Data rescue of historical wind observations in Sweden since the 1920s

John Erik Engström[1], Lennart Wern[1], Sverker Hellström[1], Erik Kjellström[1], Chunlüe Zhou[2], Deliang Chen[2], and Cesar Azorin-Molina[2,3]

[1]Swedish Meteorological and Hydrological Institute, Climate Information and Statistics, Norrköping, 601 76, Sweden
[2]Regional Climate Group, Department of Earth Sciences, University of Gothenburg, PO Box 6460, Gothenburg 40530, Sweden
[3]Centro de Investigaciones sobre Desertificación, Consejo Superior de Investigaciones Científicas (CIDE, CSIC-UV-Generalitat Valenciana), Climate, Atmosphere and Ocean Laboratory (Climatoc-Lab), Moncada, Valencia, Spain

Correspondence: Erik Engström (erik.engstrom@smhi.se)

Abstract. Instrumental measurements of wind speed and direction from the 1920s to the 1940s from 13 stations in Sweden have been rescued and digitized, making 165 additional station years of wind data available through the Swedish Meteorological and Hydrological Institute's open data portal. These stations measured wind through different versions of cup-type anemometers and were mainly situated at lighthouses along the coasts and at airports. The work followed the protocol "Guidelines on Best Practices for Climate Data Rescue" of the World Meteorological Organization consisting of (i) designing a template for digitization; (ii) digitizing records in paper journals by a scanner; (iii) typing numbers of wind speed and direction data into the template and (iv) performing quality control of the raw observation data. Along with the digitization of the wind observations, meta data from the stations were collected and compiled as support to the following quality control and homogenization of the wind data. The meta data mainly consist of changes in observer and a small number of changes in instrument types and positions. The rescue of these early wind observations can help improve our understanding of long-term wind changes and multidecadal variability (e.g., the "stilling" vs. "reversal" phenomena), but also to evaluate and assess climate simulations of the past. Digitized data can be accessed through the SMHI open data portal: https://www.smhi.se/data, last access: ~~26 December 2022~~7 April 2023, and Zenodo repository: https://doi.org/10.5281/zenodo.5850264, last access: ~~26 December 2022~~7 April 2023, (Zhou et al., 2022).

1 Introduction

The ongoing climate change, described by the Intergovernmental Panel on Climate Change (IPCC) (IPCC, 2021), raises the question of if, how, and why the wind climate in the world is changing. Winds have a large impact on many parts of the society, economy and environment (Rayner (2007); Vicente-Serrano et al. (2018); Vose et al. (2014); Pörtner et al. (2022)). Extreme winds can for example affect buildings, infrastructure, shipping (Pörtner et al., 2022), extreme cold outbreaks in winter (Zhou et al., 2021) and energy production (Pörtner et al. (2022); Castro et al. (2011)). To reduce greenhouse gas emissions there is an urgent need to transform the energy system to be more sustainable and robust, where renewable wind energy can be a part

of the solution (Pörtner et al., 2022). Of the global electricity generation, wind energy made up 6.6% in 2021 (www.iea.org, last access: ~~26 December 2022~~7 April 2023). The global annual net wind power generation was 1,870 TWh in 2021 and has to reach 7,900 TWh in 2030 to meet the Net Zero Emission by 2050 scenario (www.iea.org, last access: ~~26 December 2022~~7

25 April 2023). To secure future energy production for Sweden it is important to know how the wind climate is changing when extreme weather events are estimated to possibly cause up to 16% drop in power supply reliability for Swedish cities in future scenarios (Perera et al., 2020).

  To better understand how and why the wind climate is changing we need to improve our limited knowledge of historical wind changes and multidecadal variability. A fundamental challenge lies in rescuing old climate observations from weather

30 archives such as those at the National Weather Services and derive homogeneous and complete data series, as stated by the World Meteorological Organization (WMO) (WMO, 2016). The quality of the observations can be affected by measurement inconsistencies (Aguilar et al., 2003), changes in the surrounding environment (Dienst et al., 2017), changes of station position (Wan et al., 2010) and drifting anemometers (Azorin-Molina et al., 2018). At SMHI most of the meteorological data is available digitally since the 1960s, but before that, only a minor part of the data is in digital form.

35 Prior studies have shown a global reduction in wind speed (termed "stilling"; (Roderick et al., 2007)) over land from the 1960s to the 1980s (McVicar et al. (2012); IPCC (2021)), and from around 2010 a "reversal" with increasing wind speed (Dunn et al. (2016); Azorin-Molina et al. (2017); Azorin-Molina et al. (2018); Zeng et al. (2019). In Sweden, near-surface wind speed showed a statistically significant decreasing trend during 1956-2013 with a stronger reduction during 1979-2008 (Minola et al., 2016). But when each season is studied separately a small non-significant wind speed increase was detected for winter during

40 1979-2008, which was partly driven by the increasing North Atlantic Oscillation (NAO) index during those decades (Minola et al., 2016). In a more recent study, Minola et al. (2021) found a break in the stilling around 2010 when analyzing near-surface mean wind speed and gusts in Sweden for 1997–2019. The detected stilling-reversal was to a large extent attributed to changes in the NAO. They also found a greater increase in wind gusts compared to the mean wind speed. Additionally, changes in surface roughness were also identified as another driver of wind variability in Sweden.

45 The IPCC AR6WGI (IPCC, 2021) stated that the confidence in wind changes is "low" to "medium", highlighting that there is still much to learn about wind changes and multidecadal variability in a warming climate. One of the limitations to study the "stilling" and "reversal" periods is due ~~is~~ to the low availability of centennial wind series, and inhomogeneities of observed wind speed records (IPCC, 2021). Before the 2010s there were not many studies dealing with quality control and homogenization of wind observations (Xu et al. (2006); Rayner (2007); Wern and Bärring (2009)), but during the last decade some publications

50 have increased the knowledge in this research area (Wan et al. (2010); Azorin-Molina et al. (2014); Minola et al. (2016); Azorin-Molina et al. (2019)). Due to lack of data, earlier studies of wind conditions in Sweden could only evaluate trends and variability back to the 1960s, highlighting the need of extending the digitally available wind observations further back in time (Minola et al., 2016).

  Still, a vast amount of meteorological observational data is only present as paper journals in archives, and thus not easily

55 accessible for scientists to study climate change. WMO (WMO, 2016) stated that rescuing (digitize data and store it in a database that is backed up) old climate data from weather archives and creating homogeneous and complete data sets is a

fundamental challenge facing the climate research community. The rescue and digitization of observational data is identified as a crucial step in producing new and robust historical ~~reanalyses~~ reanalyzes (Brönnimann et al., 2018), and in better assessing and attributing wind changes. There are other ongoing data rescue initiatives focused on digitizing historical wind records from weather archives. The project WeatherRescue.org has used volunteer citizen scientists to digitize observations from the mid-1800s to the beginning of the 1900s at stations in Scotland and other locations in Europe (Hawkins et al., 2019).

SMHI holds valuable wind records only available in paper journals prior to the 1940s as the digitization of climate series at SMHI systematically started only in the 1950s–1960s. To rescue historical wind observations in Sweden from the 1920s to 1940s, the project "Assessing centennial wind speed variability from a historical weather data rescue project in Sweden" (WINDGUST) was initiated, aimed at filling the gap of data prior to 1939 and temporal inhomogeneity of wind datasets in Sweden, with the ultimate goal of bringing more wind data into the hands of scientists who seek to improve the limited knowledge on the causes of wind variability and change including that of "stilling" and "reversal" as discussed in global climate science.

Here we focus the results, challenges and lessons learned from the first part on data rescue of the WINDGUST project. Methods and results of the homogenisation are presented by and Zhou et al. (2022) and the evaluation and attribution of trends are presented by Minola et al. (Submitted).

2 A brief history of Swedish wind observations

A Swedish national station network for meteorological ground-based observations was initiated by the Swedish Royal Academy of Science in 1858 (Östman, 1928). Meteorologiska Centralanstalten (SMHI: https://www.smhi.se/, last access: ~~26 December 2022~~7 April 2023) was established in 1873 and was from then responsible for meteorological and hydrological observations across Sweden. Wind speed and direction have always been key weather observations. At the time of the initiation of the national station network in 1858, a wind speed scale from 0 to 4 was used. This scale was already introduced by Anders Celsius (1701-1744). In conjunction with the establishment of Meteorologiska Centralanstalten in 1873, a six-grade scale was taken into use. The 6 grades corresponded to every second grade in the twelve-grade Beaufort scale. The complete twelve-grade Beaufort scale was applied after the establishment of Statens Meteorologiska-Hydrologiska Anstalt (SMHA) in 1919 (SMHI, personal communication). In the early years, wind observations were usually estimated using the scales mentioned above, although there were some instrumental measurements done already during the 20th century. From 1908 to 1910 around 15 lighthouse stations were equipped with Fineman anemometers and eight coastal stations with Rung anemometers (Östman, 1928). The Fineman anemometer was a cup anemometer type, usually mounted on a 4-meter-high pole. To get the wind speed value, the observer had to climb up to the top of the pole. Rung anemometer was a tube with the wind blowing over a nozzle at the upper end of the tube. ~~This~~ The wind created a pressure drop in the tube ~~moving a pointer that~~ that moved a pointer which showed the wind speed on a scale. The Rung anemometers were mounted on 6-meter-high poles. In 1923, the Sandström cup-type anemometer (Fig. 1) was first introduced, and in 1928 this anemometer was in operation at around 30 coastal stations and 10 inland stations. This anemometer type had four cups and was installed in a mast 10–20 m above ground or in a lighthouse

tower that in some cases reach 30–40 m above the ground. During the second half of the 20th century wind measurements were performed with a Söderlund cup anemometer (Fig. 2) (SMHI, personal communication) and since 1996 with a Thies ultrasonic anemometer at automatic weather stations (AWS) (Wern and Bärring, 2009).

The early stations equipped with the above-described wind instruments (Table 1) were often lighthouses and airports (Fig. 3). The observations were done by the weather observers at these stations along with their primary duties as lighthouse keepers or airport crew. The observations were done at least three times a day and often four times a day: morning (8), midday (14), and evening (21) (Swedish normal time) before 1941. Between 1941 and 1946 the observations were done at 8, 14 and 19. From 1947 and onwards the hours were at 7, 13 and 19.

As an example, at the station Ölands norra udde the Sandström anemometer was installed in 1924 and then changed in 1938 to a SMHA No 18 anemometer. In Fig. 4 wind speed observations are plotted and the change of anemometer is marked by a vertical orange dashed line in the plot.

3 Rescue of historical wind data

The WINDGUST project rescued historical wind observations available in the old weather journals at the SMHI archives for the 1920s-1940s by applying the well-defined protocol "Guidelines on Best Practices for Climate Data Rescue" of WMO (WMO, 2016). The process consisted of (i) designing a template for digitization; (ii) digitizing paper journals by an imaging process based on scanning or photographs; (iii) typing numbers of wind speed data into the template and (iv) performing a quality control of the data digitized in step iii. Prior studies have concluded that manual typing is the quickest and most reliable data rescue method compared to other methods such as speech recognition and optical character recognition (Brönnimann et al. (2006); Ashcroft et al. (2018)).

First, an inventory of suitable and available observations in Sweden on wind speed and direction was performed in the archives of SMHI. Only stations with wind speed data in meter per second (m s^{-1}) were selected. These stations were considered to have been equipped with anemometers so the wind speed could be expected to have been measured. At stations where wind speed was only reported in the Beaufort scale, the wind speed was expected to have been estimated from signs in the surrounding environment due to the information from the journals and the wind series were discarded for the digitization. This resulted in a list of 13 stations across Sweden (Fig. 5 and Table 2). The stations were located in different parts of the country, mostly along the coast. Although it was not a criterion for selection, ten of the stations are still active making analyzes over a full century possible. The archive sheets did contain other variables that will be digitized according to the priority of stations and variables in coming years.

3.1 Weather data in paper journals

The paper journals for the selected stations were brought from the archive at SMHI. The journals used were books, each covering six months of data. Each book lookup represented observations for one day (Fig. 6). The observations were done three to four times a day (see Sect. 2)

3.2 Scanning

Each book lookup was scanned in a dedicated Bookeye 4 scanner (Fig. 7). Each image of a book lookup was compiled to a pdf file containing one year of observations. The scanner could be operated both by hand using buttons or a computer mouse, or by foot using a pedal. The scanning had to be done in the institute while the original journals were not allowed to leave the building. The paper journals were mostly in good shape and were handled with care. In a few cases, however, data were missing in the paper journals or the handwriting was difficult to read.

3.3 Digitization

A template for the digitization was designed and then used in the project (Fig. 8). The scanned image of the paper journal was viewed on one computer screen and the values from the paper journals were typed into the template on another computer screen. When the values were typed into the template, they were automatically stored in the observational MORA database at SMHI. The scanning and digitization were static work and the staff performing these duties were encouraged to often take short breaks not to get stuck in one working position for too long.

3.4 Quality control

Once the wind measurements were stored in the observational MORA database at SMHI, a first quality control check was performed to detect missing or suspicious data. If data gaps were found, the reason for missing data was investigated. High wind speeds were also checked. For instance, logical tests were performed where the maximum wind speed could not be lower than the minimum wind speed. All wind speed values higher than the known record for each station were also checked. Further details of these tests are discussed in Sect. 4.3.

3.5 Meta data

Meta data for all digitized stations was compiled in the project. The information was gathered from the protocols of the inspection of the stations done by the SMHI staff. Every station was inspected with intervals from every year to every fifth year and any changes at the stations were noted. Changes could be in the surroundings such as new buildings or old buildings that had been removed, trees or bushes that had grown up or been removed, new observers, change in instrumentation or routines. If routines decided by SMHI were not followed, it was corrected. A shortcoming of the protocols is that if a change is noted, the exact date of the change is seldom known. The change then occurred sometime between the present inspection and the prior inspection, which could mean a time span of one to five years. The lowest section in Fig. 9 is an example of meta data from the station Väderöbod from 24 September 1941. Translated to English the text is "Wind. Wind instrument S.M.H.A. No 28 and chronograph Fuess No 9753 were inspected and lubricated. Good functionality.". The meta data was compiled in digital files with codes attributed to different types of changes, which made it possible to read the meta data files automatically with computer scripts as a part of the homogenization method.

3.6 MORA database

The digitized wind observations were directly stored in the observational MORA database at SMHI. In MORA, meteorological observations are stored, both as a digital historical archive ranging from the oldest observations of temperature from Uppsala 1722 to the new real time observations. The database also contains some statistics such as mean values for month, year and 30-year periods. The database contains more than 19 billion values from around 4,300 Swedish stations, 16,000 international stations and over 100 meteorological parameters. Swedish observational data can be accessed by anyone manually through web-interfaces or by an application programming interface (API) also called machine to machine interfaces. Most of the meteorological observations in Sweden from 1960 onwards are digitally available in MORA and can be accessed through SMHI's open data portal: https://www.smhi.se/data, last access: ~~26 December 2022.~~ 7 April 2023. However, the observations before 1960 are only partly available in MORA as the digitization of this data is still ongoing.

4 Results

The work, completed by four SMHI staff members during 2,200 man hours, was finished in the spring of 2021. They had then scanned and digitized 13 selected stations spanning the years 1925 to 1948.

4.1 Data screening

A data screening was performed for the 13 stations that were digitized to visualize the data coverage, monthly variability, and wind speed range. The stations were classified into two groups, coastal and inland stations. Malmslätt (Fig. 10) represents a typical inland station with the highest monthly mean wind speed during October to March around 4 m s^{-1}. The maximum monthly mean was reported for March (8–9 m s^{-1}). For all months the lowest monthly means are about 2 m s^{-1}. Observations have been continuously registered ~~since~~ at Malmslätt since 1945. Before that storing of journals were more random which caused the data gap from July 1936 to January 1939.

Vinga (Fig. 11) represents a coastal station with the highest monthly mean wind speed generally during November and December ranging between 8 and 9 m s^{-1}. The lowest mean monthly wind speed, 5–6 m s^{-1}, is obtained in May. The maximum monthly mean wind speed is approximately 13 m s^{-1} in November and minimum around 3 m s^{-1} in August. The data gap in the 1940s ~~is~~ was due to observations that by mistake were not digitized in the project, but will be digitized after the project. To fill the data gap between 1997 and 2007, the nearby station Trubaduren is suggested to be used in a future project. Kalmar, Härnösand, and Haparanda show monthly wind speed patterns typical for inland stations, even though the stations are located close to the sea. In Kalmar the station was located on a small island, and in Härnösand and Haparanda the stations were located around 1 km from the coast. The inland character of the station can be explained by the fact that coastal stations on the Swedish east coast are somewhat shielded for the dominating westerly to the south-westerly winds and possibly some additional obstacles in the station vicinity, such as buildings or trees. For all three stations and most particularly for Haparanda,

located in the far north of Bay of Bothnia, sea ice cover can give those stations more of an inland character in some winters. Screening plots for all digitized stations can be found in Appendix A.

4.2 Meta data compilation

185 Meta data for all digitized stations were compiled into digital files with the file format shown in Fig. 12. The format of the meta data files enabled the files to be read by a computer program, which some of the homogenization methods have functionality for. First the data files contains station identification number, name of the station, as well as start and stop year for the series. After that follows a free text section summarizing identified potential homogeneity breaks. Then a section explaining the meaning of the break type codes. Last in the files, there is a list of identified break dates and types. For the 13 digitized stations, 122
190 potential homogeneity breaks were identified in the meta data, with a mean value of 17 per station. The number of potential breaks per station ranged from 6 to 30.

4.2.1 ~~Scanning and digitization~~

~~The corona pandemic has affected the whole world in many different ways. The staff at SMHI had to work from home as much as possible, which posed a challenge as the paper journals had to be handled physically at SMHI. The solution we chose was to~~
195 ~~scan the paper journals at the institute and then do the digitization at home using the electronic image files. The scanning and digitization were static work and the staff performing these duties were encouraged to often take short breaks not to get stuck in one working position for too long. A few times the scanning had to be redone several times to get acceptable quality. It was very good to have a pedal control for the scanner to free the hands to shift pages in the paper journals during the scanning. The paper journals were mostly in good shape and were handled with care. In a few cases, however, data were missing in the paper~~
200 ~~journals or the handwriting was difficult to read.~~

4.3 Quality control

The digitized values were checked for errors as soon as possible in parallel to the continuous digitization to be able to correct unclear instructions and improve the process. Suspicious values were flagged and checked an extra time in the paper journals. Several quality control checks were performed as described in Sect. 3.4. The following is a list of examples of problems found
205 and how they were dealt with:

- Sometimes the wind observation was digitized at an erroneous time of day in the database. This was discovered when the digitized data were controlled and the observations were not stored with the correct observational time in the database. This could be corrected by running a programming script.

- For some stations, identical values had been digitized at two consecutive timestamps. Again, the paper journals were
210 consulted and the value at the incorrect time was corrected.

- On some occasions, the wrong column from the paper journals had been digitized such as visibility data instead of wind speed. The missing wind data were digitized to replace the incorrect values.

- A large number of gaps in the digitized data were identified. It was found to be times marked as calm weather (no wind) that the staff performing the digitization had left empty in the digitized files due to unclear instructions. These gaps were filled in with zeros for wind speed and wind direction.

- On some occasions there was no wind speed noted in meter per second (m s^{-1}) in the paper journals, but only wind speed in Beaufort. If the gap in m s^{-1} was only one single observation time, the value in m s^{-1} was interpolated from the Beaufort value. If the gap was several observational times, values in m s^{-1} were not interpolated but marked as missing data.

- Some dates (pages) were found missing in the electronic image files from the scanning. The wind speed values for the missing dates (pages) were digitized after the scanning process.

- A number of the highest wind speed values for each station were checked against the paper journals an extra time and some values were found to be wrong and were corrected.

The performed quality control is, however, not complete. When starting to analyze data in more detail, additional suspicious values will probably appear. Then there is a possibility to check the original value in the scanned paper journal. If the value in the database turns out to be wrong then the database can be updated with the correct value and the quality of the analyzis will improve. During the time period of digitization 1920 to 1940, totally 242364 values were digitized. Out of these 242364 values, 392 values were corrected during the QC. For the ten stations with century-long records the ratio of missing monthly values from 1920 to now was 9%, 975 out of 11222 values.

5 Homogenization

The homogenization was performed and described by Zhou et al. (2022). The penalized maximal F test (PMF) method was used and a combination of CERA-20C and ERA5 was used as a reference data set for the homogenization. 71 change points for the 13 stations were detected. Most change points were found in the years 1935-1944, 1956-1964 and around 1985. Around 38% of the detected change points were confirmed by meta data from the stations. To adjust the time series for the detected change point, a mean-matching adjustment was applied using up to 5 years of data before and after each change point. The most recent segment was chosen as reference. The 13 homogenized wind speed time series showed an increase of 0.09 m s^{-1} per decade that peaked around 1990. From 1990 to 2005 a decrease by 25% in the trend followed. After 2005 the homogenized time series showed a rapid recovery in wind speed.

6 Attribution of trends

240 Further analyzis of the wind speed trends over Sweden showed a stilling during 1926-1960, a stabilization from 1960 to 1990 and then a second stilling during 1990-2003 (Minola et al., Submitted). A slight recovery was found 2003-2014. Three reanalyzis data sets were used to attribute the trends found over Sweden to large-scale atmospheric circulation patterns (e.g. the North Atlantic Oscillation). The three reanalyzis data sets were ERA20C, ERA5 and NOAA-CIRES. The analyze showed that changes in large-scale atmospheric circulations mainly drive the observed low-frequency variability of near-surface wind
245 speed, but could not fully represent the periods of strong stilling over Sweden, for which other factors are needed for attribution (e.g. changes in surface roughness).

7 Discussion

~~This work~~ The scanning of the original journals enables an additional electronic backup of the data and improve the accessibility when the electronic files can be sent to external users while the original journals have to be kept in the archive. The electronic
250 copies also made it possible to perform the digitization at any place suitable for the staff doing the digitization. It would be beneficial if the work can continue with digitization of more stations further back in time. This work provides a significant addition to the available wind observations for Sweden, but thirteen stations is far from enough to be representative for the whole country of Sweden. One possibility could be to digitize wind speed observations recorded with the different observed wind scales previously used, which would demand a profound quality evaluation to determine their usability.

255 This work contributes to satisfy the need ~~as stated by (WMO, 2016)~~ to increase the amount of digitized wind observations available for the international research community~~and enable~~, as stated by WMO (2016). The digitized observations enables additional insight into the historical variability and trends of the wind climate in Sweden~~and globally as highlighted by (IPCC, 2021)~~. There is still a vast amount of meteorological data that are only stored in paper journals that need to be digitized to be rescued for future generations and made practically accessible and usable for fellow researchers. This urges the national
260 meteorological services to continue their efforts to rescue and digitize meteorological data. In this project we acknowledge the benefit of having experienced digitization staff that work in close collaboration with climate experts to identify and correct errors in the digitization. The International Data Rescue (I-DARE; idare-portal.org; last accessed ~~26 December 2022~~7 April 2023) portal collects and provides information on a large number of data rescue activities and facilitates sharing of experience and methods in this field. Some projects have successfully tried to engage the public in citizen science contributions where
265 people could volunteer to digitize data from scanned historical records (Hawkins et al. (2019); Craig and Hawkins (2020)) using the web-service of weatherrescue.org; last accessed ~~26 December 2022~~7 April 2023, where the quality control was conducted by triple parallel digitization of three independent persons. To further accelerate the pace in rescuing and digitization of historical data the techniques of artificial intelligence (AI), machine learning (ML), and optical character recognition (OCR) could be developed and applied in this field as discussed by Campbell et al. (2021). These techniques have so far been of
270 limited value due to the low ratio of accuracy but the methods are improving fast. In the near future, there will probably be a growing number of successful projects rescuing data by these and similar methods.

8 ~~Outlook~~

At SMHI the work continues to rescue and digitize historical meteorological, hydrological and oceanographic observations from the archives of paper journals spanning over more than 150 years ~~continues~~ and the methods discussed above are developed to facilitate and speed up this work further.

8 Conclusions

With the newly digitized observations, studies of wind conditions in Sweden could be extended to the time period from around 1920 to present, enveloping one full century of wind measurements. The results of this work were fruitfully brought forward by Zhou et al. (2022) who homogenized and investigated the historical variability and trends in surface wind speed in Sweden for the last century and Minola et al. (Submitted) who investigated the causes of the detected trends in the Swedish wind climate. Digitized data can be accessed through the SMHI open data portal: https://www.smhi.se/data, last access: ~~26 December 2022~~7 April 2023, and Zenodo repository: https://doi.org/10.5281/zenodo.5850264, last access: ~~26 December 2022~~7 April 2023, (Zhou et al., 2022).

9 Data availability

Digitized data can be accessed through the SMHI open data portal: https://www.smhi.se/data, last access: 7 April 2023, and Zenodo repository: https://doi.org/10.5281/zenodo.5850264, last access: 7 April 2023, (Zhou et al., 2022).

Author contributions. Erik Engström wrote the major part of the manuscript. Lennart Wern and Sverker Hellström supported the scanning and digitization staff, performed quality control and provided valuable knowledge of historical wind measurements and stations in Sweden. Cesar Azorin-Molina, Chunlue Zhou, Erik Kjellström and Deliang Chen contributed with comments improving the manuscript.

Competing interests. The authors declare no competing interests.

Acknowledgements. Arman Dezfolian and Lena Wikvist did the digitization of the wind data and Marie Sundström and Åsa Lindgren the scanning of the paper journals. Lisette Edvinsson produced the station map and Johan Södling the data screening plots in the article. FORMAS funded this project with id 2019-00509. Cesar Azorin-Molina and Deliang Chen were also supported by the IBER-STILLING (RTI2018-095749-A-I00, MCIU/AEI/FEDER,UE) and VENTS (GVA-AICO/2021/023) projects.

References

Aguilar, E., Auer, I., Brunet, M., Peterson, T., and Wieringa, J.: Guidelines on climate metadata and homogenization. World Meteorol, Organ, 1186, 1–52, 2003.

Ashcroft, L., Coll, J. R., Gilabert, A., Domonkos, P., Brunet, M., Aguilar, E., Castella, M., Sigro, J., Harris, I., Unden, P., et al.: A rescued dataset of sub-daily meteorological observations for Europe and the southern Mediterranean region, 1877–2012, Earth System Science Data, 10, 1613–1635, https://doi.org/https://doi.org/10.5194/essd-10-1613-2018, https://doi.org/10.5194/essd-10-1613-2018, 2018.

Azorin-Molina, C., Vicente-Serrano, S. M., McVicar, T. R., Jerez, S., Sanchez-Lorenzo, A., López-Moreno, J.-I., Revuelto, J., Trigo, R. M., Lopez-Bustins, J. A., and Espírito-Santo, F.: Homogenization and Assessment of Observed Near-Surface Wind Speed Trends over Spain and Portugal, 1961–2011, Journal of Climate, 27, 3692 – 3712, https://doi.org/10.1175/JCLI-D-13-00652.1, https://journals.ametsoc.org/view/journals/clim/27/10/jcli-d-13-00652.1.xml, 2014.

Azorin-Molina, C., Dunn, R., Mears, C., Berrisford, P., McVicar, T., and Nicolas, J.: Surface winds [in "State of the Climate in 2016"], Bull. Amer. Meteor. Soc, 98, S37–S39, 2017.

Azorin-Molina, C., Asin, J., McVicar, T. R., Minola, L., Lopez-Moreno, J. I., Vicente-Serrano, S. M., and Chen, D.: Evaluating anemometer drift: A statistical approach to correct biases in wind speed measurement, Atmospheric Research, 203, 175–188, https://doi.org/https://doi.org/10.1016/j.atmosres.2017.12.010, https://www.sciencedirect.com/science/article/pii/S0169809517312401, 2018.

Azorin-Molina, C., Guijarro, J. A., McVicar, T. R., Trewin, B. C., Frost, A. J., and Chen, D.: An approach to homogenize daily peak wind gusts: An application to the Australian series, INTERNATIONAL JOURNAL OF CLIMATOLOGY, 39, 2260–2277, https://doi.org/10.1002/joc.5949, 2019.

Brönnimann, S., Annis, J., Dann, W., Ewen, T., Grant, A., Griesser, T., Krähenmann, S., Mohr, C., Scherer, M., and Vogler, C.: A guide for digitising manuscript climate data, Climate of the Past, 2, 137–144, https://doi.org/https://doi.org/10.5194/cp-2-137-2006, https://doi.org/10.5194/cp-2-137-2006, 2006.

Brönnimann, S., Allan, R., Atkinson, C., Buizza, R., Bulygina, O., Dahlgren, P., Dee, D., Dunn, R., Gomes, P., John, V. O., Jourdain, S., Haimberger, L., Hersbach, H., Kennedy, J., Poli, P., Pulliainen, J., Rayner, N., Saunders, R., Schulz, J., Sterin, A., Stickler, A., Titchner, H., Valente, M. A., Ventura, C., and Wilkinson, C.: Observations for Reanalyses, Bulletin of the American Meteorological Society, 99, 1851 – 1866, https://doi.org/10.1175/BAMS-D-17-0229.1, https://journals.ametsoc.org/view/journals/bams/99/9/bams-d-17-0229.1.xml, 2018.

Campbell, S. I., Allan, D. B., Barbour, A. M., Olds, D., Rakitin, M. S., Smith, R., and Wilkins, S. B.: Outlook for artificial intelligence and machine learning at the NSLS-II, Machine Learning: Science and Technology, 2, 013 001, 2021.

Castro, C. d., Mediavilla, M., Miguel, L. J., and Frechoso, F.: Global wind power potential: Physical and technological limits, Energy Policy, 39, 6677–6682, http://inis.iaea.org/search/search.aspx?orig_q=RN:43064173, 2011.

Craig, P. M. and Hawkins, E.: Digitizing observations from the Met Office Daily Weather Reports for 1900–1910 using citizen scientist volunteers, Geoscience Data Journal, 7, 116–134, https://doi.org/https://doi.org/10.1002/gdj3.93, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/gdj3.93, 2020.

Dienst, M., Lindén, J., Engström, E., and Esper, J.: Removing the relocation bias from the 155-year Haparanda temperature record in Northern Europe, International Journal of Climatology, 37, 4015–4026, https://doi.org/https://doi.org/10.1002/joc.4981, https://doi.org/10.1002/joc.4981, 2017.

Dunn, R., Azorin-Molina, C., Mears, C., Berrisford, P., and McVicar, T.: Global climate; atmospheric circulation] surface winds [in "state of the climate in 2015"], Bull. Amer. Meteor. Soc, 97, S38–S40, 2016.

Hawkins, E., Burt, S., Brohan, P., Lockwood, M., Richardson, H., Roy, M., and Thomas, S.: Hourly weather observations from the Scottish Highlands (1883–1904) rescued by volunteer citizen scientists, Geoscience Data Journal, 6, 160–173, https://doi.org/https://doi.org/10.1002/gdj3.79, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/gdj3.79, 2019.

IPCC, A. W.: Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, https://doi.org/10.1017/9781009157896., 2021.

McVicar, T., Roderick, D., Li, J., Van Niel, T., Thomas, P., Grieser, J., Jhajharia, D., Himri, Y., Mahowald, N., Mescherskaya, Kruger, A., and Dinpashoh, Y.: Global review and synthesis of trends in observed terrestrial near-surface wind speed: Implications for evaporation, Journal of Hydrology, 416–417, https://doi.org/10.1016/j.jhydrol.2011.10.024, 2012.

Minola, L., Azorin-Molina, C., and Chen, D.: Homogenization and Assessment of Observed Near-Surface Wind Speed Trends across Sweden, 1956–2013, Journal of Climate, 29, 7397 – 7415, https://doi.org/10.1175/JCLI-D-15-0636.1, https://journals.ametsoc.org/view/journals/clim/29/20/jcli-d-15-0636.1.xml, 2016.

Minola, L., Reese, H., Lai, H.-W., Azorin-Molina, C., Guijarro, J. A., Son, S.-W., and Chen, D.: Wind stilling-reversal across Sweden: The impact of land-use and large-scale atmospheric circulation changes, International Journal of Climatology, 42, 1049–1071, https://doi.org/https://doi.org/10.1002/joc.7289, https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.7289, 2021.

Minola, L., Lönn, J., Azorin-Molina, C., Zhou, C., Engström, E., Wern, L., Hellström, S., Zhang, G., Shenan, C., Pezzoli, A., and Chen, D.: The contribution of large-scale atmospheric circulation to variations of observed near-surface wind speed across Sweden since 1926, Climatic Change, Submitted.

Östman, C. J.: Om vindskalor och vindmätare i Svensk meteorologi, Meddelanden från Statens Meteorologisk-Hydrografiska Anstalt, 4, 1928.

Perera, A. T. D., Nik, V. M., Chen, D., Scartezzini, J.-L., and Hong, T.: Quantifying the impacts of climate change and extreme climate events on energy systems, Nature Energy, 5, 150–159, https://doi.org/10.1038/s41560-020-0558-0, https://doi.org/10.1038/s41560-020-0558-0, 2020.

Pörtner, H. O., Roberts, D. C., Adams, H., Adler, C., Aldunce, P., Ali, E., Begum, R. A., Betts, R., Kerr, R. B., Biesbroek, R., et al.: Climate change 2022: impacts, adaptation and vulnerability, IPCC AR6, 2022.

Rayner, D. P.: Wind Run Changes: The Dominant Factor Affecting Pan Evaporation Trends in Australia, Journal of Climate, 20, 3379 – 3394, https://doi.org/10.1175/JCLI4181.1, https://journals.ametsoc.org/view/journals/clim/20/14/jcli4181.1.xml, 2007.

Roderick, M., Rotstayn, L., Farquhar, G., and Hobbins, M.: On the attribution of changing pan evaporation, Geophysical Research Letters - GEOPHYS RES LETT, 34, https://doi.org/10.1029/2007GL031166, 2007.

Vicente-Serrano, S. M., Miralles, D. G., Domínguez-Castro, F., Azorin-Molina, C., Kenawy, A. E., McVicar, T. R., Tomás-Burguera, M., Beguería, S., Maneta, M., and Peña-Gallardo, M.: Global Assessment of the Standardized Evapotranspiration Deficit Index (SEDI) for Drought Analysis and Monitoring, Journal of Climate, 31, 5371 – 5393, https://doi.org/10.1175/JCLI-D-17-0775.1, https://journals.ametsoc.org/view/journals/clim/31/14/jcli-d-17-0775.1.xml, 2018.

Vose, R. S., Applequist, S., Bourassa, M. A., Pryor, S. C., Barthelmie, R. J., Blanton, B., Bromirski, P. D., Brooks, H. E., DeGaetano, A. T., Dole, R. M., et al.: Monitoring and understanding changes in extremes: Extratropical storms, winds, and waves, Bulletin of the American Meteorological Society, 95, 377–386, 2014.

Wan, H., Wang, X. L., and Swail, V. R.: Homogenization and Trend Analysis of Canadian Near-Surface Wind Speeds, Journal of Climate, 23,
370   1209 – 1225, https://doi.org/10.1175/2009JCLI3200.1, https://journals.ametsoc.org/view/journals/clim/23/5/2009jcli3200.1.xml, 2010.

Wern, L. and Bärring, L.: Sveriges vindklimat 1901-2008: Analys av trend i geostrofisk vind, SMHI, urn:nbn:se:smhi:diva-2791, 2009.

WMO: Guidelines on Best Practices for Climate Data Rescue, Tech. Rep. 1182, World Meteorological Organization, Geneva, Switzerland,
    https://doi.org/https://doi.org/10.25607/OBP-1513, https://doi.org/10.25607/OBP-1513, 2016.

Xu, M., Chang, C.-P., Fu, C., Qi, Y., Robock, A., Robinson, D., and Zhang, H.-m.: Steady decline of east Asian monsoon winds,
375   1969–2000: Evidence from direct ground measurements of wind speed, Journal of Geophysical Research: Atmospheres, 111,
    https://doi.org/https://doi.org/10.1029/2006JD007337, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2006JD007337, 2006.

Zeng, Z., Ziegler, A. D., Searchinger, T., Yang, L., Chen, A., Ju, K., Piao, S., Li, L. Z. X., Ciais, P., Chen, D., Liu, J., Azorin-Molina, C.,
    Chappell, A., Medvigy, D., and Wood, E. F.: A reversal in global terrestrial stilling and its implications for wind energy production, Nature
    Climate Change, 9, 979–985, https://doi.org/10.1038/s41558-019-0622-6, https://doi.org/10.1038/s41558-019-0622-6, 2019.

380 Zhou, C., Dai, A., Wang, J., and Chen, D.: Quantifying Human-Induced Dynamic and Thermodynamic Contributions to Severe Cold
    Outbreaks Like November 2019 in the Eastern United States, Bulletin of the American Meteorological Society, 102, S17 – S23,
    https://doi.org/10.1175/BAMS-D-20-0171.1, https://journals.ametsoc.org/view/journals/bams/102/1/BAMS-D-20-0171.1.xml, 2021.

Zhou, C., Azorin-Molina, C., Engström, E., Minola, L., Wern, L., Hellström, S., Lönn, J., and Chen, D.: HomogWS-se: A century-long
    homogenized dataset of near-surface wind speed observations since 1925 rescued in Sweden, Earth System Science Data Discussions,
385   2022, 1–24, https://doi.org/10.5194/essd-2022-29, Dataset:http://doi.org/10.5281/zenodo.5850264, 2022.
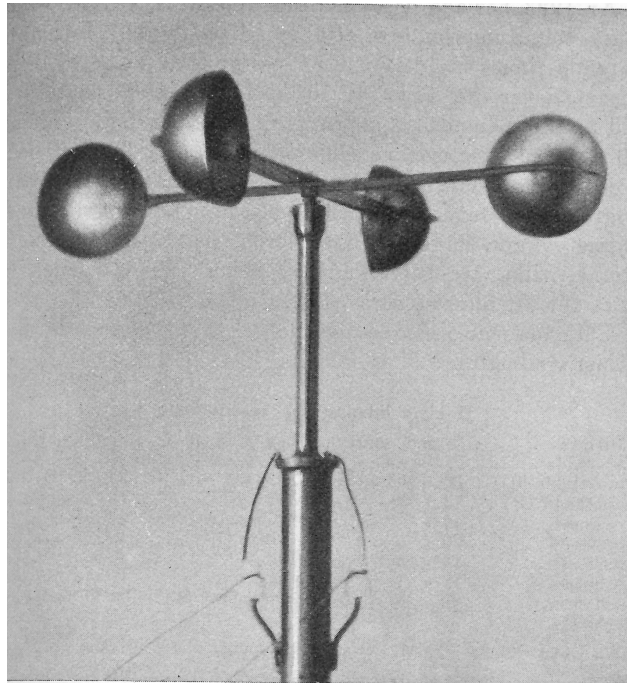
Figure 1. Sandström cup anemometer (Östman, 1928).
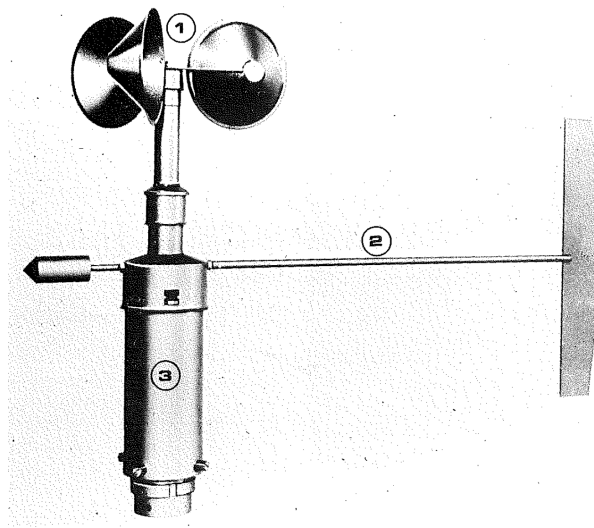
Figure 2. Söderlund cup anemometer used at SMHI observational station during the second half of the 20th century with cups (1), windvane (2) and transmission unit (3)(SMHI).

Figure 3. Sandström amemometer installations at four different Swedish stations in 1928. From left to right: Agö (mast), Grimskär (light house), Ölands södra udde (light house) and Lilljungfrun (storm warning mast) (Östman, 1928).
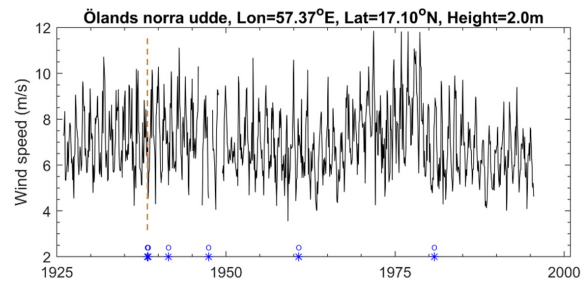
Figure 4. Wind speed observations at the station at Ölands norra udde from 1926 to 1996. Change of observer is marked in the lower part of the figure by a blue circle and star. A change of anemometer in 1938 is marked with a vertical orange dashed line (SMHI).
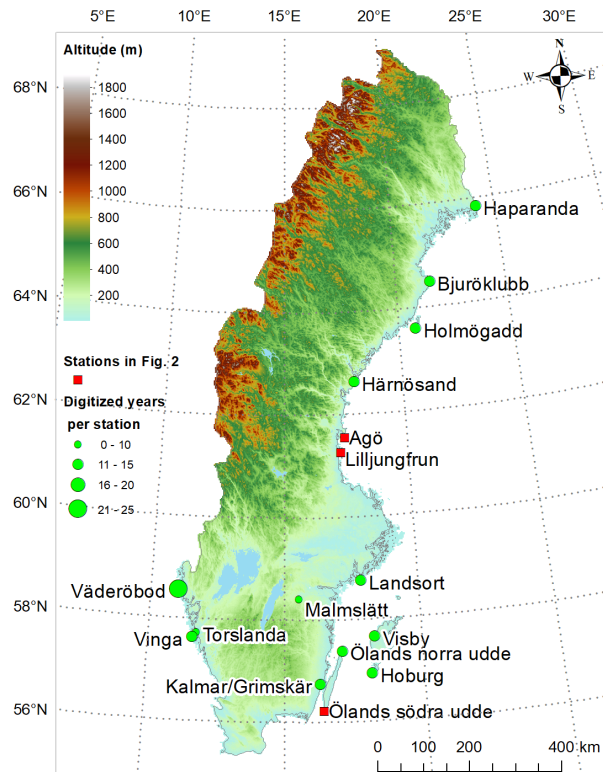
Figure 5. Elevation map of Sweden with the position of the wind stations that were digitized in green dots. Red squares show positions of the stations in the photographs in Fig. 2 (note that observations from these have not been digitized here), except Kalmar/Grimskär that was digitized (SMHI).

Figure 6. Paper journal for the Visby station, 1 January 1925. The columns are from left to right time of day (1), indoor temperature for correcting the mercury barometer reading (°C) (2), air pressure (mm hg, -700 units) (3), dry temperature (°C) (4), wet temperature (°C) (5), lower cloud type (6), lower cloud amount/cover (tenths) (7), higher cloud type (8), total cloud amount/cover (tenths) (9), wind direction (10), wind speed (Beaufort) (11), wind speed (m s^{-1}) (12), visibility (code number) (13), sea waves (14), maximum temperature (°C) (15), minimum temperature (°C) (16), precipitation amount (mm) (17), precipitation type (18). In the lower part of the page are boxes for snow cover (B (bare ground), S (snow cover), SB (mostly snow cover) and BS (mostly bare ground)) (19) and depth (cm) (20) (to the left) and present weather codes (21) (middle and right). At the bottom of the page are lines for extra notes concerning the type and time for precipitation, mist and other observations such as thunder, aurora borealis and frost (22) (SMHI).

Figure 7. Bookeye 4 scanning equipment used at SMHI (SMHI).
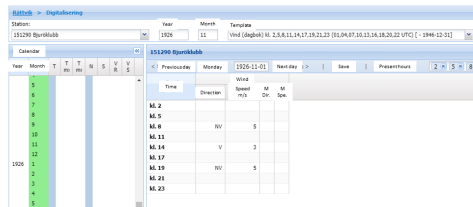
Figure 8. Digitization template for the data rescue of wind observations at SMHI (SMHI).
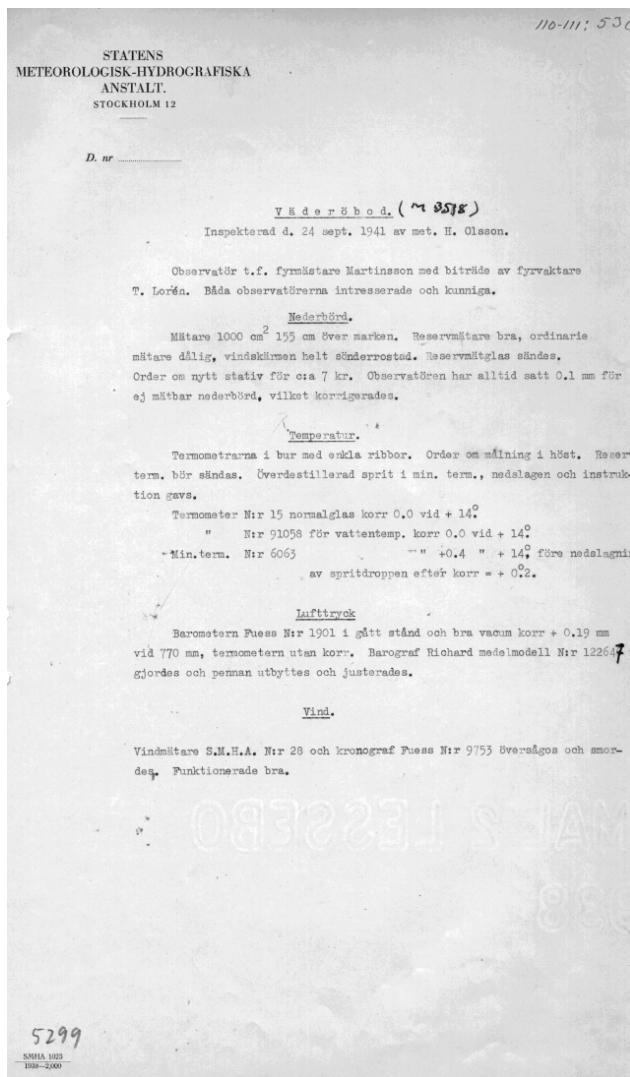
Figure 9. Meta data of the station Väderöbod from 24 September 1941. The lowest section with the title "Vind" concerns wind measurements (SMHI).
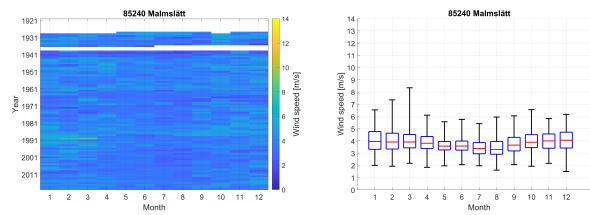
Figure 10. Data screening plot for Malmslätt as a typical inland station. Left panel shows the data coverage and the monthly mean wind speed value according to the color bar, while the right panel shows the monthly wind speed, where the red line is the monthly mean wind speed, the upper and lower limits of the box 75- and 25-percentile and the end-bars represents the maximum and minimum monthly mean wind speed (SMHI).
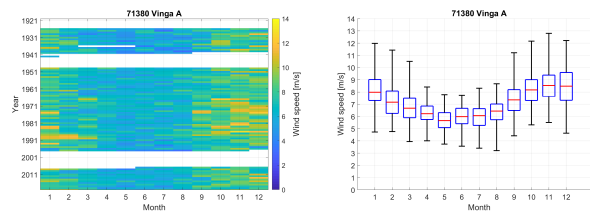
Figure 11. Data screening plot for Vinga as a typical coastal station. Left panel shows the data coverage and the monthly mean wind speed value according to the colour bar, while the right panel shows the monthly wind speed, where the red line is the monthly mean wind speed, the upper and lower limits of the box 75- and 25-percentile and the end-bars represents the maximum and minimum monthly mean wind speed (SMHI).

```
%81620 Väderöbod
%Längd på mätserie:
%1880   01      1966    01              (Mora-apps-Siljan, 2021-07-19)
%Koppling:      Nej
%
%Datum för identifierade potentiella homogenitetsbrott.
%
%1939-07-18 Nederbördsinsamling 145 cm över marken. Termometerns höjd över marken 180 cm.
Barometerns höjd över havet 9,5 meter. Vindmätarens (skålkors) höjd över marken 6,5 meter.
Observatör fyrmästare xxx, ställföreträdande fyrvaktare och biträde.
%1941-09-24 Observatör fyrmästare xxx och fyrvaktare xxx. Nederbördsmätare 1000 cm2 155 cm
över marken. Observatören har alltid angivit 0,1 mm för ej mätbar nederbörd. Termometer i
bur med enkla ribbor. Buren behöver målas om. Termometer nr 15 normalglas, termometer nr
91058 för vattentemp, min-termometer nr 6063. Lufttryck Fuess nr 1901 i bra skick.
Barograf Richard medelmodell nr 122647 justerades och ny penna monterades. Vindmätare SMHA
nr 28 och kronograf Fuess nr 9753 fungerade bra och smordes.
%
%Typ av brott:
%   Flytt temp&nbd -> TY= 10, Flytt temp -> TY= 11,  Flytt nbd -> TY= 12, Flytt vind ->
TY= 13, Flytt lufttryck -> TY= 14,
%   Observatörsbyte -> TY= 20,
%   Instrumentbyte temp&nbd -> TY= 30, Instrumentbyte temp -> TY= 31, Instrumentbyte nbd
-> TY= 32, Instrumentbyte vind -> TY= 33, Instrumentbyte lufttryck -> TY= 34
%   Förändring i omgivning temp&nbd -> TY= 40, Förändring i omgivning temp -> TY= 41,
Förändring i omgivning nbd -> TY= 42, Förändring i omgivningen vind -> TY 43, Förändring i
omgivningen lufttryck-> TY 44
%   Annat -> TY= 50.
%
%Första dag med nya förhållanden anges för brott.
%YYYY MO DA TY
1941 09 24 20
1941 09 24 40
```

Figure 12. Example of the file format for a compiled station meta data file. The example shows extracted information for Väderöbod. The names of the observers are replaced with "xxx" due to GDPR rules (SMHI).

Table 1. Technical specification of anemometer types. The higher mounting height interval marked with an * was for the anemometers mounted at light houses.

Anemometer	Type	Number of cups	Cup diameter (cm)	Arm length (cm)	Mounting height (m)	Introduction year
Rung	Pipe	-	-	-	6	1908
Fineman	Cup	4	10.4	10.4	4	1909
Sandström	Cup	4	17.5	34.7	10-20 (30-40*)	1923
Söderlund	Cup	3	14	10	10	1950

Table 2. The identification number (ID), position of the stations, start and stop years of the digitized data and for the stations.

Station	ID	Start data rescue	Stop data rescue	Latitude	Longitude	Station altitude (m.a.s.l.)	Wind observation start	Wind observation stop	Number of meta data
Bjuröklubb	151290	1926	1938	64.48	21.58	36	1879	running	28
Haparanda	163950	1925	1938	65.83	24.14	9	1859	running	19
Hoburg	68550	1926	1938	56.92	18.15	39	1879	running	8
Holmögadd	140360	1926	1938	63.59	20.76	6	1849	2008	22
Härnösand	127380	1925	1938	62.63	17.95	9	1787	running	12
Kalmar	66430	1927	1942	56.73	16.29	7	1858	running	6
Landsort	87450	1926	1938	58.74	17.87	13	1848	running	11
Malmslätt	85240	1928	1936	58.40	15.53	94	1858	running	17
Torslanda	71430	1932	1938	57.72	11.78	3	1931	1977	14
Vinga	71380	1926	1938	57.63	11.61	19	1849	running	29
Visby	78390	1925	1938	57.64	18.29	11	1859	running	7
Väderöbod	81620	1926	1948	58.55	11.03	7	1880	1966	19
Ölands norra udde	77220	1926	1938	57.37	17.10	5	1854	running	30
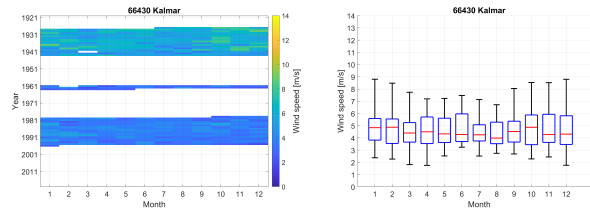
Figure A1. Data screening plot for Kalmar. Left panel shows the data coverage and the monthly mean wind speed value according to the color bar, while the right panel shows the monthly wind speed, where the red line is the monthly mean wind speed, the upper and lower limits of the box 75- and 25-percentile and the end-bars represents the maximum and minimum monthly mean wind speed.

Appendix A: Data Screening figures
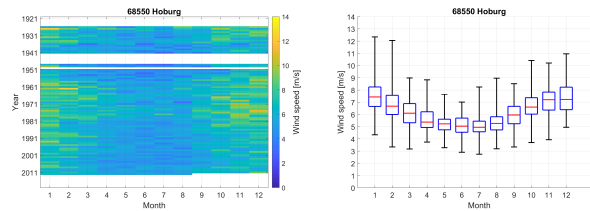
Figure A2. Data screening plot for Hoburg. Left panel shows the data coverage and the monthly mean wind speed value according to the color bar, while the right panel shows the monthly wind speed, where the red line is the monthly mean wind speed, the upper and lower limits of the box 75- and 25-percentile and the end-bars represents the maximum and minimum monthly mean wind speed.
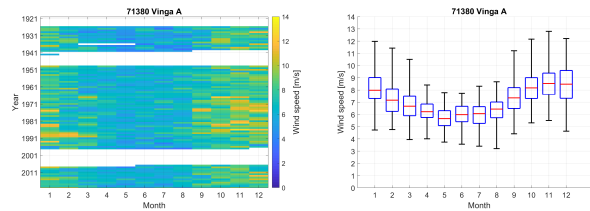
Figure A3. Data screening plot for Vinga. Left panel shows the data coverage and the monthly mean wind speed value according to the color bar, while the right panel shows the monthly wind speed, where the red line is the monthly mean wind speed, the upper and lower limits of the box 75- and 25-percentile and the end-bars represents the maximum and minimum monthly mean wind speed.
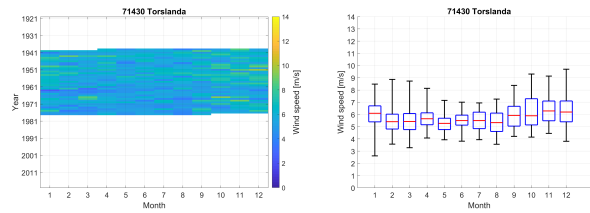
Figure A4. Data screening plot for Torslanda. Left panel shows the data coverage and the monthly mean wind speed value according to the color bar, while the right panel shows the monthly wind speed, where the red line is the monthly mean wind speed, the upper and lower limits of the box 75- and 25-percentile and the end-bars represents the maximum and minimum monthly mean wind speed.
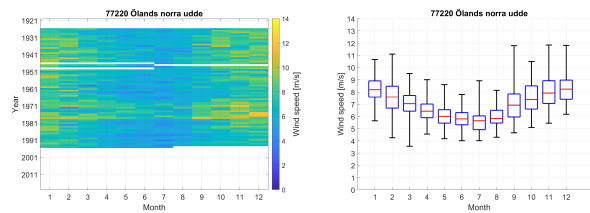
Figure A5. Data screening plot for Ölands norra udde. Left panel shows the data coverage and the monthly mean wind speed value according to the color bar, while the right panel shows the monthly wind speed, where the red line is the monthly mean wind speed, the upper and lower limits of the box 75- and 25-percentile and the end-bars represents the maximum and minimum monthly mean wind speed.
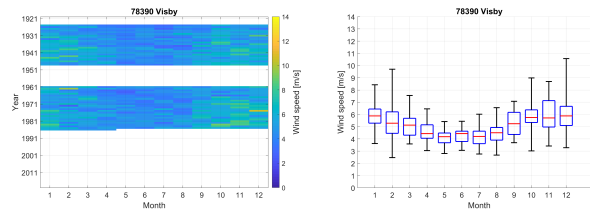
Figure A6. Data screening plot for Visby. Left panel shows the data coverage and the monthly mean wind speed value according to the color bar, while the right panel shows the monthly wind speed, where the red line is the monthly mean wind speed, the upper and lower limits of the box 75- and 25-percentile and the end-bars represents the maximum and minimum monthly mean wind speed.
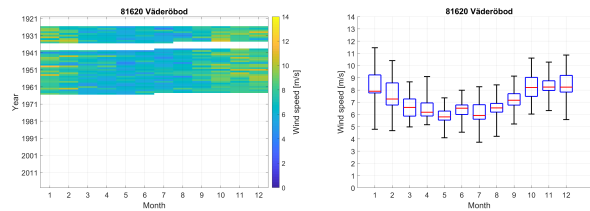
Figure A7. Data screening plot for Väderöbod. Left panel shows the data coverage and the monthly mean wind speed value according to the color bar, while the right panel shows the monthly wind speed, where the red line is the monthly mean wind speed, the upper and lower limits of the box 75- and 25-percentile and the end-bars represents the maximum and minimum monthly mean wind speed.
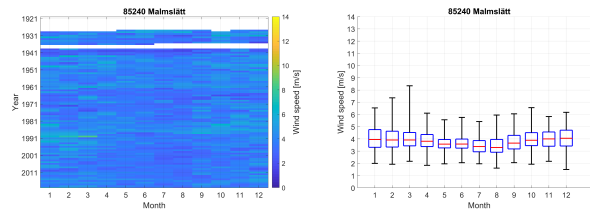
Figure A8. Data screening plot for Malmslätt. Left panel shows the data coverage and the monthly mean wind speed value according to the color bar, while the right panel shows the monthly wind speed, where the red line is the monthly mean wind speed, the upper and lower limits of the box 75- and 25-percentile and the end-bars represents the maximum and minimum monthly mean wind speed.
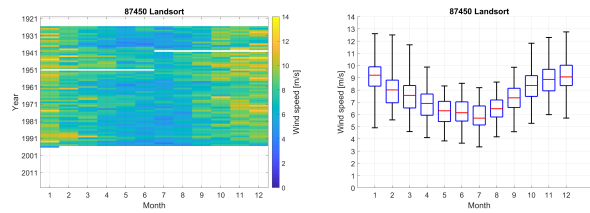
Figure A9. Data screening plot for Landsort. Left panel shows the data coverage and the monthly mean wind speed value according to the color bar, while the right panel shows the monthly wind speed, where the red line is the monthly mean wind speed, the upper and lower limits of the box 75- and 25-percentile and the end-bars represents the maximum and minimum monthly mean wind speed.
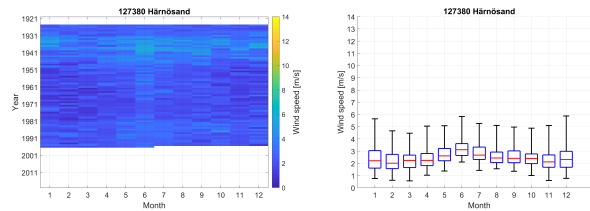
Figure A10. Data screening plot for Härnösand. Left panel shows the data coverage and the monthly mean wind speed value according to the color bar, while the right panel shows the monthly wind speed, where the red line is the monthly mean wind speed, the upper and lower limits of the box 75- and 25-percentile and the end-bars represents the maximum and minimum monthly mean wind speed.
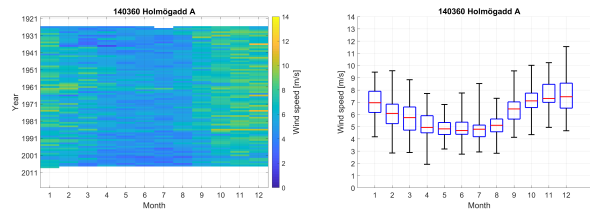
Figure A11. Data screening plot for Holmögadd. Left panel shows the data coverage and the monthly mean wind speed value according to the color bar, while the right panel shows the monthly wind speed, where the red line is the monthly mean wind speed, the upper and lower limits of the box 75- and 25-percentile and the end-bars represents the maximum and minimum monthly mean wind speed.
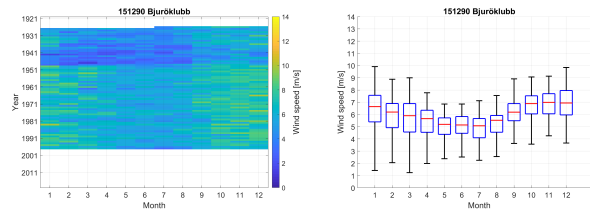
Figure A12. Data screening plot for Bjuröklubb. Left panel shows the data coverage and the monthly mean wind speed value according to the color bar, while the right panel shows the monthly wind speed, where the red line is the monthly mean wind speed, the upper and lower limits of the box 75- and 25-percentile and the end-bars represents the maximum and minimum monthly mean wind speed.
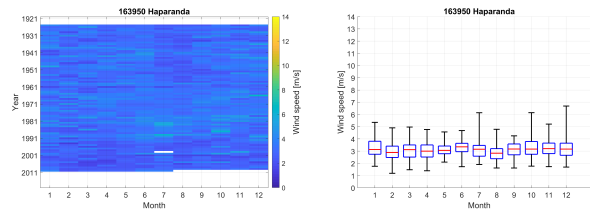
Figure A13. Data screening plot for Haparanda. Left panel shows the data coverage and the monthly mean wind speed value according to the color bar, while the right panel shows the monthly wind speed, where the red line is the monthly mean wind speed, the upper and lower limits of the box 75- and 25-percentile and the end-bars represents the maximum and minimum monthly mean wind speed.