dashboard/.

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

Open Access Earth System Science Data Discussions

1. Introduction

Datasets from meteorological stations situated across the permafrost zone of northern latitudes have great value due to their limited availability (Cowtan and Way, 2014; Folland et al., 2013; Simmons et al., 2010). Severe climate conditions, road inaccessibility, and high cost of experiments and instrument maintenance result in a low density of meteorological network (Gorokhov and Fedorov, 2018; Lytkin et al., 2021; Streletskiy et al., 2019, 2025). Although numerous climatic indices exist, the most representative indicators are near-surface air temperatures, precipitation and snow depth (Hu et al., 2021; Menne et al., 2012; Tran et al., 2024). Changes in these parameters have a significant impact on water and heat exchange processes, fauna activity, geomorphological processes, as well as on permafrost conditions (Almeida et al., 2014; Cannone et al., 2006; Oliva et al., 2017). As long observation records are indispensable for almost every research, comprehensive information from meteorological stations is the key to better understanding each side of the natural environment's evolution (Jones and Moberg, 2003; Luterbacher et al., 2024; Swift and Ragsdale, 1985).

Since the pronounced increase in climate warming began in the 1960s, for this study we selected 1966 as the inception of resulting dataset (IPCC 2023). Furthermore, the transition from four-times-daily (every 6 hours) to eight-times-daily (every 3 hours) measurements at governmental meteorological stations occurred on 1 January 1966 across the entire territory of the former Soviet Union, thus, that date is considered the beginning of standardized meteorological monitoring for the vast territories of the North-East of Russia (Handbook of …, 1969). The dataset ends on 30 September 2025 in order to correctly account for the last hydrological year, therefore, it consists of both 59 calendar and hydrological annual periods.

Except for *NOAA* (https://www.ncei.noaa.gov/), there are three main datasets used to acquire meteorological data for the North-East of Russia: the official repository of the World Data Center - *meteoru* (http://meteo.ru/data/), and two datasets deployed and maintained by enthusiasts - *rp5* (https://rp5.ru/) and *pk* (www.pogodaiklimat.ru). A unified database of referenced dataset is essential for facilitating the use of meteorological information for various scientific purposes, such as the assessing the impact of climate change and identifying the trigger of past geomorphological processes (Jones and Moberg, 2003). The appearance of a unified database eliminates ambiguity in the data, avoids differences in both data presentation and databases structure, and will be beneficial for practical use (Hu et al., 2021). The correct merging of existing datasets while preserving all advantages

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

Open Access Earth System Science Data Discussions

and minimizing disadvantages is the main challenge in creating a useful and convenient product (Evans et al., 2026).

The main objective of this study is twofold: to assess both the quality and the quantity of the available data, and to characterise the specific features of the individual services providing access to meteorological parameters. We selected the three services most widely used within the scientific community, which are referenced particularly often in studies of the permafrost zone of north-eastern Eurasia (Malkova et. al., 2011; Leibman et al., 2014; Gorokhov and Fedorov, 2018). The provision and aggregation of these data are framed by the principle of free and unrestricted data exchange established by the World Meteorological Organization (WMO Resolution 40; WMO, 1995) and subsequently broadened under the WMO Unified Data Policy (Resolution 1; WMO, 2021), and the resources employed have repeatedly served as data sources for peer-reviewed publications (Leibman et al., 2014; Gorokhov and Fedorov, 2018; Grigorieva, 2020; Tran et al., 2024). An assessment of the representativeness and quality of the records promotes the evaluation of the reliability of the meteorological network as a whole and of each individual resource in particular. Given the declared principle of free exchange of meteorological information — which under the Unified Data Policy grants the research and education communities free and unrestricted access to such data for non-commercial purposes (WMO, 2021) — raw-data aggregation services cannot shield these records from use and analysis, since their publication and open availability inherently presuppose broad utilisation. In addition, an analysis of the volume of the available data allows research to be planned more deliberately, clarifying whether the existing meteorological network can be relied upon or whether independent meteorological monitoring is required.

With regard to the user interface, given the considerable size of meteorological databases, an online dashboard representation is far easier to interpret than conventional formats such as .csv, .xlsx, or .txt (Sarikaya et al., 2019; Yigitbasioglu and Velcu, 2012). Although raw data may contain valuable measurements and observations, only processed time series are used for assessment, description, and decision-making (Kääb et al., 2018; Zhu et al., 2024). Nevertheless, seemingly straightforward operations — such as finding minimum and maximum values, averaging data, and selecting a time period — applied to raw data can take even more time than more substantive and important scientific work. To address this, the data must be cleaned and preprocessed, while a dashboard can provide interactive charts and tables that react to the selected time period.

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

In this study, we aim to present a merged database of all of the online available (167) meteorological stations across 4.27 ml km^2 on the North-East of Russia. This dataset produced and checked using *R* language and simple statistical approaches. Data availability and information quality for each station was assessed, which is essential for investigating climate variability on long timescales. That preprocessing further enhances the coherence of local nature environment. We hope the proposed dashboard becomes a robust and convenient resource for evaluating long-term and spatial distributed climate variability and simplifies the preliminary steps for any advanced research.

2. Study area

Three large administrative regions are distinguished in the North-East of Russia - the Republic of Sakha (Yakutia), Magadan Oblast, and Chukotka. The combined territory of these regions extends from 100°E to 170°W in longitude and from 55°N to 76°N in latitude and is bounded by Arctic seas to the north. It consists of the 22% of the entire territory of Russia or 8% of Eurasia and 25% of the planet's permafrost zone (Brown et al., 1997; Obu et al., 2019). The region encompasses a wide range of natural landscapes, from coniferous forests in the south, through taiga, to tundra and Arctic barrens in the north (Fedorov et al., 2018, 2017). Mountainous terrain occupies about 40% of the study area and is mainly located in its southern and eastern parts, with the highest peak reaching 3003 m (Peak Pobeda). Due to severe climate conditions the continuous permafrost almost entirely underlies the region (Fedorov et al., 2018). Temperatures at a depth of zero annual amplitude decreasing toward north and elevation rising drops to -15°C on Lena Delta vicinity, but mostly that indices vary from -6 to -2°C (Anisimov and Reneva, 2006; Kirillin et al., 2022; Sysolyatin et al., 2020; Varlamov et al., 2019).

The interaction between the movement of the Arctic air masses and the Siberian High governs the climate over continental part of the study area, whereas coastlands are strongly influenced by the nearest seas (Gong and Ho, 2002). Most of the study area belonged to the continental climate by Koppen's classification that characterized by long, harsh winters, short summer and moderate amount of liquid and solid precipitations (Kottek et al., 2006). Mountainous regions have more specific climate patterns due to their tendency to promote temperatures inversion in near-ground air masses (Kalinicheva et al., 2018). In fact of that, the Northern Hemisphere's "Pole of Cold", where air temperature drops below -70°C, is located in one of the mountainous valleys in the Verkhoyansk mountain (Ivanova, 2006). Typically observed air temperatures minimum occurs in January, while higher

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

temperatures tend to July, as well as the most precipitation occurs in warm season (Gorokhov and Fedorov, 2018).

Transport accessibility, the scarcity of amenities, and the lack of staff and equipment endurance are the main challenges to maintaining meteorological monitoring in the region (Kirillina et al., 2023). Outside of major settlements, many meteorological stations are accessible only via seasonal routes — by river in summer or over frozen ground and ice in winter — making the supply of food and fuel a complex logistical task (Stephenson et al., 2011). Due to their extreme remoteness, coastal and island stations can be resupplied only a few times per year. The shortage of qualified personnel is an equally significant obstacle. Harsh living conditions, prolonged isolation, and the absence of social infrastructure make it difficult to attract and retain trained staff at remote stations, leading to gaps in observational continuity. Moreover, instruments exposed to extreme cold, strong winds, salt air, and icing are subject to frequent failure, while the delivery of spare parts is slow and costly. Power supply presents additional difficulties, as many remote stations rely on fuel-based generators, which are directly affected by the same supply constraints (Fiebrich et al., 2010; Sabatini, 2017). These logistical, human, and technical challenges collectively undermine the consistency and quality of meteorological data collected across the region.

Figure 1 depicts the location of 167 meteorological stations which were available to download in December 2025. We have removed 3 no data and 2 automatic meteorological stations out of raw datasets due to no presence even 1-year consistent observations period. There are only 15 meteorological stations established in 19th century over that area – Yakutsk, Viluysk, Verkhoyansk, Olekminsk, Sredne-Kolymsk, Anadyr, Ust-Maya, Nyurba, Churapcha, Amga, Markovo and Suntar. The next period for encompassing meteorological observations took place in 1920-1930s years to support the reclaim of that frontier land, when 59 stations were established. In short period of 1941-1945, on 31 stations the observation was begun to maintain the Lendlease program. After WWII and next 20 years there were 54 stations established and only 11 after 1966. Note, some of the stations are not available in open-access datasets, but that amount does not reach 15 stations, and their time series is not long and clear.

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

Open Access Earth System Science Data Discussions

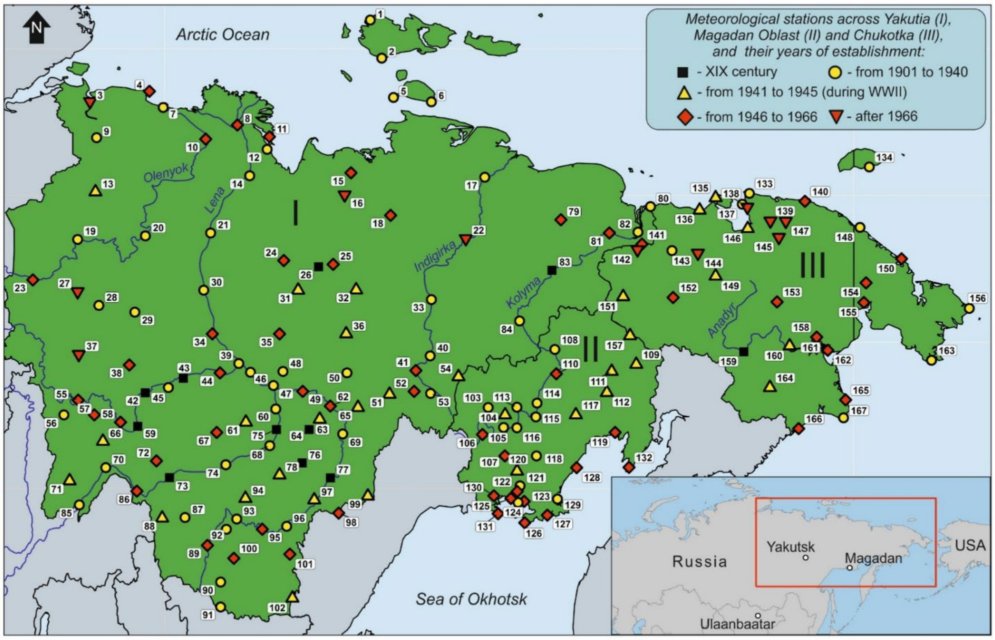


Figure 1. Schematic map showing the locations of 167 meteorological stations available from open-access datasets. Although some of the stations were established in the 19th century, this does not guarantee completeness of data or continuous records of air temperature, precipitation, or snow cover. The numbers correspond to Table S1, which provides information about each station in detail. Note: I – Yakutia, II – Magadan Oblast, III – Chukotka.

Table 1. Basic information for meteorological stations in selected regions

Region	Number of meteorological stations	Region area, km^2	Density, 1 station per km^2	Number of stations out of town or settlements, (%)	Number of coastland/island stations, (%)
Yakutia	102	3 083 523	30 231	45 (44)	12 (12)
Magadan Oblast	30	462 464	15 415	16 (53)	10 (33)
Chukotka	35	721 481	20 613	17 (49)	18 (51)
Total	**167**	**4 267 468**	**25 554**	**78 (47)**	**40 (24)**

The distribution of meteorological stations across the study area is highly uneven (Fig. 1, Table 1). Yakutia has the lowest station density relative to its total area, with approximately one station per area equivalent to the entire territory of Belgium (Bertrand et al., 2021). In the Magadan Oblast, station density is roughly twice that of Yakutia; however, a third of its stations are located in the offshore zone of the Sea of Okhotsk, while the mountainous terrain of the north and east remains almost entirely without meteorological coverage. Chukotka similarly has approximately half of its stations situated on islands or along the coastline. Nevertheless, as the

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

terrain in that region is considerably flatter than in Magadan Oblast, the spatial representativeness of the observations is comparatively more reliable.

A further issue concerns the location of stations relative to settlements. Fewer than half of all meteorological stations in the study area are situated outside of towns or populated localities. This is a significant limitation, as urban and settlement environments are well known to produce local heat island effects or create artificial shadowing, which can systematically bias temperature and humidity records (Magee et al., 1999; Oke, 1982). Consequently, despite the already critically low density of the meteorological network across the region, a considerable proportion of the available station data may be affected by local anthropogenic influences, further reducing the representativeness and reliability of observations at the regional scale.

3. Data description

As of 2015, the Russian state meteorological network comprised 1,921 stations, at which both standard meteorological parameters and more specialised variables (ground temperature, surface temperature, air pollution, etc.) are observed (Kokorev and Sherstyukov, 2015). Since 1966, observations across the state network have followed a unified methodology, harmonised with WMO standards and based on eight measurements per day (Handbook of Meteorological Instruments…, 1969, Anisimov et al., 2012). On the official portal of the All-Russian Research Institute of Hydrometeorological Information – World Data Center (VNIIGMI-WDC; http://meteo.ru/data/), 521 stations providing sub-daily observations were openly available for the territory of the Russian Federation as of 26 November 2025. As this represents only about a quarter of the existing network, enthusiast-maintained aggregator websites that facilitate access to meteorological information (pogodaiklimat.ru, rp5.ru) have gained considerable popularity.

In this study we used three open-access datasets that accumulate data from meteorological stations across the former Soviet Union (*meteoru*) and worldwide (*rp5, pk*). All of them contain stations from Yakutia, Magadan Oblast, and Chukotka, although the number of stations, data quality, and data structure are completely different. General information on each dataset is summarized in Table 2, while more detailed information can be found for each particular dataset in the Supplementary materials (for *meteoru* – Table S2, *rp5* – Table S3, *pk* – Table S4). While *meteoru* explicitly allows data utilisation and analysis, *rp5* and *pk* do not provide standard licensing terms on their websites. Correspondence with representatives of these two platforms further confirmed that our work does not conflict with their terms of use. We have ensured that the data accessed on *meteoru*,

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

Open Access Earth System Science Data Discussions

rp5 and *pk*, could be modified, aggregated and redistributed within our compilation under CC-BY license.

Table 2 Detailed data of raw and the final datasets

Dataset	Number of stations			Data availability, % *	Proportion of the final dataset, %	Numbers of parameters
	Yakutia	Magadan Oblast	Chukotka			
meteoru	47	12	20	98.36	60.8	36
rp5	97	19	17	98.88	19.8	27
pk	100	28	30	79.82	19.4	41
final	**102**	**30**	**35**	**79.55**	**100**	**5**

* - based on air temperature time series

The *meteoru* dataset was accessed on September 2, 2025, from the official repository at http://meteo.ru/data/basic-parameters/ (through http://aisori-m.meteo.ru/waisori/index.xhtml?idata=2), where a detailed description of the dataset content is also provided. As of the most recent check, the total number of available stations had decreased from 521 to 134 (on March 20, 2026). The dataset contains 36 meteorological variables in term-measurement format for 79 unique stations, with time series mostly dating back to January 1, 1966. The advantages of the *meteoru* dataset are the best data quality and availability compared to the other datasets, whereas the main disadvantages are the lowest number of stations and the absence of snow cover thickness data (Gorokhov and Fedorov, 2018; Hu et al., 2021).

The *rp5* dataset was acquired in early September 2025 by submitting individual requests for each meteorological station via a modified URL (https://rp5.ru/%D0%90%D1%80%D1%85%D0%B8%D0%B2_%D0%BF%D0%BE%D0%B3%D0%BE%D0%B4%D1%8B_%D0%B2_%D0%AF%D0%BA%D1%83%D1%82%D1%81%D0%BA%D0%B5) redirecting to the archive interface. This repository is the most widely used source of meteorological data in the Russian scientific community, as it contains the most up-to-date raw data on air temperature, precipitation, and snow depth in term-measurement format, mostly dating back to February 2005 and later. The primary drawback of the *rp5* dataset is the large number of errors and typos that require considerable time for preliminary preprocessing (Sysolyatin et al., 2020).

Like *rp5*, the *pk* repository aggregates observational records transmitted through the WMO World Weather Watch SYNOP exchange protocol; surface synoptic observations of this kind constitute essential data subject to free and unrestricted international exchange under WMO Resolution 40, which has since been

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

consolidated within the WMO Unified Data Policy (Zillman, 2019). For the *pk* dataset, an automated data parser was developed using the *httr* and *rvest* R packages, which substantially streamlined data acquisition. After a successful login, the user gains access to the summary section of the selected station, where 41 indices are available. This enthusiastically maintained resource is suitable for a brief overview of climate conditions and for acquiring satisfying information about a station's meteorological time series (Mingalev, 2021; Veselkin et al., 2019). On the other hand, the *pk* dataset provides no term-measurement data (only daily, monthly, and yearly scales), and its use for scientific purposes may be considered controversial, as no methodology is described for preprocessing the raw data. Nevertheless, it contains the largest number of unique stations (the number may vary, but as of late December 2025, it had been 163).

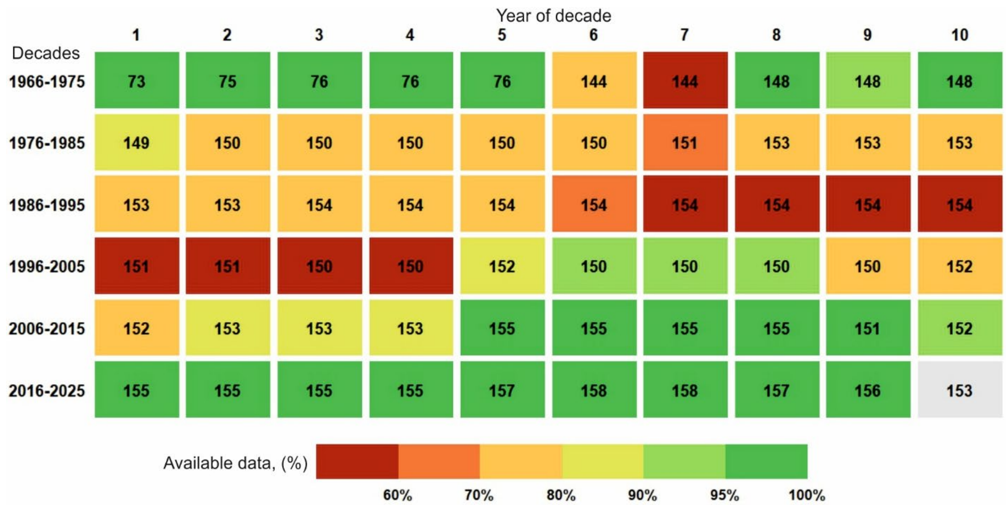


Figure 2. Temporal distribution of data availability across the study period (heat map) for the North-East of Russia meteorological network. Cell color indicates the percentage of days with observations, and the number in the middle of each cell shows the count of meteorological stations with records during that year.

After preliminary preprocessing of the *meteoru*, *rp5*, and *pk* datasets, followed by quality checking and merging (see the Methods chapter for further details), the *final* dataset comprises 167 unique meteorological stations with the following attributes: WMO ID, station name, date, mean daily air temperature, daily maximum and minimum temperatures, total precipitation, snow depth, a quality flags referencing the source dataset, the difference between the first and last measurement of the day, and the daily's number of measurements. The merging hierarchy was established such that *meteoru* served as the primary source, as it contains the fewest errors. Preprocessed data from *rp5* were then merged into it where term-measurement data

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

were available, following quality checking and duplicate cleaning. Finally, data from *pk* were incorporated. Unfortunately, the *pk* dataset lacks the quality flag of time difference between the first and last term, so these fields are empty for all records originating from that source. Several important variables such as wind speed, wind direction, and atmospheric pressure, were excluded owing to the complexity of their quality control and preprocessing.

Data loss has been a common problem on the vast North-East of Russia meteorological network due to harsh working conditions, and we assume that the best way to represent the availability of data is by plotting a heat map chart. Figure 2 shows the number of available meteorological stations and the proportion of data presence in the *final* dataset through the study period. The first five years from 1966 have perfect availability of data, then for two consecutive years it drops below 80% in 1971 and below 60% in 1972, after incorporating *pk* dataset. From 2010 to 2024 the final dataset has almost always more than 95% of the available data, while the 1990s are the worst period in the entire time series. With respect to the precipitation and snow cover time series, the raw datasets are of considerably lower quality and continuity. To cope with that and avoid uncertainty we preprocessed and cleaned the data as well and added a reference to the source dataset for the snow cover data in the *final* dataset.

4. Method

A general overview of the workflow for creating the final dataset is shown in Fig. 3. We use *RStudio* (*R* version 4.5.2), a programming environment built around the *R* language, which is well suited for processing large datasets in matrix form (Posit team, 2025). It allows reading in, standardizing, and modifying various table-format sources through the *tidyverse* package. Due to the heterogeneous structure of the datasets, different preprocessing strategies were applied. The full *R* scripts for each dataset can be provided upon request; for the *meteoru* dataset, they are available in the Zenodo repository (Sysoliatin, 2026).

Overall, the data preprocessing consisted of five major steps: (1) data reformatting and elimination of single observations, (2) duplicate identification and removal, (3) air temperature measurement quality control using simple statistical and logical approaches, (4) snow depth quality check, and (5) liquid precipitation quality check. All of these steps were necessary to produce harmonized and uniform datasets for their successful merging and to ensure the best possible data quality in the *final* dataset (Wang et al., 2018).

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

Open Access Earth System Science Data Discussions

No reanalysis products were used at any stage of the preprocessing in order to preserve the integrity of the original in-situ observations.

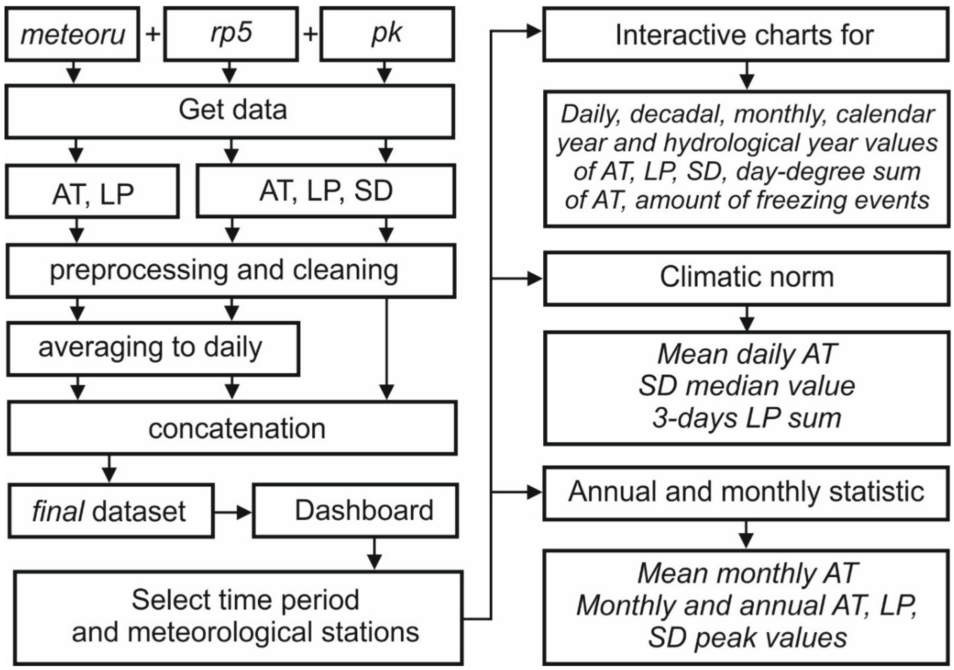


Figure 3. Schematic representation of the data preprocessing workflow used to compile the different meteorological datasets and brief description of the dashboard content. Note: AT – air temperature, LP – liquid precipitation, SD – snow depth.

4.1.1. Data harmonization

After loading the raw datasets, we converted them to *tibble* format, a special table format for *R*. We then detected an abundance of single records in the dataset, especially in the pk dataset. We decided to remove them from the final dataset due to the low reliability of such isolated observations. The best way to query the dataset is through the WMO ID, owing to wrong naming conventions across meteorological station names. Therefore, title standardization was performed through manual correction for each meteorological station.

4.1.2. Duplicates detection

Duplicate detection was performed across both meteorological parameter observation values (air temperature, humidity, precipitation, etc.) and timestamps. In the first case, this may indicate gap-filling in meteorological

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

time series by station staff or network operators. The causes of duplicate observations include instrument failures, illness or medical emergency, natural disasters, among others. Timestamp duplicates are evidence of incorrect data entry and deficiencies in dataset management practices (for example 29-07-2019 for *pk* dataset).

4.1.3. Air temperature quality control

Air temperature cleaning and quality control represented the most critical part of the preprocessing workflow. We began with a logical check of the *pk* dataset records — specifically, how many observations of minimum and maximum air temperature do not agree with the mean daily temperature. We found that up to 14% of all records at certain stations could represent minimum air temperature values exceeding the mean daily air temperature, and maximum air temperature values falling below the mean daily air temperature. These records were removed. *Meteoru* and *rp5* datasets provide extremum values of air temperature, but we did not take them into account, instead of that we derived maximum and minimum air temperature from terms daily observation.

Outlier detection in air temperature observations was the most challenging part of the preprocessing stage. Following Sherstyukov (2010), we designed a script for automatic statistical control of air temperature time series to detect potential outliers. Observations were flagged as potential outliers when they exceeded threshold values of 1.8, 2.0, or 2.5 standard deviations within a moving window of 21, 15, or 11 observations, for the *meteoru*, *rp5*, and *pk* datasets respectively. For each flagged value, we then searched for outlier dates among the 3 nearest neighboring stations. When no matching outlier dates were found within the lagged, current, or lead date, the value was considered erroneous and removed from the time series. In addition to detecting incorrect observations, this statistical control method is also suitable for typo detection.

4.1.4. Snow depth and liquid precipitation quality control

Snow depth time series control was aimed at removing spurious observations from the summer season and correcting wrong records. To address this, we checked each snow depth value for drastic increases and decreases at each timestep, and when this difference exceeded 20 cm, the value was removed from the time series. For liquid precipitation, we determined that a daily value cannot exceed 20% of the annual sum, and when such cases were identified, the corresponding values were also removed from the time series.

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

4.2. Dashboard

The dashboard is based on the *shiny* package with interactive charts from the *highcharter* package and is deployed on the *ShinyApps* hosting platform (temporary, subject to change). To produce decadal (10 days), monthly, and annual means, we used only complete time series; if even one observation in a selected time dimension is absent, the average value was not calculated. For the air temperature climate normal, we averaged the values for each particular day over the entire time series period; however, the result may be biased — for instance, if a meteorological station has more warm season observations than cold season observations. To preserve the 365-day climate normal time resolution, we decided to combine the 28th and 29th of February into a single date — the 28th of February. For the climatic normal of snow depth, we used the median value of that index over the entire time series to reduce the impact of rare anomalous events. For the liquid precipitation's climatic normal, we chose a 3-day temporal window to reduce noise in the chart and represent the general pattern of liquid precipitation distribution at each station.

5. Results

5.1. The final dataset

The *final* dataset consists of 3,688,256 daily records covering the period from 01-01-1966 to 30-09-2025 for 167 meteorological stations. However, only 2,596,754 records (70.4%) contain at least air temperature data. As a result, only 15 meteorological stations have continuous records over the entire study period: Aldan (31004), Anadyr (25563), Chokurdakh (21946), Chulman (30393), Isit (24951), Lensk (24923), Magadan (25913), Olekminsk (24944), Suntar (24738), Tiksi (21824), Ust-Maya (24966), Yakutsk (24959), Zhigansk (24343), Zhilindy (21908), and Zyryanka (25400). On the other hand, Ytyk-Kyuyol (24766) and Yekyuchyu (24361) show the lowest data availability, at 29.4% and 30.4%, respectively. Overall, 71 meteorological stations have more than 95% data availability, although not all of them cover the entire study

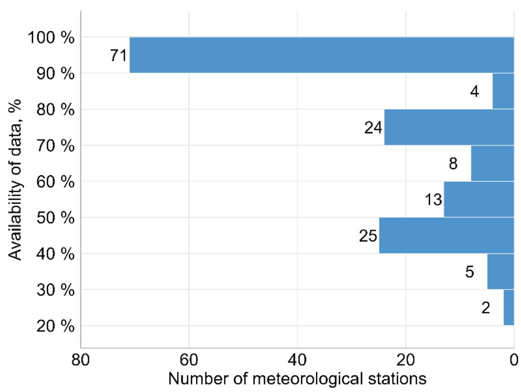


Figure 4. The availability of data for meteorological stations with not complete timeseries.

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

period. The detailed distribution of data availability is shown in Figure 4. Approximately one-third of the stations have lost more than 30% of their air temperature time series, while liquid precipitation and snow depth time series are in considerably worse condition. Daily maximum and minimum air temperatures were derived from daily observations rather than taken directly from the dataset, as the latter could not be verified. For some snow depth time series, linear extrapolation during the winter period was applied to fill small gaps.

Figure 5 shows the distribution of content sources and quality flags across the *final* dataset. The majority of the final dataset came from *meteoru*, while *rp5* and *pk*, despite their considerable differences in time series period and number of meteorological stations, contribute almost equally. Snow depth data were derived from *rp5* (41.5%) and from *pk* (58.5%). Data quality flags were added to better represent meteorological records and to preserve the user's ability to filter data. For instance, Macha-30074 has a 75% 4-terms/day observation regime and an 18% 3-terms/day observation regime, which is not suitable for contemporary requirements for meteorological observations. Consequently, 18.4% of records are flagged as reduced quality due to an incomplete number of observation terms during the day. Another quality flag captures the difference between the first and last observation terms within a day. This may indicate that part of the diurnal cycle was lost due to missed observation schedules. In practice, most stations show a 21-hour (observations every 3 hours) or 18-hour (observations every 6 hours) gap between the first and last daily observations, while only a minor portion may reflect erroneous meteorological data. Note that this quality flag was not applied to the pk dataset, as it does not contain this type of information. Overall, accounting for data absence and quality failures, approximately 60–65% of the North-East of Russia meteorological data time series from 1966 may be considered reliable for assessment.

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

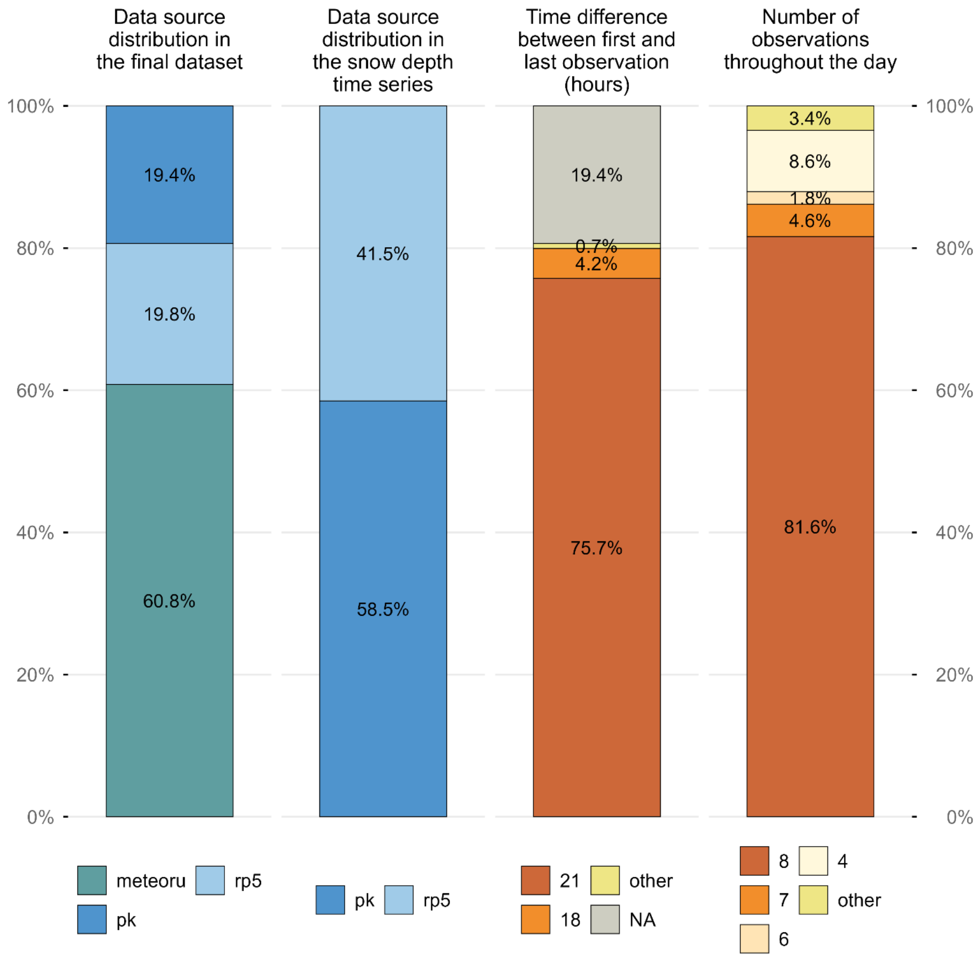


Figure 5. Distribution of data sources and quality flags throughout the *final* dataset for 167 meteorological stations from 01-01-1966 to 30-09-2025.

5.2. *Air temperature trend*

To analyze air temperature trends for the period from 1966 to 2024, we used only stations with continuous observation records. Thirteen of these stations are located in Yakutia, with one additional station in each of the Magadan Oblast and Chukotka. Prior to trend extraction, we tested the series for homogeneity — a justified step given that over a 60-year period, meteorological observation conditions may have changed substantially (Domonkos and Štěpánek, 2009; Killick et al., 2022). Homogeneity analysis identifies the breakpoint at which the difference between the mean values of the left and right segments of the time series is maximized (Alexandersson, 1986). Indeed, the homogeneity analysis divided the stations into two major

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

Open Access Earth System Science Data Discussions

subgroups — Arctic stations (breakpoint in April 2005) and continental stations (breakpoint in April 1988) — along with one subgroup affected by the 2001 flood (breakpoint in April 2001) (Fig. 6). Despite the similarity in patterns between the continental and flood-affected groups, the 2001 flood event (Kilmyaninov, 2001) and its clear signature in the homogeneity analysis of the Lensk and Isit' stations — both situated in the Lena River floodplain — confirms the validity of the applied methodology.

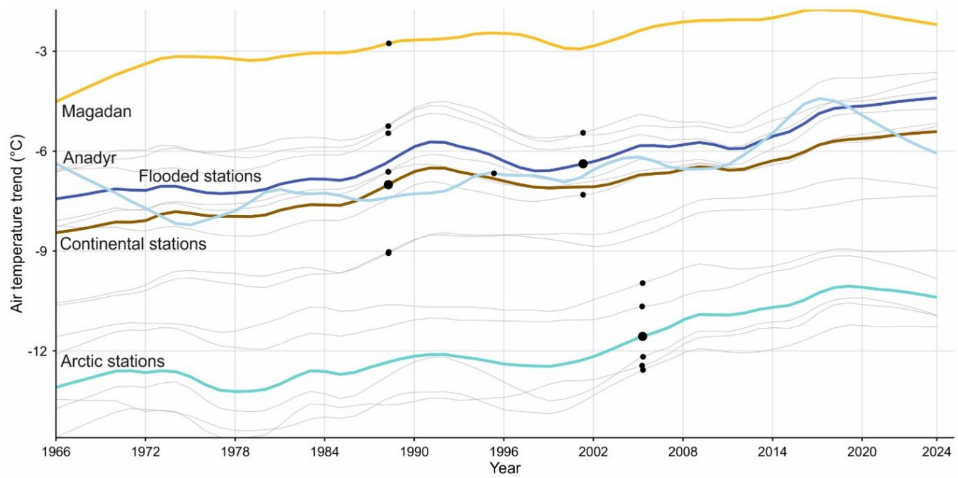


Figure 6. Air temperature trend for continuous time series from 01-01-1966 to 31-12-2024. Bold colored lines represent group averages and individual reference stations, while faint gray lines show the individual station series within each group. Black points indicate Pettitt breakpoints (the moment of a statistically significant shift in the average value) for particular stations and groups. The following groups and stations are displayed: Arctic stations (Tiksi – 21824, Chokurdakh – 21946, Zyryanka – 25400, Zhigansk – 24343, Zhilindy – 21908) Continental stations (Aldan – 31004, Chulman – 30393, Olekminsk – 24944, Suntar – 24738, Ust-Maya – 24966, Yakutsk – 24959), Flooded stations (Isit – 24951, Lensk – 24923).

Station Anadyr-25563 exhibits a distinctly individual behavior pattern, as it is strongly influenced by the waters of the Pacific Ocean. In contrast, Magadan-25913, located on the shore of the Sea of Okhotsk, shares the same breakpoint date as the continental stations in Yakutia. This may suggest a common influence of global warming on the climate of the broader macro-region. Based on the overall distribution of Pettitt breakpoints, it can be concluded that the time series of both Arctic and continental stations are homogeneous — even accounting for the undeniable urban heat island effect at station Yakutsk-24959. For the Arctic station group, the Pettitt breakpoint is shifted considerably to the left, clearly indicating a more pronounced temperature

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

increase over the past two decades. This finding once again supports the hypothesis of an amplified global warming signal at Arctic latitudes (Hu et al., 2021).

Trend extraction was performed by decomposing each time series into a trend component (10-year window), a seasonal component (365.25 days), and a residual (Fig. 6). Examining the flood-affected and continental station groups, several distinct phases can be identified: a period of stabilization (1972–1978), episodes of sharp acceleration (1985–1991 and 2010–2024), and a period of local trend decline (1991–1999). The trend for Arctic stations is steeper and exhibits a more prolonged period of temperature increase beginning in 2001. The northern Pacific coast (Anadyr-25563) is characterized by pronounced fluctuations in air temperature, whereas the Sea of Okhotsk region (Magadan-25913) displays a considerably smoother trend line.

Table 3. The average of mean air temperature for winter months (December, January, February) for initial and final periods, and air temperature trend. *p-value < 0.001, **p value < 0.005.

Meteorological stations and WMO ID	Mean air temperature for winter months for selected periods, hydrological years, °C				Overall trend*, °C/10 year
	1966-1971	**1966-1976**	2020-2025	**2016-2025**	
Aldan – 31004	-27.6	-26.7	-24.7	-24.6	0.301**
Chokurdakh – 21946	-34.0	-34.0	-31.8	-31.5	0.568
Chulman – 30393	-31.3	-30.8	-27.6	-27.4	0.385
Olekminsk – 24944	-31.2	-31.0	-28.1	-27.8	0.378**
Isit – 24951	-34.8	-34.7	-30.6	-30.5	0.567
Lensk – 24923	-29.6	-28.9	-26.1	-25.5	0.371**
Suntar – 24738	-33.1	-32.6	-29.5	-29.1	0.511
Tiksi – 21824	-30.8	-30.5	-28.1	-28.4	0.553
Zyryanka – 25400	-35.6	-35.9	-35.0	-34.0	0.467
Zhigansk – 24343	-37.5	-37.3	-36.2	-35.7	0.498
Zhilindy – 21908	-36.8	-36.8	-34.4	-34.1	0.699
Ust-Maya – 24966	-40.7	-39.7	-36.4	-36.5	0.534
Yakutsk – 24959	-40.4	-39.7	-35.3	-35.1	0.665
Anadyr - 25563	-19.6	-21.3	-20.5	-19.1	0.453
Magadan - 25913	-17.0	-16.2	-15.4	-15.3	0.342
Average	**-32.0**	**-31.7**	**-29.3**	**-29.0**	**0.52**

We also analyzed the numerical magnitude of the temperature trend alongside aggregated mean winter-month temperatures at the beginning and end of the study period (Table 3). Hydrological years were selected as the basis for this analysis, since the use of calendar years would have excluded December from

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

consideration. The majority of researchers have noted that the effects of global warming are most pronounced during winter months, a finding that is also confirmed by the meteorological network of northeastern Russia. Both short 5-year and more representative 10-year period comparisons yield an average difference in mean winter-month temperature of 2.7°C. Station Yakutsk-24959 stands out notably, with a difference exceeding 5°C — an increase undoubtedly driven by the urban heat island effect, given that the city's population has doubled over the study period. A less pronounced climatic signal in winter temperatures is observed at Anadyr-25563, where winters during 2020–2025 were on average colder than those of the late 1960s. The statistically significant temperature trend exceeds 0.3°C/decade at all stations, with a mean value of 0.52°C/decade across the 15 stations. This supports the conclusion that winter months make the dominant contribution to changes in mean annual temperature: at the derived trend magnitude, temperatures rose by 3.1°C over six decades, of which 2.7°C can be attributed to the winter months alone (IPCC, 2023b; Screen and Simmonds, 2010; Sun et al., 2024).

5.3. Liquid precipitation and snow depth data

Liquid precipitation and snow depth measurements are recorded at stations with considerable inconsistency (Zhilcova and Anisimov, 2009). Furthermore, the automatic weather stations that have been integrated into the observation network since the early 2010s do not capture these parameters at all. In the final dataset, we provide precipitation and snow depth values aggregated to daily resolution. Their analysis, quality assessment, and comparison are complicated by the low density of the meteorological network. We made every effort to remove obvious errors, such as snow depth records occurring during the summer period, abrupt jumps in snow depth values, and residual code entries in the format 999 or 9999 — indicating missing values. In its current form, the data can be used for a general characterization of the local climate at a given station; however, homogenization of the observation series is required before trend extraction or more detailed analysis can be conducted (Guijarro et al., 2023).

In the dashboard, these parameters are presented for each weather station, along with the calculated climatic normals. It should be noted, however, that both the length of the observation period and the quality of measurements vary between stations, which may lead to inconsistencies in the results.

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

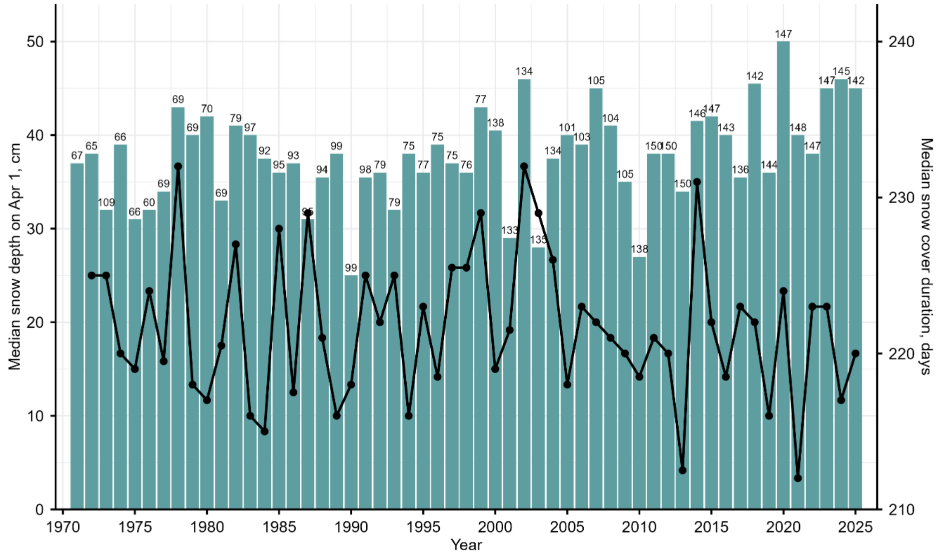


Figure 7. Median snow depth on April 1 (bars, left axis) and median snow cover duration (black line, right axis) across the study area for 1971–2025. The duration metric is computed for the August 1 – July 31 season and labelled by the year of season completion; only station-seasons with at least 200 days of non-missing snow depth observations were used. The 1971 season is omitted from the duration series because of incomplete coverage at the beginning of the dataset. Numbers above the bars indicate the number of stations with non-missing snow depth records on April 1.

To provide a regionally aggregated overview of snow conditions across the study area, we computed two complementary metrics for each year: the median snow depth on April 1 — close to the climatological maximum of seasonal snow accumulation in northeastern Russia — and the median duration of the snow cover season, defined as the number of days with *snow_cover* value above the zero within the August 1 – July 31 window and labelled by the year of season completion. The 1971 season was excluded from the duration series, because the dataset begins on 1 January 1971 and covers only the second half of that winter. To limit the influence of station-seasons with substantial gaps, only stations with at least 200 days of non-missing snow depth observations were retained for the duration calculation. Across the 1971–2025 period (Fig. 7), both metrics exhibit pronounced interannual variability superimposed on weak long-term tendencies: a slight increase in median April 1 snow depth and a slight decrease in median snow cover duration. This pattern is broadly consistent with the response of high-latitude snow regimes to ongoing warming, in which higher atmospheric moisture content tends to enhance winter snow accumulation while rising shoulder-season temperatures shorten the period of continuous snow cover

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

(e.g., Bulygina et al., 2011; Brown and Mote, 2009). Given the heterogeneity of the network (flat, mountain, continental, coastline areas) and the absence of formal homogenization, these tendencies should be interpreted as a qualitative regional signal rather than as quantitative trend estimates.

6. Discussion

6.1. Meteorological parameters quality check and some disadvantages

Despite comprising a smaller number of stations, the *meteoru* dataset accounts for the largest share of the final merged dataset. This is attributable to its longest observation period and the highest overall data quality among the sources used (Parfenova and Tchebakova, 2023). However, the majority of stations included in the *meteoru* dataset are located in urban areas, which may significantly affect measurement accuracy due to artificial shading and the urban heat island effect (Magee et al., 1999; Oke, 1982). In addition, the *meteoru* dataset does not include snow depth records. Furthermore, at the time of writing, the administration of the World Data Centers (WDC) in Russia (http://meteo.ru/data/) has reduced the number of available stations from the original 521 to 134, with coverage for the study region now not exceeding 15 stations.

To compare the *final* dataset with widely used meteorological databases, we selected the Copernicus solution at daily and sub-daily temporal resolution and the Northern Hemisphere dataset of Tran et al. (Copernicus Climate Change Service, 2026; Tran et al., 2024). Both reference datasets contain up to 10% duplicated stations within the study area; in the Tran et al. dataset, individual stations occasionally appear in up to three copies. The Tran et al. dataset additionally includes water-balance gauges (identified by 4-digit codes), at which meteorological observations are conducted under a reduced protocol. Furthermore, after a careful and thorough cross-check, we did not identify any unique meteorological stations in the Tran et al. dataset relative to the *final* and Copernicus datasets.

To illustrate the relative usefulness and temporal completeness of each dataset, we compared the number of stations providing air temperature time series across the study period from 1966 to 2024 (Fig. 8). For the *final* dataset, two data-availability thresholds were applied (≥80% and ≥99% of daily records present per year), whereas for the Copernicus and Tran et al. datasets all non-empty time series were counted. Overall, the Tran et al. dataset and the daily Copernicus dataset display nearly identical station counts throughout the period, varying between 68 and 107 and exhibiting a gradual decline towards the present, with both falling below 75 stations after 2021. The *final* dataset

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

surpasses both reference products from the late 1970s onward, which confirms the value of enthusiast-maintained repositories such as *rp5.ru* and *pogodaiklimat.ru* in supplementing the official record. A pronounced dip in the early 1990s is visible in the *final* dataset and is consistent with the post-Soviet contraction of the state observation network (Groisman et al., 2014). A recovery begins in the mid-2000s, followed by a sharp increase after 2009 associated with overall socio-economic improvement. The persistent gap between the ≥80% and ≥99% curves indicates that a non-negligible fraction of station-years contains intra-annual gaps, although this gap narrows in the most recent decade, suggesting an overall improvement in data completeness. The sub-daily Copernicus dataset is the least suitable option for the study area in terms of long-term continuity, as station counts collapse from ~170 in 1969–1971 to fewer than 20 stations by the mid-1990s and remain at that level thereafter; nevertheless, the unique stations available in the early period (1969–1971) constitute a potentially valuable resource for machine-learning-based gap-filling of the historical time series.

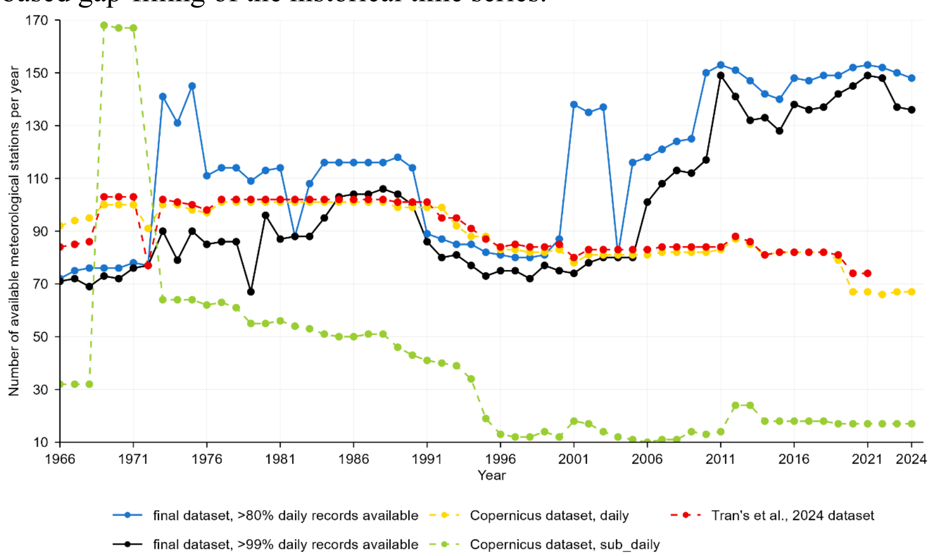


Figure 8. Number of unique meteorological stations available per year from 1966 to 2024 in the *final* dataset, the Copernicus daily and sub-daily datasets, and the Tran et al. (2024) dataset. For the *final* dataset, two data-availability thresholds are shown (≥80% and ≥99% of daily air temperature records present per year); for the Copernicus and Tran et al. datasets, all non-empty annual time series are counted regardless of completeness.

The scripted error detection approach applied in this study is not without limitations and is not particularly demanding in terms of computational resources. A viable alternative would have been the use of the *climatol* package; however, its authors explicitly note its limitations when applied to

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

synoptic observations recorded six or eight times per day (Guijarro et al., 2023). In our case, the priority was to preserve as much of the available data as possible. Nevertheless, we plan to carry out homogenization of precipitation totals in a future update, which will subsequently be reflected in the dashboard. It should also be acknowledged that isolated errors have a negligible effect once data are aggregated to decadal and monthly timescales, which renders both the original datasets and our merged product suitable for climatic characterization of the region (Gorokhov and Fedorov, 2018).
It is also important to recognize the inherent methodological limitations of obtaining meteorological information (Handbook of Meteorological Instruments…, 1969). Observations at weather stations are conducted at a single measurement point, and given the very low density of the station network in the study region, the resulting picture of climatic conditions may be distorted (Anisimov and Orttung, 2019; Sherstyukov and Sherstyukov, 2015). A clear illustration of this is the climatological snow depth distribution. It is well established that the Magadan-25913 region, situated along the Sea of Okhotsk, is characterized by substantial snow depths, whereas central Yakutia — and Yakutsk-24959 in particular — belongs to a zone of low precipitation, both liquid and solid (Bednorz, 2016). Yet when comparing median snow depth values, Yakutsk-24959 exceeds Magadan-25913. This discrepancy may be partly explained using the median statistic, which suppresses the contribution of the heavy snowfall events characteristic of Magadan-25913, but is more likely attributable to the high exposure of the Magadan-25913 observation site to wind and the generally windier conditions in that region. These considerations underscore the need for a critical approach when interpreting the climatic characteristics of the study region.
A further potential limitation is the exclusion of stations located outside the boundaries of the study region. Many of these stations have more complete observation records, which could have assisted in the quality control of time series at the target stations. This will be addressed in future expansions of the final dataset. An additional weakness lies in the application of the quality control and error detection methodology separately to each of the three original datasets, which individually contain fewer stations than the merged final product. This results in greater distances between neighboring stations, which in turn reduces the precision of error and outlier detection.
A common objective is the general scarcity of data. The figures presented in Table 2 may be misleading, as they do not account for the full extent of the potentially available observation period. The quality of the data also warrants

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

scrutiny: more than 20% of air temperature observations do not meet contemporary meteorological standards and are therefore considered non-compliant. The situation is considerably worse for other meteorological parameters. Across the entire dataset, fewer than 10% of stations maintain a continuous observation record dating back to 1966, and only 4 of these are located outside of major populated areas — a circumstance that severely limits a comprehensive understanding of the natural climatic conditions of such an extensive region (Gorokhov and Fedorov, 2018).

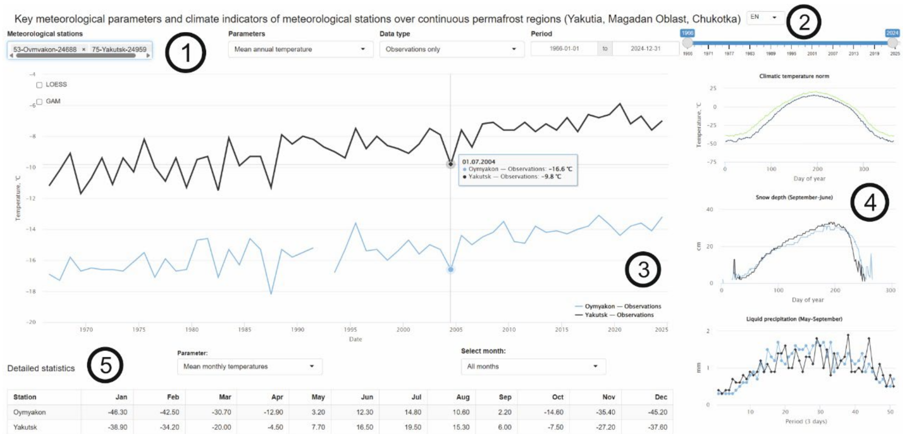
Figure 9. Spatial distribution of all meteorological stations across the study area, compiled from the final dataset (current article), the Copernicus daily dataset, and the Copernicus sub-daily dataset. For each station, the source providing the longest continuous time series is shown. Station numbers correspond to the entries in Table S5, where additional metadata are provided. Note: I – Yakutia, II – Magadan Oblast, III – Chukotka.

To assess the spatial coverage of meteorological observations across northeastern Eurasia, we mapped all unique stations from the three reference datasets described above (Fig. 9). After duplicate removal and time-series consolidation, the combined inventory comprises 238 meteorological stations, of which 151 (63.5%), 65 (27.3%), and 22 (9.2%) originate from the *final* dataset and from the Copernicus daily and sub-daily datasets, respectively (Table S5). The resulting observation density for the study region reaches approximately one station per 18,000 km², which is not comparable even with another one Russia region — for example, the Kaliningrad region maintains 10 meteorological stations across 15,000 km². The spatial distribution is, in

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

addition, strongly inhomogeneous: stations are concentrated along the major river valleys (Lena, Indigirka, Kolyma) and along the southern coast and inland transport corridors, whereas vast inland areas of central Yakutia and the interior Chukotka Peninsula remain effectively unobserved. To improve the representativeness of the dataset, we plan to incorporate all remaining unique stations from Copernucus dataset identified through this comparison into the *final* dataset and to update the accompanying dashboard accordingly.

6.2. *The dashboard solution*



Station	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Oymyakon	-46.30	-42.50	-30.70	-12.90	3.20	12.30	14.80	10.60	2.20	-14.60	-35.40	-45.20
Yakutsk	-38.90	-34.20	-20.00	-4.50	7.70	16.50	19.50	15.30	6.00	-7.50	-27.20	-37.60

Figure 10. Main components of the dashboard. 1 – control panel for selecting meteorological stations, climate parameters, data type (observations or modelled data), and time period; 2 – language switcher (Russian and English available); 3 – main interactive time series chart displaying selected parameter for chosen stations, with an interactive tooltip, a legend, and optional LOESS and GAM smoothing overlays; 4 – three supplementary interactive charts showing climatological norms for the selected stations over the full available period: mean annual temperature cycle, snow depth (September–June), and liquid precipitation (May–September); 5 – detailed statistics table values for the selected parameter across all chosen stations, with additional facilities. A map for spatial navigation across the study area is located below the table.

Usability is as important as quality for climate data (Sarikaya et al., 2019; Yigitbasioglu and Velcu, 2012). Researchers debate reanalysis accuracy, while non-specialists rely on the most accessible indicators (Biskaborn et al., 2019). For our region, comprehensive climatic information is hard to obtain: beyond data-quality issues, it is fragmented across three structurally different databases. The most convenient, pogodaiklimat.ru, supplies pre-aggregated results of questionable quality and still requires reformatting before temporal change can be assessed.

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

To simplify access, we developed an interactive dashboard delivering user-defined aggregates of air temperature, liquid precipitation, and snow depth in both graphical and numerical form (charts, tables). It also supports direct comparison of climatic zones; the two-station limit is adjustable but balances readability of the graphical output.
Stations are selected by WMO ID or name, with 40 indicators currently available. A data-type toggle switches between raw observations and gap-filled reconstructions, and optional LOESS and GAM overlays reveal long-term trends in the main time-series panel. Smaller plots show climatic normals — daily mean temperature (365 days), liquid precipitation (May–September), and snow depth (October–May) — while a lower panel reports summary statistics and an integrated map supports spatial navigation across the study area.
The *final* dataset may retain minor errors, chiefly in maximum and minimum temperature. We deferred comparison with reanalysis products, as suitable evaluation criteria remain unclear. Planned developments include more stations and broader coverage (West Siberia and the Qinghai–Tibet Plateau, China), biannual updates (March–April and October–November), and a possible mobile version. Comments on both the dataset and the dashboard are welcome.

Data availability

The *final* dataset and is available at https://doi.org/10.5281/zenodo.19883706 (Sysoliatin, 2026). The R preprocessing pipeline (meteoru_preproc_pipeline.R) and its documentation (README.md) are included in the repository root.

7. Conclusions

This article describes the steps undertaken to consolidate meteorological data from the station network of northeastern Russia — a remote and climatically severe region. Despite the considerable challenges involved in working with three distinct data sources, the final dataset delivers the following outcomes:

1. None of the raw datasets individually can provide complete, high-quality meteorological records for North-East Eurasia. The *meteoru* dataset covers fewer than half of the available stations; *rp5* holds records only from 2005 onward and its time series are of comparatively lower quality; and *pk* is a preprocessed dataset whose methodology is not transparent.

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

Only the integration of these heterogeneous sources, combined with a multi-stage preprocessing workflow, makes it possible to obtain a coherent picture of the regional climate and its ongoing changes. The *final* dataset encompasses key climatic indicators spanning the period from 1 January 1966 to 30 September 2025 and will be updated on an ongoing basis.

2. Time series of key meteorological parameters — daily mean air temperature, daily precipitation totals, and snow depth — have been compiled for the period from 1 January 1966 to 30 September 2025. Each observation is accompanied by an indication of its source and an objective assessment of data quality.
3. A comprehensive evaluation of data availability and quality across the meteorological observation network has been carried out. The greatest data losses occurred during the 1990s. Of the entire dataset, only 15 stations maintain a continuous air temperature record. The quality and completeness of precipitation and snow depth data are considerably poorer.
4. The magnitude of observed climatic change varies across the region; however, the overall trend points toward a widespread increase in air temperature. At the current rate of warming, positive mean annual temperatures may begin to be recorded at individual stations within the permafrost zone.
5. An interactive system for accessing the *final* dataset has been developed and launched, enabling rapid retrieval of information on 40 aggregated meteorological parameters from 167 stations across the region. This system provides the detailed meteorological information required by a wide range of users and researchers.

We invite the scientific community to refine our approach and contribute missing data from currently inactive meteorological stations.

Supplement. The supplement is published alongside this article and is available online. The dashboard is available at https://robertseesaw.shinyapps.io/meteo-dashboard/. The preprocessing *R*-scripts and sub-daily data of each source datasets are available upon request from Robert Sysoliatin.

Author contributions. RS wrote the first draft of the paper with contributions from ZW, QG, KX, and YT. The preprocessing and analysis of the meteorological observations were performed by RS, and YT, and the quality check and methodic improved were performed by RS, ZW, QG, KX, and YT.

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

RS, ZW, QG, and KX constructed the combined gridded analysis. All of the authors reviewed the final text.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Acknowledgements. The authors thank colleagues at Northeastern Institute of Eco-Environment Research (CAS) and Melnikov Permafrost Institute (RAS) for their valuable comments and advice during the preparation of the *final* dataset. Special thanks are due to Tonghua Wu and Xiaofan Zhu for their assistance in organizing the workflow. Special gratitude is extended to Professor Lin Chen for the keyboard.

Financial support. This research project was supported by the Major Science and Technology Project of Gansu Province (23ZDWA005 and 24ZD13FA003), National Natural Science Foundation of China (42371140 and 42272332) and the program of the State Key Laboratory of Cryospheric Science and Frozen Soil Engineering, CAS (CSFSE-ZZ-2411 and CSFSEFX-2502). This work was also supported as part of the main project of the Melnikov Permafrost Institute, "Deep ground temperature and permafrost state of the Asian part of Russia. Response of the thermal regime of permafrost to climate change" (Registration No. 126020216324-9; led by corresponding member of the Siberian Branch, Russian Academy of Sciences, Dr. M.N. Zheleznyak).

References

Alexandersson, H.: A homogeneity test applied to precipitation data, Journal of Climatology, 6, 661–675, https://doi.org/10.1002/joc.3370060607, 1986.

Almeida, I. C. C., Schaefer, C. E. G. R., Fernandes, R. B. A., Pereira, T. T. C., Nieuwendam, A., and Pereira, A. B.: Active layer thermal regime at different vegetation covers at Lions Rump, King George Island, Maritime Antarctica, Geomorphology, 225, 36–46, https://doi.org/10.1016/j.geomorph.2014.03.048, 2014.

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

Anisimov, O. and Orttung, R.: Climate change in Northern Russia through the prism of public perception, Ambio, 48, 661–671, https://doi.org/10.1007/s13280-018-1096-x, 2019.

Anisimov, O. and Reneva, S.: Permafrost and changing climate: The Russian perspective, Ambio, 35, 169–175, https://doi.org/10.1579/0044-7447(2006)35[169:PACCTR]2.0.CO;2, 2006.

Anisimov, O. A., Anokhin, Yu. A., Lavrov, S. A., Malkova, G. V., Myach, L. T., Pavlov, A. V., Romanovsky, V. A., Streletskiy, D. A., Kholodov, A. L., and Shiklomanov, N. I.: Methods for assessing the consequences of climate change for physical and biological systems, Chap. 8: Continental permafrost, Roshydromet, Moscow, Russia, 301–359, 2012 (in Russian).

Bednorz, E.: Seasonal cycle of snow cover changes in Eastern Siberia and its synoptic preconditions, Russian Meteorology and Hydrology, 41, 648–656, https://doi.org/10.3103/S1068373916090089, 2016.

Bertrand, C., Ingels, R., and Journée, M.: Homogenization and trends analysis of the Belgian historical precipitation time series, International Journal of Climatology, 41, 5277–5294, https://doi.org/10.1002/joc.7129, 2021.

Biskaborn, B. K., Smith, S. L., Noetzli, J., Matthes, H., Vieira, G., Streletskiy, D. A., Schoeneich, P., Romanovsky, V. E., Lewkowicz, A. G., Abramov, A., Allard, M., Boike, J., Cable, W. L., Christiansen, H. H., Delaloye, R., Diekmann, B., Drozdov, D., Etzelmüller, B., Grosse, G., Guglielmin, M., Ingeman-Nielsen, T., Isaksen, K., Ishikawa, M., Johansson, M., Johannsson, H., Joo, A., Kaverin, D., Kholodov, A., Konstantinov, P., Kröger, T., Lambiel, C., Lanckman, J.-P., Luo, D., Malkova, G., Meiklejohn, I., Moskalenko, N., Oliva, M., Phillips, M., Ramos, M., Sannel, A. B. K., Sergeev, D., Seybold, C., Skryabin, P., Vasiliev, A., Wu, Q., Yoshikawa, K., Zheleznyak, M., and Lantuit, H.: Permafrost is warming at a global scale, Nat. Commun., 10, 264, https://doi.org/10.1038/s41467-018-08240-4, 2019.

Brown, J., Ferrians, O. J., Heginbottom, J. A., and Melnikov, E. S.: Circum-Arctic map of permafrost and ground-ice conditions, https://doi.org/10.3133/cp45, 1997.

Cannone, N., Ellis Evans, J. C., Strachan, R., and Guglielmin, M.: Interactions between climate, vegetation and the active layer in soils at two Maritime Antarctic sites, Antarct. Sci., 18, 323–333, https://doi.org/10.1017/S095410200600037X, 2006.

Copernicus Climate Change Service, Climate Data Store, (2021): Global land surface atmospheric variables from 1755 to present from comprehensive in-situ

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

observations. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). (Accessed on 01-05-2026), 10.24381/cds.cf5f3bac

Cowtan, K. and Way, R. G.: Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends, Quarterly Journal of the Royal Meteorological Society, 140, 1935–1944, https://doi.org/10.1002/qj.2297, 2014.

Domonkos, P. and Štěpánek, P.: Statistical characteristics of detectable inhomogeneities in observed meteorological time series, Studia Geophysica et Geodaetica, 53, 239–260, https://doi.org/10.1007/s11200-009-0015-9, 2009.

Evans, M. N., Lücke, L. J., Fan, K. J., and Zhu, F.: A database of databases for Common Era paleoclimate applications, Earth Syst. Sci. Data, 18, 1185–1202, https://doi.org/10.5194/essd-18-1185-2026, 2026.

Fedorov, A. N., Iwahana, G., Konstantinov, P. Y., Machimura, T., Argunov, R. N., Efremov, P. V., Lopez, L. M. C., and Takakai, F.: Variability of Permafrost and Landscape Conditions Following Clear Cutting of Larch Forest in Central Yakutia, Permafr. Periglac. Process., 28, 331–338, https://doi.org/10.1002/ppp.1897, 2017.

Fedorov, A., Vasilyev, N., Torgovkin, Y., Shestakova, A., Varlamov, S., Zheleznyak, M., Shepelev, V., Konstantinov, P., Kalinicheva, S., Basharin, N., Makarov, V., Ugarov, I., Efremov, P., Argunov, R., Egorova, L., Samsonova, V., Shepelev, A., Vasiliev, A., Ivanova, R., Galanin, A., Lytkin, V., Kuzmin, G., and Kunitsky, V.: Permafrost-Landscape Map of the Republic of Sakha (Yakutia) on a Scale 1:1,500,000, Geosciences (Basel)., 8, 465, https://doi.org/10.3390/geosciences8120465, 2018.

Fiebrich, C. A., Morgan, C. R., McCombs, A. G., Hall, P. K., and McPherson, R. A.: Quality Assurance Procedures for Mesoscale Meteorological Data, J. Atmos. Ocean. Technol., 27, 1565–1582, https://doi.org/10.1175/2010JTECHA1433.1, 2010.

Folland, C. K., Colman, A. W., Smith, D. M., Boucher, O., Parker, D. E., and Vernier, J.: High predictive skill of global surface temperature a year ahead, Geophys. Res. Lett., 40, 761–767, https://doi.org/10.1002/grl.50169, 2013.

Gong, D.-Y. and Ho, C.-H.: The Siberian High and climate change over middle to high latitude Asia, Theor. Appl. Climatol., 72, 1–9, https://doi.org/10.1007/s007040200008, 2002.

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

Gorokhov, A. N. and Fedorov, A. N.: Current Trends in Climate Change in Yakutia, Geography and Natural Resources, 39, 153–161, https://doi.org/10.1134/S1875372818020087, 2018.

Grigorieva, E.: Evaluating the Sensitivity of Growing Degree Days as an Agro-Climatic Indicator of the Climate Change Impact: A Case Study of the Russian Far East. Atmosphere, 11(4), 404. https://doi.org/10.3390/atmos11040404Guijarro, J. A., López, J. A., Aguilar, E., Domonkos, P., Venema, V. K. C., Sigró, J., and Brunet, M.: Homogenization of monthly series of temperature and precipitation: Benchmarking results of the MULTITEST project, International Journal of Climatology, 43, 3994–4012, https://doi.org/10.1002/joc.8069, 2023.

Handbook of the USSR climate. Air humidity, precipitation, snow cover. In 34 issues. Iss. 16. Part 4). Leningrad: Gidrometeoizdat, 1969. 195 p. (in Russian)

Hu, G., Zhao, L., Wu, T., Wu, X., Park, H., Fedorov, A., Wei, Y., Li, R., Zhu, X., Sun, Z., Ni, J., and Zou, D.: Spatiotemporal variations and regional differences in air temperature in the permafrost regions in the Northern Hemisphere during 1980–2018, Science of The Total Environment, 791, 148358, https://doi.org/10.1016/j.scitotenv.2021.148358, 2021.

Intergovernmental Panel on Climate Change (IPCC): Climate Change 2021 – The Physical Science Basis, Cambridge University Press, https://doi.org/10.1017/9781009157896, 2023a.

Intergovernmental Panel on Climate Change (IPCC): Climate Change 2021 – The Physical Science Basis, Cambridge University Press, https://doi.org/10.1017/9781009157896, 2023b.

Ivanova, R. N.: Extremely low air temperatures in Eurasia, Vestnik YSU, 3, 13–19, 2006 (in Russian).

Jones, P. D. and Moberg, A.: Hemispheric and Large-Scale Surface Air Temperature Variations: An Extensive Revision and an Update to 2001, J. Clim., 16, 206–223, https://doi.org/10.1175/1520-0442(2003)016<0206:HALSSA>2.0.CO;2, 2003.

Kääb, A., Leinss, S., Gilbert, A., Bühler, Y., Gascoin, S., Evans, S. G., Bartelt, P., Berthier, E., Brun, F., Chao, W.-A., Farinotti, D., Gimbert, F., Guo, W., Huggel, C., Kargel, J. S., Leonard, G. J., Tian, L., Treichler, D., and Yao, T.: Massive collapse of two glaciers in western Tibet in 2016 after surge-like instability, Nat. Geosci., 11, 114–120, https://doi.org/10.1038/s41561-017-0039-7, 2018.

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

Kalinicheva, S., Fedorov, A., and Zheleznaik, M.: Mapping Mountain Permafrost Landscapes in Siberia Using Landsat Thermal Imagery, Geosciences (Basel)., 9, 4, https://doi.org/10.3390/geosciences9010004, 2018.

Killick, R. E., Jolliffe, I. T., and Willett, K. M.: Benchmarking the performance of homogenization algorithms on synthetic daily temperature data, International Journal of Climatology, 42, 3968–3986, https://doi.org/10.1002/joc.7462, 2022.

Kilmyaninov, V. V. "Disastrous flood on the Lena River near Lensk in 2001. Russian Meteorology and Hydrology 12 (2001): 55-58. (in Russian)

Kirillin, A. R., Zheleznaik, M. N., and Zhizhin, V. I.: New data on thickness of permafrost at the Lena-Aldan watershed, Earth's Cryosphere, 26, 3–11, https://doi.org/10.15372/KZ20220301, 2022.

Kirillina, K., Tananaev, N., Savvinova, A., Lobanov, V., Fedorova, A., and Borisov, A.: Climate change impacts the state of winter roads connecting indigenous communities: Case study of Sakha (Yakutia) Republic, Clim. Serv., 30, 100356, https://doi.org/10.1016/j.cliser.2023.100356, 2023.

Kokorev, V. A. and Sherstyukov, A. B.: On meteorological data for studying contemporary and future climate change over the territory of Russia, Arktika XXI Vek. Estestvennye Nauki, 2(3), 5-23, 2015 (in Russian).

Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F.: World Map of the Köppen-Geiger climate classification updated, Meteorologische Zeitschrift, 15, 259–263, https://doi.org/10.1127/0941-2948/2006/0130, 2006.

Kuklina, V. V., & Osipova, M. E. (2018). The role of winter roads in ensuring transport accessibility of the Arctic and subarctic regions of the Republic of Sakha (Yakutia). *Soc. Environ. Dev*, *2*, 107-112. (in Russian)

Leibman, M. O., Kizyakov, A. I., Plekhanov, A. V., and Streletskaya, I. D.: New permafrost feature – deep crater in central Yamal (West Siberia, Russia) as a response to local climate fluctuations, Geogr. Environ. Sustain., 7, 68–79, https://doi.org/10.24057/2071-9388-2014-7-4-68-79, 2014.

Luterbacher, J., Allan, R., Wilkinson, C., Hawkins, E., Teleti, P., Lorrey, A., Brönnimann, S., Hechler, P., Velikou, K., and Xoplaki, E.: The Importance and Scientific Value of Long Weather and Climate Records; Examples of Historical Marine Data Efforts across the Globe, Climate, 12, 39, https://doi.org/10.3390/cli12030039, 2024.

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

Lytkin, V., Suleymanov, A., Vinokurova, L., Grigorev, S., Golomareva, V., Fedorov, S., Kuzmina, A., and Syromyatnikov, I.: Influence of Permafrost Landscapes Degradation on Livelihoods of Sakha Republic (Yakutia) Rural Communities, Land (Basel)., 10, 101, https://doi.org/10.3390/land10020101, 2021.

Magee, N., Wendler, G., and Curtis, J.: The Urban Heat Island Effect at Fairbanks, Alaska, Theor. Appl. Climatol., 64, 39–47, https://doi.org/10.1007/s007040050109, 1999.

Malkova, G.V., Pavlov, A.V., Skachkov, Yu. B.: Assessment of permafrost stability under contemporary climatic changes. Eaths's Cryosphere., XV (4), 29-32.

Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E., and Houston, T. G.: An Overview of the Global Historical Climatology Network-Daily Database, J. Atmos. Ocean. Technol., 29, 897–910, https://doi.org/10.1175/JTECH-D-11-00103.1, 2012.

Mingalev, D. E.: Agroclimatic Zoning of Russia and Kazakhstan under Current Climate Change, Geography and Natural Resources, 42, 115–121, https://doi.org/10.1134/S1875372821020086, 2021.

Obu, J., Westermann, S., Bartsch, A., Berdnikov, N., Christiansen, H. H., Dashtseren, A., Delaloye, R., Elberling, B., Etzelmuller, B., Kholodov, A., Khomutov, A., Kaab, A., Leibman, M. O., Lewkowicz, A. G., Panda, S. K., Romanovsky, V., Way, R. G., Westergaard-Nielsen, A., Wu, T. H., Yamkhin, J., and Zou, D. F.: Northern Hemisphere permafrost map based on TTOP modelling for 2000-2016 at 1 km(2) scale, Earth. Sci. Rev., 193, 299–316, https://doi.org/10.1016/j.earscirev.2019.04.023, 2019.

Oke, T. R.: The energetic basis of the urban heat island, Quarterly Journal of the Royal Meteorological Society, 108, 1–24, https://doi.org/10.1002/qj.49710845502, 1982.

Oliva, M., Hrbacek, F., Ruiz-Fernández, J., de Pablo, M. Á., Vieira, G., Ramos, M., and Antoniades, D.: Active layer dynamics in three topographically distinct lake catchments in Byers Peninsula (Livingston Island, Antarctica), Catena (Amst)., 149, 548–559, https://doi.org/10.1016/j.catena.2016.07.011, 2017.

Parfenova, E. I. and Tchebakova, N. M.: Potential Forest Distribution over the South Siberian and North Mongolian Mountains Related to Predicted Climate Change by the Midcentury, Izvestiya Rossiiskoi Akademii Nauk Seriya Geograficheskaya, 87, 1019–1031, https://doi.org/10.31857/S2587556623070129, 2023.

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

Posit team (2025). RStudio: Integrated Development Environment for R. Posit Software, PBC, Boston, MA. URL http://www.posit.co/.

Sabatini, F.: Setting up and Managing Automatic Weather Stations for Remote Sites Monitoring: From Niger to Nepal, 21–39, https://doi.org/10.1007/978-3-319-59096-7_2, 2017.

Sarikaya, A., Correll, M., Bartram, L., Tory, M., and Fisher, D.: What Do We Talk About When We Talk About Dashboards?, IEEE Trans. Vis. Comput. Graph., 25, 682–692, https://doi.org/10.1109/TVCG.2018.2864903, 2019.

Screen, J. A. and Simmonds, I.: The central role of diminishing sea ice in recent Arctic temperature amplification, Nature, 464, 1334–1337, https://doi.org/10.1038/nature09051, 2010.

Sherstyukov, A. B. and Sherstyukov, B. G.: Spatial features and new trends in thermal conditions of soil and depth of its seasonal thawing in the permafrost zone, Russian Meteorology and Hydrology, 40, 73–78, https://doi.org/10.3103/S1068373915020016, 2015.

Simmons, A. J., Willett, K. M., Jones, P. D., Thorne, P. W., and Dee, D. P.: Low-frequency variations in surface atmospheric humidity, temperature, and precipitation: Inferences from reanalyses and monthly gridded observational data sets, Journal of Geophysical Research: Atmospheres, 115, https://doi.org/10.1029/2009JD012442, 2010.

Stephenson, S. R., Smith, L. C., and Agnew, J. A.: Divergent long-term trajectories of human access to the Arctic, Nat. Clim. Chang., 1, 156–160, https://doi.org/10.1038/nclimate1120, 2011.

Streletskiy, D. A., Maslakov, A., Grosse, G., Shiklomanov, N. I., Farquharson, L., Zwieback, S., Iwahana, G., Bartsch, A., Liu, L., Strozzi, T., Lee, H., and Debolskiy, M. V: Thawing permafrost is subsiding in the Northern Hemisphere—review and perspectives, Environmental Research Letters, 20, 013006, https://doi.org/10.1088/1748-9326/ada2ff, 2025.

Streletskiy, D. A., Suter, L. J., Shiklomanov, N. I., Porfiriev, B. N., and Eliseev, D. O.: Assessment of climate change impacts on buildings, structures and infrastructure in the Russian regions on permafrost, Environmental Research Letters, 14, 025003, https://doi.org/10.1088/1748-9326/aaf5e6, 2019.

Sun, F., Li, Y., Chen, Y., Fang, G., Duan, W., Li, B., Li, Z., Hao, X., Yang, Y., and Zhang, X.: The dominant warming season shifted from winter to spring in the arid

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

Open Access Earth System Science Data Discussions

region of Northwest China, NPJ Clim. Atmos. Sci., 7, 178, https://doi.org/10.1038/s41612-024-00724-z, 2024.

Swift, L. W. and Ragsdale, H. L.: Meteorological Data Stations at Long-Term Ecological Research Sites, in: The Forest-Atmosphere Interaction, Springer Netherlands, Dordrecht, 25–37, https://doi.org/10.1007/978-94-009-5305-5_2, 1985.

Sysolyatin, R., Serikov, S., Zheleznyak, M., Tikhonravova, Y., Skachkov, Y., Zhizhin, V., and Rojina, M.: Temperature monitoring from 2012 to 2019 in central part of Suntar-Khayat Ridge, Russia, J. Mt. Sci., 17, 2321–2338, https://doi.org/10.1007/s11629-020-6175-3, 2020.

Sysoliatin, R. (2026). Air temperature, precipitation and snow depth for 167 meteorological station across Republic of Sakha (Yakutia), Magadan Oblast, and Chukotka [Data set]. In Earth System Science Data. Zenodo. https://doi.org/10.5281/zenodo.19883706

Tran, V. N., Zhou, W., Kim, T., Mazepa, V., Valdayskikh, V., and Ivanov, V. Y.: Daily station-level records of air temperature, snow depth, and ground temperature in the Northern Hemisphere, Sci. Data, 11, 645, https://doi.org/10.1038/s41597-024-03483-x, 2024.

Varlamov, S. P., Skachkov, Y. B., and Skryabin, P. N.: Evolution of the thermal state of permafrost under climate warming in Central Yakutia, Holocene, 29, 1401–1410, https://doi.org/10.1177/0959683619855959, 2019.

Veselkin, D. V., Kuyantseva, N. B., Chashchina, O. E., Mumber, A. G., Zamshina, G. A., and Molchanova, D. A.: Levels of Leaf Damage by Phyllophages in Invasive Acer negundo and Native Betula pendula and Salix caprea, Russ. J. Ecol., 50, 511–516, https://doi.org/10.1134/S1067413619060134, 2019.

Wang, K., Jafarov, E., Overeem, I., Romanovsky, V., Schaefer, K., Clow, G., Urban, F., Cable, W., Piper, M., Schwalm, C., Zhang, T., Kholodov, A., Sousanes, P., Loso, M., and Hill, K.: A synthesis dataset of permafrost-affected soil thermal conditions for Alaska, USA, Earth Syst. Sci. Data, 10, 2311–2328, https://doi.org/10.5194/essd-10-2311-2018, 2018.

WMO: Resolution 40 (Cg-XII) – WMO policy and practice for the exchange of meteorological and related data and products including guidelines on relationships in commercial meteorological activities, in: Twelfth World Meteorological Congress: Abridged Final Report with Resolutions, WMO-No. 827, World Meteorological Organization, Geneva, Switzerland, 1995.

https://doi.org/10.5194/essd-2026-346
Preprint. Discussion started: 15 July 2026

WMO: Resolution 1 (Cg-Ext(2021)) – WMO Unified Data Policy, in: Abridged Final Report of the Extraordinary Session of the World Meteorological Congress, WMO-No. 1281, World Meteorological Organization, Geneva, Switzerland, 2021.

Yigitbasioglu, O. M. and Velcu, O.: A review of dashboards in performance management: Implications for design and research, International Journal of Accounting Information Systems, 13, 41–59, https://doi.org/10.1016/j.accinf.2011.08.002, 2012.

Zhilcova E.L., Anisimov O.A. On the accuracy of air temperature and precipitation replicated by global climate archives, Russian Meteorology and Hydrology 10 (2009) , 79–90 (in Russian).

Zhu, X., Wu, T., Chen, J., Wu, X., Wang, P., Zou, D., Yue, G., Yan, X., Ma, X., Wang, D., Lou, P., Wen, A., Shang, C., and Liu, W.: Summer heat wave in 2022 led to rapid warming of permafrost in the central Qinghai-Tibet Plateau, NPJ Clim. Atmos. Sci., 7, 216, https://doi.org/10.1038/s41612-024-00765-4, 2024.

Zillman, J.W., 2019. Origin, impact and aftermath of WMO Resolution 40. WMO Bulletin, 68(2), 69–71.