https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

SEA-HISTIN: An inventory of historical weather observations of Southeast Asia

Praveen Teleti[1,2], Fiona Williamson[1]

[1]College of Integrative Studies, Singapore Management University, 10 Canning Rise, Singapore
[2]Dept of Meteorology, University of Reading, United Kingdom

Correspondence to: Praveen Teleti (praveent@smu.edu.sg)

Abstract. Understanding past climate is essential for extreme weather risk preparedness and planning. Southeast Asia suffers from an acute scarcity of historical weather observations, partly due to a lack of information on the existence and location of historical data sources. Here, we address this challenge by presenting SEA-HISTIN, the largest known inventory of historical weather data sources for Southeast Asia, cataloguing diverse materials ranging from government and institutional records, bulletins, and ship logbooks, identified across archives in the UK, Malaysia, Singapore, Indonesia, the Netherlands, Japan, and the United States. The inventory documents over 1,900 terrestrial station records and more than 10,000 marine voyages spanning roughly two centuries, with detailed metadata on location, observation period, variables, instrumentation, and data source quality compiled following international best-practice guidelines. Its relational structure separates source documents from individual stations and ships, supporting flexible extension as new record types are discovered. Beyond its scientific value for reconstructing rainfall trends, droughts, and floods, the inventory connects historical weather to broader social history, including epidemics, famines, and migration patterns shaped by climate extremes. Released as an open, versioned dataset, SEA-HISTIN is designed to guide future data-rescue priorities, foster international collaboration through initiatives such as ACRE, and make historical climate records more accessible to researchers and the public alike.

Short summary. Southeast Asia suffers from acute scarcity of historical weather observations due to lack of information on existence and location of historical data-sources. Here, we address the challenge by presenting the largest inventory of historical data-sources for Southeast Asia, cataloguing diverse sources. The inventory can record comprehensive meta-data yet flexible to accommodate new data-sources. We hope this inventory will encourage more data-rescue efforts and fill in the existing data gap.

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

1 Introduction

Long, continuous records of past weather are the empirical foundation of our understanding of climate variability and change. They allow the frequency and severity of extreme events such as floods, droughts to be characterised and long-term trends to be quantified. The scientific value of historic observational data for reconstructing long-term climate patterns has long been recognised (Brönnimann et al., 2019; Brunet and Jones, 2011), and demand for such records has grown sharply with the development of century-scale dynamical reanalyses. Products such as the NOAA-CIRES-DOE Twentieth Century Reanalysis, ECMWF CERA-20C (Laloyaux et al., 2018) reconstruct the past climate, depend directly on the density and quality of the historical observations available for assimilation (Compo et al., 2011; Slivinski et al., 2019). Where observations are dense, these reanalyses are well constrained; where they are sparse, uncertainty grows and systematic biases emerge. The recovery of historical observations is therefore not merely an exercise in preservation but essential to improving our quantitative understanding of past climate.

Southeast Asia is among the most poorly constrained regions in these global products (Slivinski et al., 2019). While millions of people are exposed to monsoon variability, and flood and drought hazards, it is critically important to identify not only climatological norms but the likelihood and scale of extreme rainfall events using past events (Su et al. 2026). Improving the historical baseline of climate data for long-term modelling and prediction is therefore essential for policy makers and agriculturalists to improve disaster planning and preparation (Lessa et al. 2016, Lassa, 2013). Widespread collection of historical weather observations in Southeast Asia has been difficult however, not least because many of the countries and provinces within this region have experienced unstable political histories, affecting the collation and retention of historical records. The region contributes comparatively few observations to international databanks for the period before the mid-twentieth century.

Crucially, this scarcity is not, mostly, due to observations never having been made. Systematic instrumental observation began relatively early across the region, driven by the scientific curiosity and administrative needs of colonial governments. For example, in the British Straits Settlements, meteorological observation dates to at least the 1840s, sustained in its early decades by the enthusiasm of individual military and medical officers rather than by dedicated scientific institutions (Williamson et al., 2015). In the Philippines, the Jesuit-run Observatorio Meteorológico de Manila began systematic daily observations in 1865 and became a pre-eminent centre for typhoon forecasting in the Far East before its transfer to the American colonial administration in 1901 (Williamson et al., 2018). Comparable colonial observing efforts existed under Dutch administration in the East Indies and French administration in Indochina. The present-day scarcity is thus largely a product of historical and political circumstance rather than of absent observations.

Several intertwined factors explain why so much of this record is now difficult to access. First, the region's fragmented colonial past means that surviving records for a single territory are frequently dispersed across archives on several continents, having changed custody as territories passed between imperial powers (Williamson, 2021). Second, the Second World War and the Japanese occupation of 1942 to 1945 caused the deliberate and accidental

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

destruction of records, as well as their removal to Japan; the near-total destruction of the Manila Weather Bureau's holdings during the 1945 Battle of Manila is an especially stark example of irrecoverable loss (PAGASA, 2026). Third, in the turbulent post-war and decolonisation periods, meteorological records were often not treated as a preservation priority, and the high-humidity tropical climate has hastened the physical deterioration of paper documents held in unsuitable conditions. As a result, the historical observational record for Southeast Asia is fragmentary, scattered, and in places entirely absent, with the gaps deepening further back in time (Williamson et al., 2015).

1.1 Data rescue in Southeast Asia: progress and its limits

Recognition of this problem has motivated a growing number of international data-rescue efforts. Data rescue is an established area of archival fieldwork, informing several important climate studies on historical climate and its changes by providing the basis for many global climate datasets and generated models (Teleti et al., 2026; Hawkins et al., 2023; Slivinski, 2021; Nash et al. 2021; Brugnara et al., 2019; Wilkinson et al. 2011; Brunet and Jones, 2011; Compo et al., 2011). It is a threefold process of locating, imaging, and digitising instrumental weather observations from historical documentary records created by societies, such as weather stations, ship logbooks, diaries, private and public institutional records, including hospitals and botanic gardens, and official meteorological records.

Figure 1. Study region of the SEA-HISTIN inventory. The main panel shows the Southeast Asian countries together with the maritime corridors that cross the region that define the inventory's focus region (dark green); the inset places the region's extent (red box) on the globe.

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

These recovery efforts are coordinated globally by the Atmospheric Circulation Reconstructions over the Earth (ACRE; Allan et al., 2011) and its regional chapters that connect national meteorological and hydrological services, archives, and academic institutions to recover historical observations and channel them into international repositories. The regional focus ACRE Southeast Asia was established within this framework specifically to build capacity within regional institutions and to recover the region's dispersed instrumental documentary records (Williamson et al., 2015).

Several data-rescue projects have since made important progress within Southeast Asia, though each has necessarily been bounded in geographic or institutional scope. For Indonesia, the most substantial effort has been the Digitisasi Data Historis (DiDaH) project, a collaboration between the meteorological services of Indonesia (BMKG) and the Netherlands (KNMI), which has digitised large volumes of historical Indonesian climate data covering the period from around 1850 and made them available through the Southeast Asian Climate Assessment & Dataset (SACA&D), a regional analogue of the European Climate Assessment & Dataset (van den Besselaar et al., 2015). SACA&D now holds many thousands of daily observation series, and its digitised station data underpin the SA-OBS gridded product for the region (van den Besselaar et al., 2017). However, its scope is confined largely to Indonesia and to the digitised daily record, and the majority of its constituent series remain access-restricted, limiting its usefulness as a region-wide picture of what source material survives.

Williamson (2021) made the first systematic effort to compile fragmented records of historical observations in the former British colony of the Straits Settlements, which included Malacca, Penang and parts of Perak, and Singapore and an analysis of some of the historic rainfall data for Singapore was presented by Gao et al in 2018. This work has demonstrated both the richness of the surviving Straits Settlements record and the labour-intensive, cross-disciplinary effort required to recover it, combining archival, historical, and meteorological expertise. More recently, the long-neglected countries of former French Indochina, have been addressed in the Climate data rescue of Vietnam, Cambodia and Laos database, covering 1867 to 1973 (Thomas et al. 2025). Drawing on nine archives across Vietnam and France, filling a large gap in the regional record. Together with the Indonesian and Straits Settlements data-rescue efforts, this database extends the recoverable frontier into the French mainland territories, though its core data remain embargoed to project researchers until 2029.

For marine records, there are no marine data-rescue projects dedicated specifically to Southeast Asia. However, global marine data-rescue projects such as RECLAIM (RECovery of Logbooks and International Marine data; Wilkinson et al. 2011) project and CLIWOC (Climatological Database for the World's Oceans; García-Herrera et al. 2005) coordinates to facilitate the data-rescue through imaging and digitisation of archived global marine data sources. Recently, a marine dataset was published by Teleti et al. (2026) including observations taken in the Southeast Asian waters e.g. Malacca Straits, South China Seas in the nineteenth century. A team of researchers at the Japan-Asia Climate Data Program made marine data available from Dutch ship logs collated in a series of joint-collaborative projects between Japanese and Dutch researchers to recover and digitise extant logbook information covering waters around Japan, South China Seas (Tsukahara et al., 2026).

1.2 A regional source inventory

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

Despite this progress, regional data-rescue activity remains fragmented along national and institutional lines, and a critical prerequisite for coordinating it across Southeast Asia as a whole has been missing: a comprehensive, region-wide inventory of what historical weather records exist, where they are held, what variables and periods they cover, and how accessible they are. Individual projects have compiled inventories for specific countries or institutions, but no single, systematically structured catalogue has spanned the region's terrestrial and marine sources across national and colonial boundaries.

This distinction between a source inventory and a digitised dataset is important. Existing regional datasets such as SACA&D catalogue observations that have already been digitised; they do not document the far larger body of source material that survives in archives but has yet to be imaged or transcribed. Without an inventory of this kind, scarce data-rescue resources cannot be prioritised effectively, and there is a persistent risk of duplicated effort, overlooked sources, and the continued loss of records whose very existence is undocumented. The scale of the undocumented material is considerable: the number of undiscovered or uncatalogued historical weather observations for the region has been estimated to run into the millions (Wilkinson et al., 2019).

1.3 Scope and aim of this work

Here we address this gap by presenting SEA-HISTIN, a relational inventory of historical weather observation sources for Southeast Asia. We define the scope of the inventory deliberately along three dimensions. Geographically, it covers both mainland and maritime Southeast Asia broadly from the Malay Peninsula through the Indonesian archipelago to the Philippines together with the maritime corridors that cross the region, reflecting the fact that historical observing networks were organised by colonial administrations and shipping routes rather than by modern national borders (Figure 1). Temporally, it spans the early instrumental period from the late nineteenth century to the mid-to-late twentieth century, capturing the transition from ad hoc observation by ports, missions, plantations, and ships to standardised reporting by national meteorological services. In terms of source type, it encompasses government and institutional records, bulletins and compendia, newspapers, and both naval and mercantile ship logbooks.

It is important to be clear about the scientific aim of this work. SEA-HISTIN is an inventory of sources, not a database of digitised observations: rather than transcribing observations directly, it systematically catalogues the source documents from which observations can be recovered, recording standardised metadata on location, temporal coverage, observed variables, data format, provenance, and data-rescue readiness for each source. It is intended as a planning and coordination instrument for the data-rescue community, establishing what exists and where, so that subsequent imaging, transcription, and quality control can be directed where they yield the greatest scientific return. In this respect it is complementary to climate datasets such as ISPD (Cram et al. 2015, Copernicus Climate Change Service (C3S 2026), and is designed to feed material toward them and toward international repositories.

The inventory brings together previously catalogued sources for Singapore and Malaysia (Williamson 2021) with a large body of newly identified material spanning a wider geographic area including Sumatra and the wider Indonesian archipelago, the Philippines, Thailand, and Borneo alongside marine records. The inventory is built on a

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

flexible relational structure that separates source documents from individual observing platforms, allowing it to be extended with new record types as they are discovered, and it is released as a versioned, machine-readable resource compliant with FAIR data principles. Although the data sources tend to emphasize on numerical weather observations, but these data sources are also excellent record of socio-economic history of the region. This inventory, is hoped, will be used to discover and understand impact of past climate on Southeast Asian society, and will be basis for many inter-disciplinary studies in the future.

The remainder of this paper is organised as follows. Section 2 describes the historical context of weather observation in Southeast Asia and surveys the major data sources incorporated into the inventory. Section 3 sets out the methodology, covering source discovery, inclusion criteria, metadata extraction, and the relational structure of the inventory. Section 4 summarises the coverage and characteristics of the assembled inventory, discusses its scientific and societal applications, and outlines priorities for future data rescue across the region.

Figure 2. Meteorological air balloon launch at Kallang and Paya Lebar Meteorological Station, 1955. Image courtesy of National Archives of Singapore (Ministry of Information and the Arts Collection, Media Image No. 19980001113).

2 Data Sources

Records of historical weather observations from Southeast Asian countries and regions are stored and archived at a variety of different places. In the case of Singapore and Malaysia, for example, weather observations were contained in documents that changed hands on multiple occasions due to those nation's particular colonial histories. Early

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

records until the mid-1860s belonged to the East India Company but, after the company's dissolution, those records were sent to repositories in India and Britain, much eventually ending up at the British Library in London. From the late 1860s, weather data were collated by different government departments, including the Medical Department and Harbour Department, records from which are now scattered across different repositories in Singapore and Malaysia, with a few in the UK.

By the 20th century, meteorological records were more standardised under the official institutions of the British Colonial Office. These included the Straits Settlements and Federated Malay States Agricultural Department and the official Malayan Meteorological Service (the latter from 1929). However, many of these records were removed to Japan in 1942 due to the occupation of Malaya by the Japanese military or held in situ in Syonan-to (Singapore), the location of the Japanese meteorological headquarters for Southeast Asia from 1942-1945. In 1945, these records were returned officially but, many had either been deliberately or unintentionally destroyed or discarded during this turbulent period of history. Post-independence weather observatories were passed to Meteorological Service Singapore (MSS), a visual showing launch of weather balloon at Kallang and Payar Lebar Meteorological Station in 1955 (Figure 2).

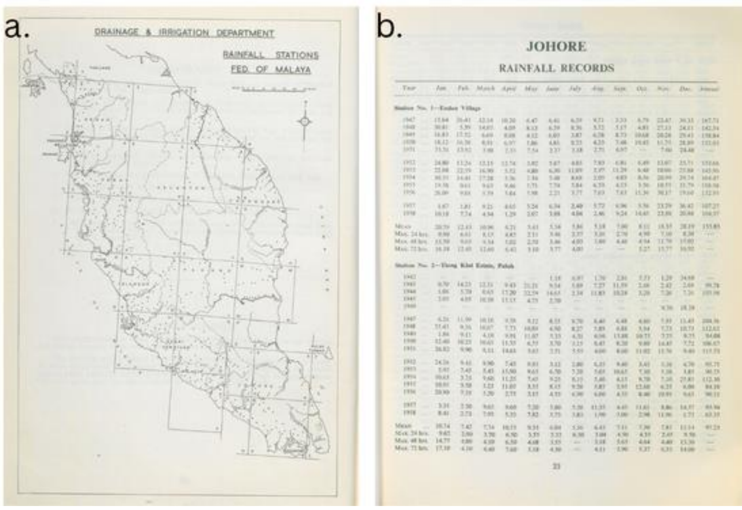

Figure 3. Rainfall Records, 1879–1958, was published by the Drainage and Irrigation Department, Federation of Malaya. It is an official government compilation presenting nearly 80 years of rainfall observations collected across the Federation of Malaya. Map of all stations tabulated in the report (a) and a typical data page of the report (b). Image courtesy of NLB Singapore (Malaya. Drainage and Irrigation Department)

Some of the major data sources are colonial government publications (Federated Malay States Agricultural Bulletins, Malayan Agricultural Journal, Straits Settlements Gazette, Singapore Botanic Gardens bulletins), hydrological compilations (Federation of Malaya rainfall records 1879–1958; Figure 3), British Borneo rainfall

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

statistics 1896–1957, Dutch East Indies observatory records 1900–1939), regional compendia (the ASEAN Compendium of Climate Statistics covering more than 900 stations across five countries) and local newspapers. For the scarce World War Two period, there are Japanese observations archived at JAMSTEC and at the UK Meteorological Office climatological returns for Malayan and Singapore stations. A substantial body of oceanic records survive in ship logbooks for the British Royal Navy held at the UK Hydrographic Office and for the US Navy at the US National Archives and Records Administration (NARA). The marine sources presented here contains ships' traversing across the globe not just Southeast Asian waters, however, many voyages included Southeast Asian waters. This collection thus represents one of the most comprehensive assemblages of pre-modern instrumental climate records for tropical Southeast Asia, with particular emphasis on climate records during the colonial and early post-colonial periods.

All these records were stored in paper form or scanned, but over the time some of these records have been transcribed in machine readable format. To produce a simple and clear inventory of newly discovered and known data sources, we have grouped these sources as un-transcribed and transcribed. The Table 1 provides the list of major data-sources included in the inventory with various details. More sources in the inventory and their detailed description and historical context are provided in Appendix A.

Data Source	Region/Country	Source Type	Author	Archive	Years(s) Covered	Previously catalogued	Data Rescue Status
Un-transcribed							
Federated Malaya States' Agricultural Bulletin	Malaysia and Singapore	Terrestrial	Dept. of Agriculture, Federated Malay States	NLB Singapore	1912-1921	No	Imaged
Malayan Agricultural Journal	Malaysia and Singapore	Terrestrial	Ministry of Agriculture for Malaya	NLB Singapore	1922 - 1963	No	Imaged
Gardens' Bulletin Straits Settlements	Malaysia and Singapore	Terrestrial	Singapore Botanic Gardens	National University of Singapore Libraries/Singapore Botanic Gardens Library of Botany and Horticulture	1912 - 1939	No	Imaged
Botanic Gardens Monthly observations	Malaysia and Singapore	Terrestrial	Malayan Meteorological Service	Singapore Botanic Gardens Library of Botany and	1929-1957	No	Imaged

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

				Horticulture			
Rainfall Statistics of the British Borneo Territories	Malaysia	Terrestrial	Department of Civil Aviation and Meteorological Services	Personal Collection, Rory Walsh	1896 - 1957	No	Imaged
Meteorological Observatory in Batavia	Indonesia	Terrestrial	Meteorological and Geophysical Service; Royal Magnetic and Meteorological Observatory (Batavia)	Leiden University Library	1900 - 1945	No	Imaged
Monthly Meteorological Bulletin	Thailand	Terrestrial	Meteorological Section, Hydrographic Service Thailand	NOAA Library	1930s	No	Imaged
Climate summaries and Compendium	ASEAN	Terrestrial	ASEAN Meteorological Service	Personal Collection, Rory Walsh	1951 - 1975	No	Imaged
JAMSTEC meteorological observations	Asia and Oceania	Terrestrial	Japan Agency for Marine-Earth Science and Technology (JAMSTEC)	JAMSTEC archive	1940 - 1941	Yes	Imaged
Meteorological Office climatological returns	Singapore	Terrestrial	Meteorological Office Singapore	UKMO archive	1860 - 1971	Yes	Imaged
Straits Settlements Government Gazette	Malaysia and Singapore	Terrestrial	Straits Settlements Government	National Library of Singapore, National University of Singapore	1860s-1930s	No	Imaged
Captains' logbooks	Global	Marine	British Royal Navy	National Meteorological Archive, Exeter	1860 - 1900	No	Imaged
Meteorological records	High-latitude Southern Ocean and Global	Marine	British Antarctic Survey, Cambridge	British Antarctic Survey Archive	1900-1960	No	Imaged
Royal Navy Harbour Vessels' logbooks	China and Southeast Asia	Marine	British Royal Navy	National Archives, Kew	1860-1920	No	Imaged
Mercantile ships'	Global	Marine	Åland Maritime	Åland	1880 - 1940	No	Imaged

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

logbooks			Museum	Maritime Museum Archives			
Mercantile and fishing ships' logbooks	Global	Marine	Maritime Institute of the Åbo Akademi - Turku Finland	Maritime Institute of the Åbo Akademi - Turku Finland	1890 - 1930	No	Imaged
Royal Navy Yangtze Gunboats' logbooks	Coastal China	Marine	British Royal Navy	National Archives Kew	1850 - 1900	No	Imaged
Royal Navy ships' logbooks	Japan	Marine	British Royal Navy	National Archives Kew	1880 - 1900	No	Imaged
Individual Remark Books	Global	Marine	British Royal Navy	UK Hydrographic Office, Taunton	1860 - 1920	No	Imaged
Masters' ship logbooks	Oceania and Asia	Marine	British Royal Navy	National Meteorological Archive, Exeter	1855 - 1925	No	Imaged
Captains' logbooks	Southeast Asia	Marine	British Royal Navy	National Meteorological Archive, Exeter	1865 - 1872	No	Imaged
US Navy ships' logbooks	Asia	Marine	US Navy ships	National Archives and Records Administration (NARA)	1940 - 1970	No	Imaged
Transcribed							
Weather Observations	Penang, Malaysia	Terrestrial	Captain Francis Light, superintendent, Penang	NLB Singapore	1786	Yes	Transcribed
Meteorological observations	Maleka, Malaysia	Terrestrial	William Farquhar, British resident at Malacca	NLB Singapore	1809	Yes	Transcribed
Weather observations	Fort Canning, Singapore	Terrestrial	Charles Edward Davis, military staff officer, EEIC	NLB Singapore	1820- 1824	Yes	Transcribed
Daily weather	Singapore	Terrestrial	The Singapore	NLB	1835 - 1840	Yes	Transcribed

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

observations			Free Press and Mercantile Advertiser	Singapore			
MacRitchie Reservoir Monthly Rainfall observations	Singapore	Terrestrial	Municipal Engineer	NLB Singapore	1879 - 1948	Yes	Transcribed
Weather observations	Penang, Malaysia	Terrestrial	T. Irvine Rowell, principal civil medical officer	NLB Singapore	1885 - 1917	Yes	Transcribed
Weather observations	Province Wellesley, Malaysia	Terrestrial	The colonial surgeon (various)	NLB Singapore	1896 - 1915	Yes	Transcribed
Meteorological Observations	Christmas Island, Australia	Terrestrial	Straits Settlements Govt Gazette	NLB Singapore	1901–1952	Yes	Transcribed
Kuching Observatory (Museum) weather observations	Sarawak, Malaysia	Terrestrial	Sarawak Government Gazette	NLB Singapore	1906 - 1955	No	Transcribed
Bintulu Sarawak Climatological Summary	Sarawak, Malaysia	Terrestrial	Bintulu Sarawak Climatological Summary	NLB Singapore	1954 - 1955	No	Transcribed

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

Table 1. The main historical weather data sources identified and recorded in the inventory, detailed information and historical context can be found in the Appendix A. The table does not contain all data sources recorded in the inventory to keep the table concise.

3 Methodology

Constructing a comprehensive inventory of historical weather data sources for Southeast Asia requires confronting challenges that are at once archival, geographic, and technical. The region's colonial history has distributed records across multiple national archives on several continents, often without any transfer trail. Institutional disruptions particularly the Japanese occupation of 1942–1945 and the subsequent decolonisation process further fragmented the observational record, dispersing holdings and, in some cases, destroying documents entirely. Hence, a deliberate broad discovery approach was chosen to build this inventory.

The construction methodology described below was designed with two overarching priorities. First, to maximise discovery of new historical weather documents, we cast as wide a net as possible across institution types, archive locations, and document formats. Second, to ensure that the resulting inventory is internally consistent, extensible, and aligned with internationally recognised data-rescue standards, so that it can serve as a practical planning tool for future digitisation efforts rather than merely a bibliographic catalogue.

As one of the first structured attempts to catalogue multi-institutional historical weather data sources for Southeast Asia, the schema was designed with built-in flexibility. The inventory is implemented in three parts: primary table, secondary tables and platform-level tables. The primary table holds meta-data of the historical documents catalogued and forms core component of the inventory. For example, names of holding institution, producing agency/author, call number or URL at the archive, document data rights and digitization status, publication year etc. However, a single document may contain several stations or ships, to capture those details in the data-sources secondary tables are used. For example, original and standardised station name, region and country currently located in, geographical location (latitude-longitude), elevation and distance to coast, instrument type etc. for terrestrial stations are stored in the terrestrial secondary table.

Similarly for marine platforms, e.g. ships, title and description of documents, vessel/ship type, owner or port of registry, vessel movements, notes on meteorology and oceanography are stored in the marine secondary table. Any number of secondary tables can then be attached to primary table to accommodate more platform types, for example, early aerial observations' data sources can be added without altering the underlying relational structure. This extensibility is especially important in a region where archival discovery is ongoing and where new source categories continue to emerge. Furthermore, platform-level attributes can be added to platform-level tables to enhance meta-data for stations or marine vessels, for example, using additional tables for observers, instrument attributes (e.g., rain-gauge rim height, screen orientation) etc.

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

The focus region was defined to include both mainland and maritime Southeast Asia (broadly from the Malay Peninsula through the Indonesian archipelago to the Philippines as defined by Britannica, 2026), acknowledging that many historical observation networks were organized by colonial administrations and maritime routes rather than by modern national borders. Also, by including marine data sources, the inventory would be useful to recover marine and coastal weather observations within Southeast Asia. Temporal coverage was set to the early instrumental period through the mid twentieth century, capturing the transition from ad hoc observations (e.g., ports, missions, plantations, ships) to national meteorological services and standardized reporting.

3.1 Source Discovery

The data sources and archives listed in this inventory has been slowly and painstakingly researched and collated over last two decades by the authors. The library catalogues and repositories were searched for following terms to find historical weather documents "meteorological observations", "climatological returns", "weather register", "rainfall register", "rain gauge records", "temperature returns", "climatological summary", "climatological bulletin", "weather report", "meteorological service", "observatory", "meteorological station", "weather station", "agricultural bulletin", "agricultural department", "medical officer", "colonial surgeon" (common authors of early observations), "government gazette" etc.

The primary language of search and items catalogued were in English, however, Japanese, Thai, and Dutch medium documents have been consulted and added to the inventory. During this project, our team have discovered, identified and collated large number of previously unknown and uncatalogued weather records in many archives and unexpected places across the UK, Netherlands, Malaysia, Singapore and the U.S.

We employed a multi-archival discovery approach designed to minimise single-institution bias and to capture both government and non-government records across the full diversity of source types present for the region. Relying on a single institutional stream, for example, national meteorological service holdings alone would systematically underrepresent records held in other repositories, private collections, and online historical archives, and would limit our understanding of past climate to weather stations' environs, miss lesser-known times and locations elsewhere.

Negative Searches

Historical weather data sources turn up at unexpected places but also not found in very likely places. To illustrate, we looked for documents with weather observations taken at once numerous rubber plantations across Malaya region, e.g. rubber plantation reports and bulletins at various places including Malaysia and Singapore. No useful documents were found in these places, but we could only find in 2023 a report on rainfall impact on rubber plantations and their yield (Rainfall in Malaysia. A study of its occurrence, with tables of probability of rainfall at selected stations, and an introduction to hydrology in rubber plantations (Wycherley 1967) at Cambridge University

Library published by Rubber Research Institute of Malaya (RRIM, Kaula Lumpur) in 1967. However, the report only contained climate summaries, but it contained references to raw observations used to construct the summaries. Those references helped us to find Malaysian Irrigation Dept's rainfall data compilation at NLB Singapore. Discovery proceeded iteratively: references encountered within identified documents frequently pointed to further sources, and catalogue cross-references in library holdings were followed systematically. This cascading discovery process was essential for recovering sources that were not discoverable through any single keyword search or subject catalogue, particularly for early-twentieth-century records where cataloguing practices were inconsistent.

Similarly, we searched for weather data published in Planters' bulletin published by RRIM now archived at National Library of Scotland (NLS) but couldn't find any relevant material. Also, 'The Planter' journal published by International Society of Planters was searched for relevant data, but none was found. At the same archives NSL Catalogue of Manuscripts, historically significant work on magnetism and meteorology by Allan Broun was consulted but did not yield any relevant weather observations.

For unofficial weather observations or reports, we turned to archives of missionary societies, as Christian missionaries were active in the region during 19th and 20th Century living and working in many rural and interior regions of Malaya where neither plantation nor govt weather stations existed. Similar missionary archives have been used previously to reconstruct past climate elsewhere (Nash et al. 2021). For Malaysia and Singapore, we consulted archives at Salvation Army International Heritage Centre, William Booth College, London and Malaya Mission Council, Presbyterian Church at SOAS London. Unfortunately, none of the materials archived contained any weather observations.

Sometimes, historical weather data sources are found at unexpected places. Multi-national cooperative agency ASEAN complied climate atlas of whole Southeast Asian region and published a compendium of weather observations from across the region in 1975. The compendium was never institutionally archived, it was by chance discovered to be with Rory Walsh, researcher at University of Swansea, UK. We have now imaged it and added to our inventory. A climate statistic of British Borneo and Sarawak was found with the same researcher, which has been now imaged and catalogued, is part of the presented inventory (See Appendix A for more details).

Four complementary discovery streams were used and described below:

1. **National meteorological services and government archives**: We screened annual reports, monthly bulletins, station registers, and technical manuals for station operations, instrument types, observation schedules, and tabulated data. Naturally, we started off our search for historical weather records at Centre for Climate Research Singapore (CCRS) archives under Meteorological Services Singapore (MSS), Malaysian Meteorological Department (METMalaysia). MetOffice UK maintained stations and published weather observations before local weather services were formed, we searched MetOffice UK' National Meteorological archive (NMA) for relevant documents.

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

Where records were not directly available for inspection, catalogue entries and finding aids were used to identify candidate holdings for follow-up.

2. **Colonial and overseas administrative records**: Because much of Southeast Asia's early meteorological documentation was produced under colonial infrastructures, British in Malaya, Singapore, Borneo, and Burma; Dutch in the East Indies; American in the Philippines; French in Indochina; and Japanese across much of the region from 1941–1945, a substantial proportion of surviving records is held in colonial repositories rather than in the countries of origin. We therefore searched colonial repositories in the United Kingdom (British Library, National Archives Kew, UK Meteorological Office archive), the Netherlands (Leiden University Library, KNMI), the United States (NARA, NOAA Library), and Japan (JAMSTEC) for references to station operations, instrument types, observation schedules, and tabulated data. During the search monthly weather observations were found in Straits Settlement govt gazette, weather observations in Agricultural and garden bulletin

3. **Maritime and naval records**: Ship logbooks and voyage meteorological registers constitute a distinct and frequently underutilised source category. Unlike fixed terrestrial stations, shipborne observations provide spatially continuous records along maritime routes and can fill observational gaps in the data-sparse ocean regions of Southeast Asia. We targeted Royal Navy and mercantile vessel logbooks held at the UK Hydrographic Office (Taunton), the National Archives (Kew), and the US National Archives and Records Administration (NARA), as well as holdings at the Åland Maritime Museum and Åbo Akademi Maritime Institute in Finland. These sources were treated as distinct mobile platforms rather than fixed stations, reflecting their mobile observation taking and the different metadata fields required for their catalogue representation. The whole ship journeys are catalogued in the inventory rather than sections were vessel entered into Southeast Asian waters.

4. **Libraries and digital repositories**: We systematically queried university library collections, national libraries National Library Board (NLB) Singapore, National University of Singapore Libraries), and digitised historical book databases including HathiTrust (HathiTrust 2026), the NOAA Central Library (NOAA 2026), Leiden University Library (LUL 2026), and the Biodiversity Heritage Library (BHL 2026) for meteorological bulletins, regional atlases, and agricultural journals that reprinted station data. Regional atlases and institutional bulletins are particularly valuable because they aggregate data from multiple stations in a single document, effectively condensing the discovery effort for an entire network into a single source.

3.2 Inclusion criteria

Materials were included in the inventory if they contained quantitative or semi-quantitative meteorological observations either instrumental readings or formally coded observations accompanied by at least positional or location reference. Specifically, an eligible data-source had to satisfy three conditions: (i) a temporal reference of at least a year and month, enabling placement within a calendar timeline; (ii) an identifiable observation location, whether with geographic coordinates for a fixed station or a position derivable from ship track; and (iii) at least one meteorological element, such as rainfall, temperature, atmospheric pressure, wind speed or direction, humidity, or sunshine duration.

Narrative-only weather descriptions, without associated measurable quantities, were excluded. For example, newspaper reports describing weather as 'heavy rains' or 'unusually dry' without accompanying quantitative data do not satisfy criterion (iii), but newspaper that contained daily weather reports were included in the inventory. Similarly, documents containing only climatological statistics derived from earlier primary observations were included in the main inventory but noted as secondary sources references to original observations where available. These distinctions are important for downstream data-rescue users, who need to know whether a source contains raw instrumental data suitable for digitisation or only processed summary statistics.

No minimum temporal duration was imposed as an inclusion threshold. Single-year observations, isolated monthly records, and incomplete series are valuable precisely because they fill spatial and temporal gaps that are characteristic of the pre-twentieth-century observational network in Southeast Asia, and excluding short records would systematically bias the inventory toward established long-running stations at the expense of rare or opportunistic observations.

3.3 Metadata extraction

Metadata extraction follows established international guidelines for historical climate data rescue, specifically the WMO Guidelines on Climate Data Rescue (DARE; WMO-No. 1182) and the Copernicus Climate Change Service Best Practice Guidelines for Climate Data Rescue (Wilkinson et al., 2019) and for Keying Data from Historic Marine Documents (Wilkinson and Freeman, 2021). These guidelines specify minimum metadata standards for observing platform documentation, quality flags, and FAIR data principles, and adopting them ensures interoperability with global climate data rescue initiatives including ACRE and ICOADS.

Given the heterogeneity of data-sources in the current collection, we implemented a qualitative but systematic quality screening. Rather than assigning a single quality score for a data-source, which would obscure the distinct characteristics of a particular data-source, we created quality flags to assess features relevant to different downstream uses. As mentioned earlier, inventory consists of three components: primary table, secondary tables and platform-level tables. We store metadata related to the original document and its location, document quality flags in the primary table, secondary tables contain information on the stations, ships, whereas platform-level tables contain platform-level attributes, such as details of observers, instrument specifications (e.g., rain-gauge rim height, screen orientation) etc.

Primary table

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

Each record in the primary table contains minimum level of meta-data to reliably and uniquely identify the said data-source in an archive or a library. Each data-source entry contains information on type of data-source, quantitative flag, format of data, known gaps, language, name of the archive or library, call number, copyright status, digitized status, published date etc. The full custodial history of each source was recorded where known, including the original producing institution, subsequent transfers of custody, and current holding institution and access conditions (open access, restricted, or copyright-encumbered). Data-rescue status either digitised or un-digitised was recorded, with a link to the digitised version or transcription repository where available.

Records were classified by format as tabular printed or hand-written ledger, charts or diagrams, logbooks, or mixed format. The language or script of the document was recorded, together with a legibility and scan quality assessment (good, fair, or poor), which directly informs the transcription readiness rating described below. Transcription readiness rates the practical feasibility of digitisation from the existing physical or scanned record. Data-sources were classified as high (clear printed tables, good scan quality), medium (some ambiguity in figures, faded ink, or partial scan coverage), or low (damaged pages, dense handwriting, non-Latin script, or poor scan resolution). This rating is explicitly designed to support the prioritisation of data-rescue resources by identifying sources where digitisation is both scientifically valuable and operationally feasible.

Secondary tables

For more granular information, secondary table is populated from information extracted from stations or ships contained within the data-source. Each station, ship, or observing platform was described using a standardised template comprising five thematic categories:

a. Location and identification: Historical place names presented a particular challenge across the multilingual and multi-colonial context of Southeast Asia. Names were harmonised through systematic cross-referencing with the Getty Thesaurus of Geographic Name (Getty Research Institute 2026), colonial-era administrative atlases (Bartholomew 1912), and the GeoNames database (GeoNames 2026). When coordinates were not explicitly provided in the source document, locations were inferred from contextual descriptions — such as references to administrative centres, port facilities, or named estates — and flagged as 'inferred,' with a categorical uncertainty estimate (<5 km, 5–25 km, or >25 km) reflecting the specificity of the available contextual information.

Location metadata included geographic coordinates (decimal degrees), coordinate precision flag (exact, estimated, or inferred), elevation above sea level where recorded, and proximity descriptors (distance to coastline). Undocumented station relocations are a major source of artificial inhomogeneity in long historical series and must be flagged to enable downstream homogenisation. Historical name variants were recorded alongside modern place names to support downstream geocoding and deduplication. In case of marine data-source, name and movements of vessel are recorded.

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

b. Consistency: Each data-source is given internal consistency flag that captures the uniformity of units, notation, and format within a source document. Data-sources were classified as: (a) consistent: units and format are uniform throughout; (b) unit changes present: the document shifts between measurement systems (e.g., inches to millimetres) at a discernible point; (c) ambiguous notation: recording conventions are unclear or inconsistently applied; or (d) mixed conventions: multiple systems coexist with or without clear demarcation. This dimension is particularly relevant for sources spanning the transition from imperial to metric measurement systems, which occurred at different times across different colonial administrations in Southeast Asia.

c. Temporal coverage: Metadata fields captured station record or ship voyage start and end dates, calendar used, weather observation frequency (hourly, synoptic, daily, or monthly), stated observation times, and time standard (local mean time, standard time, or GMT where indicated). Time standard is frequently ambiguous in pre-twentieth-century sources; where not explicitly stated, it was flagged as unknown and noted in the provenance field.

d. Variables and units: The list of observed meteorological elements was recorded for each source, along with their original units (e.g., degrees Fahrenheit or Celsius, inches or millimetres, inches of mercury or millibars) and any stated reductions or corrections applied to the data (e.g., pressure reduced to mean sea level).

Platform-level tables

Platform-level tables attempt to answer, "what recorded the data and who recorded it," at a per-station or platform, time-resolved granularity, distinct from the secondary tables (which describe the observing entity as a whole) and the primary table (which describes the document). These tables store the details of physical measuring equipment at a platform or details of observers etc. linked to a specific station or platform.

a. The variable the instrument measures (rainfall, temperature, pressure, and so on), the instrument type (e.g. 8-inch rain gauge, Stevenson screen, syphon recorder), manufacturer or pattern where known, and its physical characteristics gauge diameter, rim height, sensor or screen height above ground, elevation, exposure and siting notes, original units, and any corrections applied. It also records calibration and service dates and a standardisation flag.

b. Observer attributes contain details of who made the observations at a platform, the observer's name, their role or office (medical officer, ship captain or master, estate manager, and so on), the responsible agency, and their period of tenure. Like the instrument table, it holds multiple dated rows per station, so a change of observer can be lined up against a suspected discontinuity in the record as observer changes being one of the most common causes of undocumented inhomogeneity, since a new keeper often meant a shift in observation time, exposure, or recording diligence. It also allows the individuals who kept the registers frequently named in colonial records to be credited and traced.

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

Together, these metadata fields and quality flags provide data-rescue teams with the information needed to locate the data-source and assess not only whether a source is worth rescuing, but also what challenges they are likely to encounter and what post-processing steps will be required before the extracted data can be used in scientific analyses.

3.5 Inventory structure and relational linking

The inventory is structured as a lightweight relational database comprising a primary table and in this instance two secondary tables (terrestrial and marine), with provision for additional secondary tables as new record categories are identified. A platform-level tables are also included containing instrument and other fine-grained attributes (Figure 4). This design was motivated by the need to represent many-to-many relationships between physical documents and individual observation platforms and their attributes, a structural complexity that flat tabular formats cannot accommodate without either data duplication or loss of provenance.

Each entity was assigned persistent identifier to enable stable citation and future updates. To support interoperability, fields were aligned with widely used climate data rescue conventions (e.g., station metadata and observation fields consistent with common climate database schemas), while preserving original notation in "raw text" fields to maintain provenance.

Primary table. Each row in the primary table represents a single physical document or archival item, identified by a unique Source ID that is never reused or modified once assigned. The primary table records document-level metadata: title, authoring institution, date of production, holding archive, call number or catalogue reference, format, and access conditions. A Series ID provides a unique identifier for each row independently of the Source ID, supporting internal cross-referencing.

Secondary tables. Each secondary table extends the primary table by cataloguing the individual observational platforms either terrestrial stations or ships documented within each source. In the terrestrial secondary table, each row represents a unique station-source combination, identified by a Station ID and linked to the corresponding primary table entry via Source ID as a foreign key. In the marine secondary table, each row represents a unique ship voyage or ship logbook, identified by a Ship ID and similarly linked via Source ID. Separating data sources (in primary table) and recorded stations/ships (in secondary table) allows multiple stations/ships published in a single physical document to be individually catalogued in the inventory.

Platform-level tables. The Platform-level tables hold fine-grained attributes for stations or marine vessels, e.g. tables for observers, instrument attributes (e.g., rain-gauge rim height, screen orientation) etc. For example, in the instrument attributes table, each row contains attributes of the instrument linked to a unique station via Station ID as a foreign key and to unique document source in the primary table via Source ID as a foreign key. Linking of

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

instrument attributes to all the way to source document preserves provenance of the meta-data recorded in the tables for the down-stream data-rescue users.

Instances where same station is observed and published in different documents, for example, a single station could have multiple series (e.g., pre-relocation and post-relocation segments, or separate variables recorded in different documents), or same stations appeared in multiple editions (e.g., reprints, summaries), multiple rows in the secondary table are generated but one per source document, each independently linked to the corresponding primary table entry. This foreign-key architecture means that all stations or voyages contained within a single physical document can be retrieved by filtering on their shared Source ID, and platform-level attributes can be retrieved by filtering on their shared Station ID or Ship ID. This process supports downstream digitization planning and prevents duplication (Figure 4).

This inventory operates with FAIR principles by using unique reference identifier for each data-source within it and persistent DOI link if available, for whole inventory, making it discoverable and reusable. Each inventory entry stores links to original scans and/or images where possible. Because entries are tracked at the station/ship level, downstream users can retrieve the exact scans or transcriptions used to make informed data-rescue decisions.

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

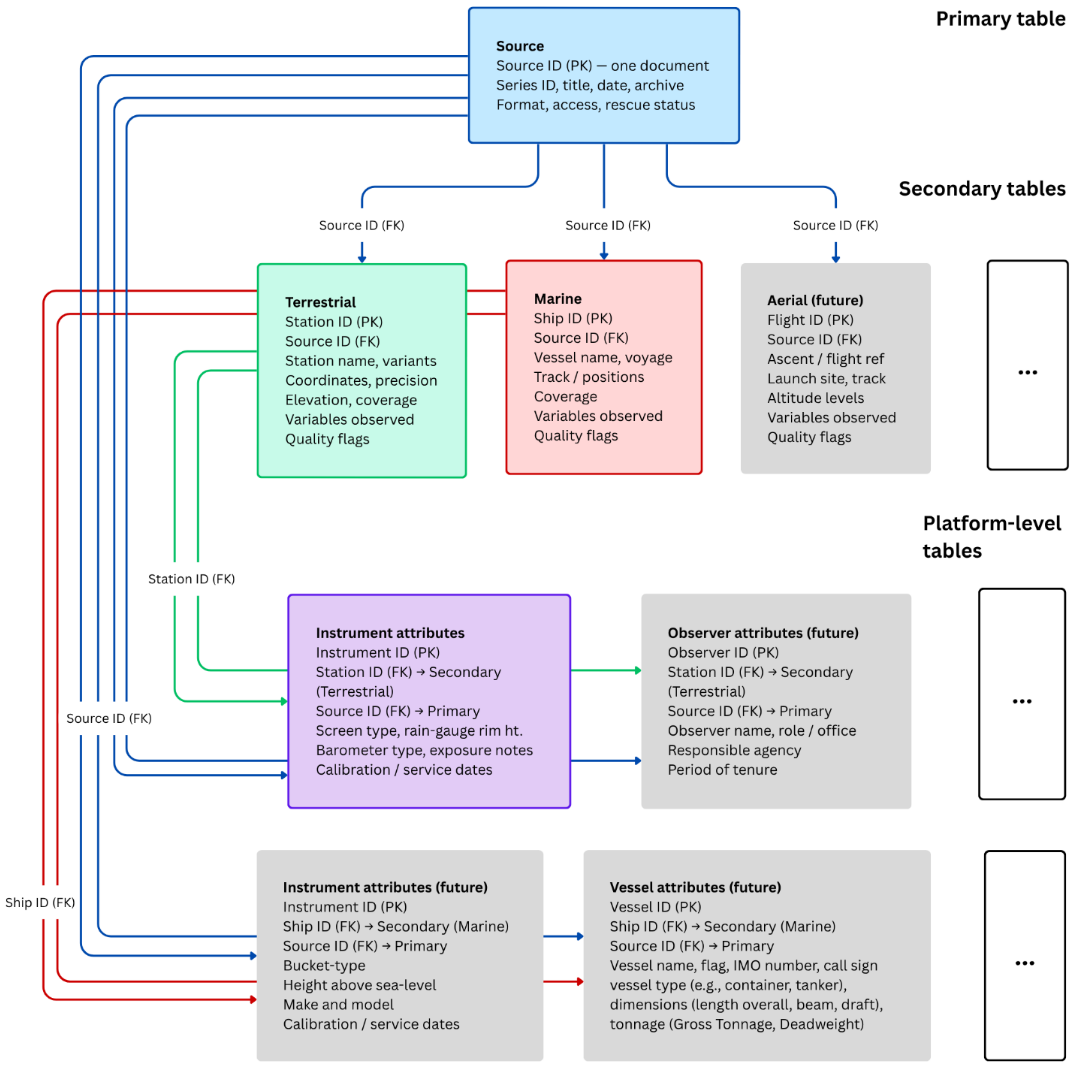

Figure 4. Relational structure of the SEA-HISTIN inventory. Physical documents are catalogued in the primary Source table (Source ID, PK). Observing platforms are recorded in secondary tables — Terrestrial and Marine — that reference the Source table via Source ID (FK), giving many-to-many linkage between documents and platforms. Platform-level metadata records: the Instrument attributes and Observer attributes table each reference individual stations via Station ID (FK) and Source ID (FK), allowing time-varying instrument and observer histories to be recorded per station or ship. The grey tables (Aerial, Observer, Vessel attributes) illustrate future extensibility new platform types, platform-level attributes can be added without altering the existing schema.

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

4 Summary

We have presented SEA-HISTIN, a relational inventory of historical weather observation sources for Southeast Asia. In this first published version (v0), the inventory comprises 183 source documents catalogued in the primary table, from which 1,984 individual terrestrial station records and 10,729 marine records (ship voyages) are documented in the two secondary tables. Of the source documents, 170 are terrestrial and 13 are marine in origin; 106 have already been transcribed and a further 77 imaged. Every secondary record resolve to a valid parent document, and the geographic coordinates place 1,980 of the 1,984 station records within the Southeast Asian domain, spanning approximately 10° S to 27° N and 93° E to 145° E. To demonstrate the analytical utility of the inventory, and to illustrate the kinds of diagnostic that its structure supports, we summarise its spatial, temporal, and variable-based characteristics below (Figure 5).

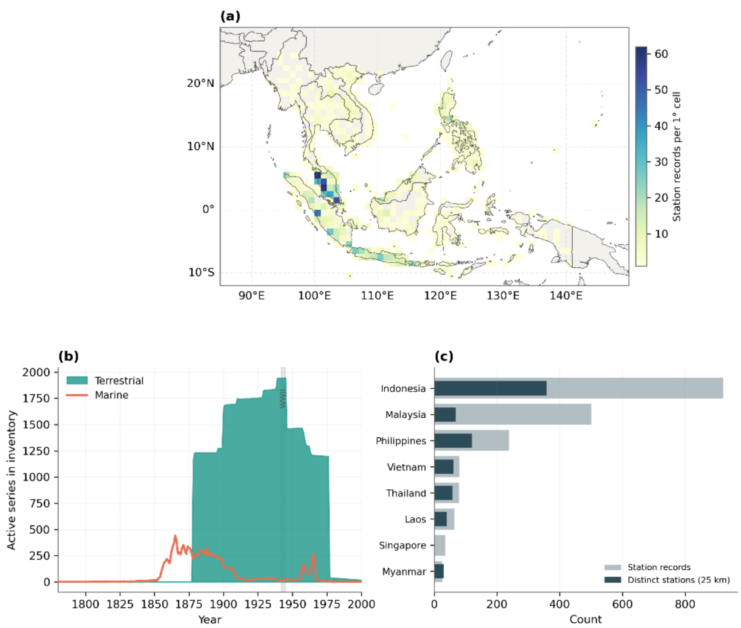

Figure 5. Spatial and temporal diagnostics of the SEA-HISTIN inventory. (a) Spatial density (binned 1° × 1° cell) of the 1,984 catalogued terrestrial stations across the Southeast Asian domain. (b) Temporal density, expressed as the number of active terrestrial (filled) and marine (line) series in each year from 1780 to 2000; the shaded band marks the Second World War. (c) Distribution of station records by modern country. Diagnostics are shown for the first published version of the inventory and reflect its current source base rather than the full universe of historical observations.

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

4.1 Spatial coverage

The catalogued stations span the full breadth of the region, from the northern mainland of Myanmar and Vietnam to the southern islands of the Indonesian archipelago, and from Sumatra eastward to the Philippines (Figure 5a). Grouped by modern nation, Indonesia is the most heavily represented, with 922 station records, followed by Malaysia (500), the Philippines (238), Vietnam (80), Thailand (79), and Laos (65); Singapore, Myanmar, Cambodia, and Brunei together account for a further 93 records, with a small number of stations falling in adjacent territories. West coast of Peninsular Malaysia has the highest density of stations per cell (Figure 5a). This distribution reflects both the historical density of colonial observing networks and the current state of source discovery, and it should be regarded as a map of catalogued sources rather than of all observations that once existed.

4.2 De-duplication

The 1,984 rows of the terrestrial secondary table represent station–source combinations rather than physically distinct stations: a single station reported across several source documents generates one row per document, and minor differences in inferred coordinates or transliterated names mean that the same location can appear under slightly different entries. The number of physically distinct stations in the inventory is therefore smaller than the raw record count, and its precise value depends on the spatial tolerance used to decide when two nearby records refer to the same station, an unavoidable ambiguity when working with historical coordinates that are themselves inferred.

Spatial tolerance	Unique stations	Reduction from raw
Raw station records	1,984	—
Unique standardised names	1,714	14 %
5 km	1,468	26 %
10 km	1,244	37 %
25 km	754	62 %

Table 2. Number of distinct terrestrial stations in the SEA-HISTIN inventory (v0) under successive deduplication criteria. Spatial tolerances merge records whose inferred coordinates lie within the stated great-circle distance of one another (single-linkage).

To quantify this, we grouped station records by geographic proximity clusters on the great-circle (haversine) distance between inferred coordinates, merging any two records separated by less than a chosen threshold into a single station. Reducing first to unique standardised names lowers the count from 1,984 to 1,714, indicating that roughly 14% stations are a repeat appearance of an already-named station in a different source. Applying a spatial

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

threshold reduces the count further and does so strongly as the threshold widens (Table 2): a 5 km tolerance yields 1,468 distinct stations, a 10 km tolerance 1,244, and a 25 km tolerance 754 (Figure 5c). The inventory therefore documents on the order of 750 to 1,250 physically distinct observing locations, depending on whether a conservative (25 km) or a strict (10 km) matching criterion is adopted.

The sensitivity of the count to the chosen threshold is greatest for the countries with the densest historical networks e.g. Indonesia, Malaysia, and the Philippines, where many stations lie within 10 to 25 km of a neighbouring station and progressively coalesce as the tolerance widens; for Indonesia, for example, the distinct station count falls from 658 at 10 km to 358 at 25 km (Table 3; Figure 5c). By contrast, the sparser mainland networks of Myanmar, Cambodia, and Vietnam are almost unaffected, their stations already lying far enough apart that few merge at either threshold. This behaviour is expected, and it underlines that the deduplicated station count is best reported as a threshold-dependent range rather than a single figure. For the diagnostics presented elsewhere in this section we retain all station–source records, since each represents a distinct catalogued source entry; the deduplicated counts above should be used when the quantity of interest is the number of physical observing locations rather than the number of inventory records.

Country modern	**Raw records**	**Unique names**	**Dist. (5km)**	**Dist. (10km)**	**Dist. (25km)**
Indonesia	922	867	792	658	358
Malaysia	500	329	215	166	68
Philippines	238	229	201	181	120
Vietnam	80	66	67	65	61
Thailand	79	69	68	64	58
Laos	65	62	59	56	40
Singapore	35	32	7	3	2
Myanmar	30	30	30	30	30
Cambodia	20	18	18	16	15
Brunei	8	8	8	5	4
Others	7	6	6	5	5

Table 3. Number of catalogued terrestrial stations by modern country under successive deduplication criteria. Raw records are individual station–source entries in the secondary table; unique standardised names collapse repeated appearances of the same named station across sources; the 5, 10, and 25 km columns merge records whose inferred coordinates lie within spatial clustering and distinct clusters are then tallied per country. Countries are ordered by raw record count; "Others" comprises Australia and one record each in the Federated States of Micronesia, China, the United States.

Furthermore, the relative sparseness of records for the mainland states of Cambodia, Laos, and Myanmar is consistent with the documented history of data rescue in the region and identifies these countries as priorities for future archival work. In v0 all coordinates are flagged as inferred and all location confidence as high; refining these flags against primary sources, and distinguishing genuinely precise from approximate locations, is a planned enhancement for subsequent versions.

4.3 Temporal density

The start and end dates recorded for each series allows the temporal density of the inventory is quantified as the number of active series in each calendar year (Figure 5b). The terrestrial and marine records display markedly different temporal signatures. The marine record begins earliest, with sporadic Royal Navy logbook coverage from the early nineteenth century, rising to a pronounced peak of over 400 concurrently active voyages in the 1860s, after which naval logbook coverage in the catalogued sources declines. The terrestrial record, by contrast, builds steadily from the late nineteenth century, with the number of active station series exceeding 1,200 by 1880 and rising to a maximum of nearly 2,000 around 1945, before falling away through the second half of the twentieth century as the inventory's current source base thins. The two records are therefore strongly complementary: the marine sources extend the region's observational reach back into the early instrumental period, while the terrestrial sources dominate the century from 1880 to 1970.

The temporal profile also carries the imprint of the region's history. The plateau in terrestrial coverage across the early 1940s, coinciding with the Japanese occupation, and the subsequent decline through the decolonisation period, are consistent with the documented disruption and loss of records during and after the Second World War. Care is required in interpreting the exact year-to-year values, since the date fields in this version are recorded at annual resolution in a compact numeric format and a subset of records exhibit start and end dates whose ordering requires reconciliation against the source documents; the density curves shown here are constructed from the full temporal span of each series and are robust to this ordering, but precise sub-annual analysis awaits the date-field standardisation planned for version 1. This temporal-density diagnostic provides a direct, quantitative means of identifying the periods in which targeted data rescue would yield the greatest marginal gain.

4.4 Observation frequency and meteorological variables

The inventory is designed to record observation frequency: hourly, synoptic, daily, and monthly and, for each station, the specific meteorological variables observed (rainfall, temperature, pressure, humidity, sunshine, cloud, wind, and coded weather). These fields are what allow a prospective data-rescue user to target, for example, the daily rainfall series required for flood reconstruction or the sub-daily pressure observations required for reanalysis assimilation. In the present version, however, these two fields are not yet sufficiently differentiated to support the intended diagnostics: temporal resolution is recorded as monthly for all 1,984 station records, and the per-variable flags are populated at or close to 100% for most variables, reflecting default values applied during initial data entry rather than variable-by-variable verification against each source.

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

This is the most important immediate priority for the inventory's development. Once the frequency and variable fields are verified against primary sources and updated in the inventory, the inventory will support the full range of variable-based diagnostics, including maps of rainfall-only versus multivariate stations, and time series of the availability of each meteorological element.

For the marine records, the recorded content is more fully differentiated: of the 10,729 voyages catalogued, 9,675 are noted as containing meteorological observations, 9,810 as containing oceanographic observations, and 9,939 as recording vessel movements, with the Royal Navy (6,189 voyages) and US Navy (669 voyages) the principal identified sources alongside a substantial body of mercantile logbooks. This confirms the marine component as a rich and internally consistent resource for the recovery of both atmospheric and oceanographic variables across the region's seas.

4.5 Applications

By establishing what records exist, where they are held, what they contain, and how ready they are for digitisation, SEA-HISTIN is intended to serve as a planning and coordination instrument for the regional data-rescue community. Its most direct application is the prioritisation of scarce data-rescue resources: the spatial and temporal diagnostics demonstrated here allow effort to be directed toward the geographic and temporal gaps where recovery would most improve the observational record, while the transcription-readiness and data-quality fields allow it to be directed toward the sources where recovery is most feasible. Abundant historical weather observations for Southeast Asia would directly improve the skill of century-scale reanalyses in a region where they are currently poorly constrained, supporting the reconstruction of past monsoon variability, droughts, floods, and tropical cyclones.

The generated inventory is made available as machine-readable interlinked files (Primary, Secondary and Platform-level tables) in CSV/JSON format and maintaining an updateable versioned repository so that additional collections and newly digitized materials can be incorporated without restructuring the database. We have documented methodological protocol to ensure reproducibility, including inclusion rules, metadata definitions, and uncertainty coding. The technical documentation (see Supplementary Information S1) describing complete schema, field definitions, provenance information, unique identifiers and guidance on how to interpret access-status fields is stored in the repository along with the inventory datafiles (version v0) and can be found at Teleti and Williamson (2026).

Historical weather is interesting to scientists and climate-enthusiasts alike. Beyond the science, many climate-enthusiasts are drawn to historical weather data because it connects climate to human history explaining epidemic disease outbreaks, crop failures, and migration patterns that shaped the modern world especially in Southeast Asia. The discovery that a particular monsoon failure contributed to a colonial famine, or that unusual weather influenced

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

a widespread migration, makes the historical weather far beyond specialist meteorology. We hope with the publication of this inventory, searching historical weather records will be more accessible to wider public.

5. Future work

Several lines of development follow directly from this first version. We can fully utilise the powerful relational structure of the inventory to store more granular meta-data from data-sources, stations, ships, and instruments. As shown in the Figure 4, there is much broad scope to accommodate finer details. Similarly, for existing fields, immediate step is the verification and standardisation of the temporal-resolution, observation-frequency, and per-variable fields against primary sources, which will unlock the full variable-based diagnostic capability described above. In parallel, the compact numeric date fields will be converted to a standard calendar format and their internal ordering reconciled, and the coordinate-precision and location-confidence flags will be refined to distinguish precisely located from approximately located stations. Beyond data curation, the inventory will be extended geographically and thematically: continued archival discovery, particularly for the under-represented mainland states and for Dutch and French colonial holdings, will expand its coverage, while its extensible relational structure allows entirely new source categories e.g. early upper-air and aerial observations, additional instrument and observer attribute tables to be incorporated without disruption to the existing schema.

Finding newer sources of historical weather data-sources is only a start to the data-rescue journey leading to standardised, error-corrected extracted weather observations. Digitising these new and previously unknown records for Malaysia and Singapore can be achieved using a combination of citizen science and student participation. This is a commonly established method of digitising large quantities of data, relatively inexpensively and quickly (Teleti et al. 2026 and Lorrey, 2022).

Lastly, building large community asset such as this inventory cannot be accomplished without partnership of regional and international meteorological services. Often these national meteorological services and their archives hold historical weather records. As the aim of the inventory is to collate a regional historical weather data-sources, partnership initiatives are being taken under ACRE leadership with Myanmar, Cambodia, Malaysia and Indonesia. We envision that this inventory will be updated regularly as new sources of historical weather observations come to light through data-rescue community, we intend SEA-HISTIN to grow into an active, community-maintained resource that accelerates the recovery of Southeast Asia's historical climate.

Appendix A

This appendix provides detailed descriptions and historical context for the data sources summarised in Table 1. Sources are grouped into un-transcribed and transcribed categories, reflecting their current state of data-rescue progress. For each source, we describe its provenance, custodial history, observed variables, spatial and temporal coverage, and any notable features relevant to assessing its scientific value and digitisation priority. This information

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

is intended to support data-rescue teams and prospective users in identifying sources of interest and planning future recovery efforts. These sources can be identified by matching against Source ID column in the inventory file; title, authoring institution, date of production, holding archive, call number or catalogue reference, format, and access conditions can be found for all data sources in the inventory file.

A1. Un-transcribed data sources

These sources have been found in the form of paper records, some have been photographed, but are still waiting to be transcribed.

A1.1 Supplements from the Agricultural Bulletins of the Federated Malay States (FMS) (1912-1921)
[Source ID: BF6XuPPJ]

One of the major sources of weather observations in the present collection is the Supplements from the Agricultural Bulletins of the Federated Malay States (FMS) (1912-1921). The Agricultural Bulletins of the FMS (four British-protected states on the Malay Peninsula operating under a single centralized colonial administration from 1896 until 1946) contain technical articles and meteorological data relevant to tropical agriculture and pest management. The FMS was comprised of Perak, Selangor, Negri Sembilan and Pahang but these bulletins also note observations made within what was known as the Unfederated Malay States, including Kelantan, as well as the Straits Settlements (Singapore, Malacca, Penang, Province Wellesley and parts of Perak). The weather data variables include air temperature-related measurements, such as mean dry bulb and wet bulb temperature, maximum air temperature, minimum air temperature, and range. Humidity and precipitation variables include dew point, total rainfall and maximum daily rainfall. These bulletins provided detailed climatic data for agricultural planning in the tropical environment.

A1.2 Malayan Agricultural Journal (1922 - 1963)
[Source ID: BF6XuPPJ]

The Malayan Agricultural Journal was published (v17n1-12) by the Ministry of Agriculture for Malaya, continuing briefly after independence from Britain in 1957. The journal contains the latest research articles on crop cultivation, disease and pest management and other topics related to agriculture. It also contains monthly meteorological summaries e.g. min, max air temperature observations, humidity and rainfall totals. Several stations across multiple states and districts in the FMS, UnFMS and Straits Settlements were represented, for example, the northerly UnFMS of Kedah and the datasets also included observations made in towns, as well as in agricultural areas.

A1.3 Gardens' Bulletin Straits Settlements (1912 – 1939)
[Source ID: Bpx2g9dL]

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

The Gardens' Bulletin, published by the Singapore Botanic Gardens, includes scientific notes, botanical observations, entomological reports, and horticultural updates relevant to Malaysia in the early 20th century. Prepared by the staff of the Singapore Botanic Gardens, the bulletin documents ongoing research into economically important crops, destructive insect pests, plant diseases, and field observations from botanical expeditions across the Malay Peninsula. The bulletin also published daily and monthly rainfall data from Singapore (observed at the Botanic Gardens, Singapore) and Penang (Waterfall Gardens).

The current collection holds many volumes of the bulletins, though not continuous, from 1912 to 1932. Crucially, the locations of the observing sites were not moved throughout the observation period. This is important for studying long-term variability of rainfall in Singapore and Penang.

A1.4 Federation of Malaya Irrigation Department' Rainfall Records
[Source ID: B2iRvJNt]

The compilation of rainfall data as Hydrological Data: Rainfall Records, 1879–1958, was published by the Drainage and Irrigation Department, Federation of Malaya. It is an official government compilation presenting nearly 80 years of rainfall observations collected across the Federation of Malaya. It represents one of the most comprehensive early efforts to consolidate and standardise rainfall observations for the Federation of Malaya spanning colonial and post-colonial periods. The data were collected by various departments, agencies, to produce this document, all records were brought together, quality controlled and statistical analysis was prepared by the Drainage and Irrigation Department. This volume brings together nearly eight decades of meteorological data collected from stations distributed across the peninsula, records are organised by state and region, including: Johore, Kedah, Kelantan, Malacca, Negri Sembilan, Pahang, Penang Island, Perak, Perlis, Province Wellesley, Selangor and Trengganu.

The document also contains important meta-data required to make adjustments and detect biases in the recorded data. For example, it states that some earlier observations were made on locally manufactured 8-inch diameter, whereas post-war standardisation introduced 5-inch diameter gauges and tipping-bucket or float-type rainfall recorders. Attention was given to gauge exposure, siting height, protection from wind and obstructions, calibration and inspection routines to standardise the observations. Alongside observations, graphs, maps and summaries of recorded observations is also provided for quick summary.

A1.5 Rainfall Statistics of the British Borneo Territories
[Source ID: Bzy1MqaW]

The publication *Rainfall Statistics of the British Borneo Territories (Sarawak – Brunei – North Borneo),* 1896–1957 constitutes a landmark compilation of historical rainfall data for the island of Borneo during the late

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

nineteenth and early twentieth centuries. Prepared by the Department of Civil Aviation and Meteorological Services, the volume brings together records from a wide network of rainfall stations distributed across Sarawak, North Borneo, and Brunei. The volume covers the entirety of British Borneo, with rainfall stations distributed across coastal areas, river basins, interior and upland regions, including Kuching (capital of Sarawak), Sibu, Miri (Lutong), Bintulu, and Mukah.

A notable feature of the publication is its inclusive use of observations from both government-run stations and those maintained by private estates and commercial enterprises. These contributions greatly expanded geographical coverage, particularly in remote and interior regions, and significantly enhanced the climatological understanding of the territories. Mean monthly and calculated yearly totals are given and spatial distribution of monthly rainfall on the Island is also shown.

A1.6 Meteorologische en Geophysische Dienst; Koninklijk Magnetisch en Meteorologisch Observatorium (Batavia): Regenwaarnemingen in Nederlandsch-Indië; Batavia 1900-1939 (Meteorological and Geophysical Service; Royal Magnetic and Meteorological Observatory (Batavia): Rain observations in the Dutch East Indies)

[Source ID: BA5EJRYJ]

This official issue of the Meteorological Observatory in Batavia contains weather observations from Batavia (Jakarta) but also the Dutch East Indies more widely. Data comprises monthly and annual rainfall totals, with latitude, longitude, elevation, and distance from the coast, number of rainy days and maximum rainfall in 24 hours for more than two hundred stations distributed throughout Java, Madura, and what were then called the Outer Possessions. The compilation was a systematic effort by the Royal Magnetic and Meteorological Observatory in Batavia to document and analyse rainfall patterns across one of the world's most climatically diverse regions. Beyond the tabulated rainfall totals, the publication includes a crucial descriptive component: a concise gazetteer titled *Korte Overzicht van de Ligging der Stations* ("Short overview of the location of the stations").

This meta-data provides geographic and administrative context for each rain-gauge site, noting whether a station was situated at a residency headquarters, port, railway station, military post, or commercial installation, and frequently recording elevation above sea level and distance from the coast. Such metadata enables meaningful comparison of rainfall across contrasting environments—from low-lying coastal settlements and small offshore islands to inland rail corridors and mountain foothills. Understanding the local environment and topography, coastal influence enables users to adjust and offset biases if necessary. The publication combines meticulous station-level rainfall measurements with regional statistical analyses extending back to 1879 through to 1941. The data series is also replicated in the Verhandelingen, Koninklijk Magnetisch en Meteorologisch Observatorium te Batavia [Proceedings of the Royal Magnetic and Meteorological Observatory in Batavia].

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

A1.7 Thailand's Monthly Meteorological Bulletin

[Source ID: BsadwKor]

The Monthly Meteorological Bulletin, issued by the Section of Meteorology within the Hydrographic Service, the bulletin documents weather conditions using standardized instruments and internationally recognized methods at the Central Meteorological Station in Bangkok, supplemented by observations from selected stations across the country. The publication presents hourly, daily, and monthly summaries of atmospheric pressure, temperature, wind, rainfall, and cloud conditions, together with graphical analyses and detailed rainfall reports.

The Central Meteorological Station of the Meteorological Section, Hydrographic Service is located within the Grand Palace enclosure, near the Naval Cadet School.
Five daily observations are made at 0700, 1000, 1300, 1600, and 1900 Siam Standard Time; additional hourly data are obtained from automatic instruments. Air pressure is corrected for temperature, latitude, and mean sea level. Wind direction and speed were recorded hourly using an automatic Dines pressure-tube anemograph. All thermometers were housed in a Stevenson screen.

A1.8 ASEAN compendium

[Source ID: BhmD2u5C]

The ASEAN Compendium on Climate Statistics represents a significant regional initiative aimed at aggregating, consolidating, and disseminating climate-related statistical information across the Association of Southeast Asian Nations (ASEAN). Developed under the coordination of the ASEAN Secretariat, in collaboration with national statistical offices and meteorological and environmental agencies. While the Compendium is not intended to replace detailed climatological datasets, it plays a complementary role by translating complex climate information into statistically robust indicators.

As the data is principally contributed by National Statistical Offices (NSOs) and National Meteorological and Hydrological Services (NMHSs) of ASEAN Member States, the compendium offers rare historical weather observations, not present in international climate datasets. More importantly, it enhances reproducibility and traceability for the methods and data, as they are tied to official national data sources and documented methodologies. Typically, the compendium published in 1976 contains records of monthly and annual rainfall, mean and extreme rainfall, mean and extreme temperature, humidity, dew point, sunshine, cloudiness, pressure, surface and upper-air winds from more 900 stations in five countries namely Indonesia, Malaysia, Philippines, Singapore and Thailand. The stations recorded in the compendium naturally have different lengths of records and are detailed in different tables having less than 20 years in the 1951-1975 period.

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

A1.9 JAMSTEC meteorological observations

[Source ID: BIPznyqD]

Meteorological observations collected in Southeast Asia during the early 1940s constitute an important but still under-utilised source of the global historical climate record. Among these, observations associated with Japanese scientific, naval, and colonial activities in Asia and Oceania. The period 1940–1941 lies at a critical juncture: immediately prior to, and during the early stages of, the Pacific War, when observational coverage in much of Southeast Asia was unusually dense but later became fragmented, or lost making them inaccessible to international climate databases.

These records are now archived and made available by the Japan Agency for Marine-Earth Science and Technology (JAMSTEC). The data source contains station metadata inventory, listing more than 100 stations across Southeast Asia and nearby regions, exact coordinates and elevations. Representing more than eleven modern-day countries including Philippines, Malaysia, Singapore, Thailand, Indonesia, Vietnam, Laos, Cambodia, Papua New Guinea (New Guinea) and others. The document contains a reference sheet (in Japanese) for an international meteorological reporting code. It lays out the variable sequence used in a synoptic observation group and provides lookup tables for each code (e.g., cloud types, visibility, temperature, pressure, wind, past weather, etc.).

A1.10 UK Met Office climatological returns

[Source ID: BGoL8zQx]

The UK Met Office (UKMO) climatological returns for Singapore, Malacca, Penang and other locations in Malayan Peninsula are stored in UKMO Archives in Exeter, UK. These forms (3211 and 3212) were part of standard UKMO weather observations aggregation. The Meteorological Observations constitutes a detailed monthly and annual record of weather conditions observed at Singapore and selected Malayan stations from 1925 to 1964 and for Christmas Island, which was a British Crown Colony (1900-1958) from 1901 to 1911. The records contain both monthly summaries and annual totals for atmospheric pressure (mean, maxima, minima), air temperature (daily maxima, minima, mean temperature, absolute extremes), humidity, rainfall (monthly totals, departures from normal, extreme daily falls), cloud amount, wind, weather phenomena and sunshine duration. All temperatures are expressed in degrees Fahrenheit, rainfall primarily in millimetres, and pressure in millibars, with instrument heights and exposure carefully specified.

A1.11 Royal Navy Ship Logbooks

[Source ID: Blc5veHJ]

Throughout the 19th century, the British Royal Navy maintained meticulous logbooks as part of its operational requirements. These logbooks often contained detailed meteorological observations, including air and sea

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

temperatures, wind force (speed) and direction (frequently using the Beaufort wind scale), barometric pressure, and descriptive weather conditions. Logbooks archived at the UK Hydrographic Office (UKHO) and imaged under ACRE project, and have been prioritised for digitisation in this project, especially those from the early-industrial period, in order to fill gaps for the 1860s and 1870s from the Atlantic, Indian, and Pacific oceans.

Hundreds of these British Royal Navy ships would drop anchor at ports including George Town, Malacca and to replenish supplies and for essential repairs on their way to and from Britain and Asia in the 18th and 19th centuries, making the Malacca Straits one of the world's busiest shipping lanes. Some of these ships would anchor at trading docks or a few miles out in the sea for several days or weeks. While anchored, they would continue taking weather observations. To take advantage of this high-volume of shipping traffic, near-shore weather observations can cover gaps in adjacent terrestrial observations.

A1.12 National Archives and Records Administration (NARA)
[Source ID: BHBJwDTc]

The National Archives and Records Administration (NARA) have made available several hundreds of thousands of logbook pages of US Navy ships that operated in Southeast Asian waters from the 1940s to the 1960s. For this project our focus has been only the ships travelling to, or transiting through, Malacca Straits, Singapore, containing meteorological observations. Of a potential 693 ships, at least 504 have been identified as containing relevant data, including wind, pressure, air temperature (dry and wet bulb) and clouds.

A2 Transcribed data sources

Many of the data sources have been digitised over several years by the authors, volunteers and paid staff across multiple projects, including the international Atmospheric Circulations over the Earth (ACRE) initiative. The quality and quantity of weather observations contained in these sources are varied. So are the weather variables recorded, frequency and duration. Though the observations are digitised they still require thorough standardisation, quality checks and formatting if they are to be assimilated into existing climate datasets or to be used for climate studies.

Weather observations recorded at terrestrial stations in Malaysia and Singapore for the 19th and early 20th centuries can be erratic and piecemeal in their survival. Often, the location and state in which they are stored has also not been standardised, especially for the pre-1940s period. A sample list is below. For further information on smaller datasets and for historical context for the 19th century, see Williamson (2021).

A2.1 Singapore Magnetic Observatory Yearbooks, 1841-45
[Source ID: BJmzGNHd]

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

Ostensibly a magnetic observatory set up under the authority of the British Society for the Advancement of Science and managed by the East India Company, the observatory made by default observations of the temperature at the site on the Kallang River on Singapore Island. These observations were published contemporaneously in Madras and were digitised posthumously by the British Geological Survey.

A2.2 Straits Settlements Government Gazette
[Source ID: BkUSukC9]

One of the most extensive sources of information is the Straits Settlements' Government Gazette. This was published monthly from 1859 to 1940. Many editions contain annual, monthly or daily weather observations for different stations, although there is no station or temporal continuity. The gazette covered information pertinent to the running of the colony, from judicial notices, official appointments, major events and collated annual reports from different government departments.

It is under the reports of the various official iterations of the medical services, agricultural services, harbour authorities and other departments that meteorological summaries can be found. Data – including rainfall, pressure and temperature - has been digitised for the years 1864-67, 1869-1874, 1884-1915 mainly for Singapore but also for various stations in Penang, Province Wellesley and Malacca. Data for Singapore contains more daily observations over longer temporal spans. Digitised data includes that taken from Raffles and Horsburgh lighthouses in the Singapore Strait.

It is from the Government Gazettes that we also have information from Christmas Island, which was administered under the Straits Settlements government from 1900 to 1958, with a dataset starting from 1901. Recovered and digitised data covers the years 1901 to 1952 and includes variables for humidity, air temperature, cloud, rainfall, and wind.

A2.3 Private weather diaries/newspapers
[Source ID: Ba4Oi9BF]

Several weather diaries survive, usually due to publication in contemporary newspapers and pseudo-scientific journals. A newspaper correspondents' private weather diary was published in The Singapore Free Press and Mercantile Advertiser, 1837, for example. The observations record sub-daily temperatures at Sambas, Borneo during the months of June and July 1837. During the 1860s, there was renewed interest in weather observations by plantation owners, and two important datasets from this period survive and have been digitised: the Killiney Estate from 1863-65 and Mount Pleasant Estate from 1864 to 1869, both in Singapore.

A2.4 Kuching Museum observations

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

[Source ID: BxfefqEH]

A meteorological station was established at Kuching Museum, Sarawak, East Malaysia in 1891. Sarawak operated semi-autonomously from the authorities in the Straits Settlements until being ceded in 1946 as a British Crown Colony. Several daily weather observations were recorded including temperature, dry/wet-bulb temperatures, air pressure, rainfall and Sunshine hours. The current data starts in 1901 ending in 1954, though is not continuously recorded. Digitised data covers 1906 to 1927, and 1954.

A2.5 Others

[Source ID: B0o8n6PW, BJmzGNHd, BaxUTc6B, BrI9qpW1, Bdgn4d6j, BQEv9l4Q, BatVlDq2, Bykcnvsf, Be4iThCB, BnXXNSvA, B91SHnL1 ,BKwyDJxv, Ba6GpT6V, Bu8SqEgQ, BI2Pn4UX]

A significant amount of data has been recovered and digitised for Singapore and for the FMS for the early to mid-20th century. This includes records from the Straits Settlements' Medical Department from the 1880s to 1917 (various stations including in Singapore, Penang and Province Wellesley); the Singapore Botanic Gardens from 1916 to 1924 and data from the McRitchie Reservoir (formerly Impounding Reservoir) from 1875 to 1975. The hospitals however provide the most comprehensive datasets for all settlements throughout this period.

Data availability

The SEA HISTIN inventory (version v0) and technical documentation (Supplementary information S1) are available to view and download from the Zenodo repository at the DOI https://doi.org/10.5281/zenodo.20663521 (Teleti and Williamson, 2026). The python code and scripts used to generate the figures in this paper are available at Teleti (2026).

Author contributions

PT and FW designed the study. FW collected the data. PT developed the code and performed the analysis. Both authors contributed equally to writing the manuscript.

Competing interests

We declare no conflicts of interest.

Acknowledgements

None

Financial support

This research is supported by the Ministry of Education, Singapore, under MOET2EP40123-0021, Building a Deeper Long-Term Climate Understanding of Rainfall Trends in Singapore, West Malaysia, and Northern Sumatra.

References

Allan, R., Brohan, P., Compo, G. P., Stone, R., Luterbacher, J., and Brönnimann, S.: The International Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative, B. Am. Meteorol. Soc., 92, 1421–1425, https://doi.org/10.1175/2011BAMS3218.1, 2011.

Bartholomew, J. G.: A Literary and Historical Atlas of Asia, Everyman's Library No. 633, J. M. Dent and Sons, London, available at: https://www.gutenberg.org/ebooks/40900 (last access: 22 July 2026), 1912.

Biodiversity Heritage Library: available at: https://www.biodiversitylibrary.org/ (last access: 22 July 2026)

Britannica (Encyclopaedia): Southeast Asia, available at: https://www.britannica.com/place/Southeast-Asia (last access: 22 July 2026), 2026.

Brönnimann, S., Allan, R., Ashcroft, L., Baer, S., Barriendos, M., Brázdil, R., Brugnara, Y., Brunet, M., Brunetti, M., Chimani, B., Cornes, R., Domínguez-Castro, F., Filipiak, J., Founda, D., García Herrera, R., Gergis, J., Grab, S., Hannak, L., Huhtamaa, H., Jacobsen, K. S., Jones, P., Jourdain, S., Kiss, A., Lin, K. E., Lorrey, A., Lundstad, E., Luterbacher, J., Mauelshagen, F., Maugeri, M., Maughan, N., Moberg, A., Neukom, R., Nicholson, S., Noone, S., Nordli, Ø., Ólafsdóttir, K. B., Pearce, P. R., Pfister, L., Pribyl, K., Przybylak, R., Pudmenzky, C., Rasol, D., Reichenbach, D., Řezníčková, L., Rodrigo, F. S., Rohr, C., Skrynyk, O., Slonosky, V., Thorne, P., Valente, M. A., Vaquero, J. M., Westcott, N. E., Williamson, F., and Wyszyński, P.: Unlocking pre-1850 instrumental meteorological records: A global inventory, B. Am. Meteorol. Soc., 100, ES389–ES413, https://doi.org/10.1175/BAMS-D-19-0040.1, 2019.

Brugnara, Y., et al.: The EUSTACE global land station daily air temperature dataset, Geosci. Data J., 6, 235–251, https://doi.org/10.1002/gdj3.81, 2019.

Brunet, M. and Jones, P.: Data rescue initiatives: bringing historical climate data into the 21st century, Clim. Res., 47, 29–40, https://doi.org/10.3354/cr00960, 2011.

Buckley, B. M., Fletcher, R., Wang, S.-Y. S., Zottoli, B., and Pottier, C.: Monsoon extremes and society over the past millennium on mainland Southeast Asia, Quaternary Sci. Rev., 95, 1–19, https://doi.org/10.1016/j.quascirev.2014.04.022, 2014.

Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Matsui, N., Allan, R. J., Yin, X., Gleason, B. E., Vose, R. S., Rutledge, G., Bessemoulin, P., Brönnimann, S., Brunet, M., Crouthamel, R. I., Grant, A. N., Groisman, P. Y., Jones, P. D., Kruk, M. C., Kruger, A. C., Marshall, G. J., Maugeri, M., Mok, H. Y., Nordli, Ø., Ross, T. F., Trigo, R. M., Wang, X. L., Woodruff, S. D., and Worley, S. J.: The Twentieth Century Reanalysis Project, Q. J. Roy. Meteor. Soc., 137, 1–28, https://doi.org/10.1002/qj.776, 2011.

Copernicus Climate Change Service (C3S): Data Rescue Service, available at: https://datarescue.climate.copernicus.eu/ (last access: 22 July 2026).

Cram, T. A., Compo, G. P., Yin, X., Allan, R. J., McColl, C., Vose, R. S., Whitaker, J. S., Matsui, N., Ashcroft, L., Auchmann, R., Bessemoulin, P., Brandsma, T., Brohan, P., Brunet, M., Comeaux, J., Crouthamel, R., Gleason, B. E., Groisman, P. Y., Hersbach, H., Jones, P. D., Jónsson, T., Jourdain, S., Kelly, G., Knapp, K. R., Kruger, A., Kubota, H., Lentini, G., Lorrey, A., Lott, N., Lubker, S. J., Luterbacher, J., Marshall, G. J., Maugeri, M., Mock, C. J., Mok, H. Y., Nordli, Ø., Rodwell, M. J., Ross, T. F., Schuster, D., Srnec, L., Valente, M. A., Vizi, Z., Wang, X. L., Westcott, N., Woollen, J. S., and Worley, S. J.: The International Surface Pressure Databank version 2, Geosci. Data J., 2, 31–46, https://doi.org/10.1002/gdj3.25, 2015.

Encyclopaedia Britannica: Southeast Asia, available at: https://www.britannica.com/place/Southeast-Asia (last access: 22 July 2026), 2026.

Gao, E., Timbal, B., and Williamson, F.: Creating Singapore's longest monthly rainfall record from 1839 to the present, MSS Research Letters, 2, 3–11, 2018.

GeoNames: GeoNames Geographical Database, available at: https://www.geonames.org/ (last access: 22 July 2026), 2026.

Getty Research Institute: Getty Thesaurus of Geographic Names (TGN) Online, J. Paul Getty Trust, Los Angeles, available at: https://www.getty.edu/research/tools/vocabularies/tgn/index.html (last access: 22 July 2026), 2026.

García-Herrera, R., Können, G. P., Wheeler, D. A., Prieto, M. R., Jones, P. D., and Koek, F. B.: CLIWOC: a climatological database for the world's oceans 1750–1854, Climatic Change, 73, 1–12, https://doi.org/10.1007/s10584-005-6952-6, 2005.

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

HathiTrust: HathiTrust Digital Library, HathiTrust, available at: https://www.hathitrust.org/ (last access: 22 July 2026).

Hawkins, E., Brohan, P., Burgess, S. N., Burt, S., Compo, G. P., Gray, S. L., Haigh, I. D., Hersbach, H., Kuijjer, K., Martínez-Alvarado, O., McColl, C., Schurer, A. P., Slivinski, L., and Williams, J.: Rescuing historical weather observations improves quantification of severe windstorm risks, Nat. Hazards Earth Syst. Sci., 23, 1465–1482, https://doi.org/10.5194/nhess-23-1465-2023, 2023.

Laloyaux, P., de Boisseson, E., Balmaseda, M., Bidlot, J.-R., Brönnimann, S., Buizza, R., Dahlgren, P., Dee, D., Haimberger, L., Hersbach, H., Kosaka, Y., Martin, M., Poli, P., Rayner, N., Rustemeier, E., and Schepers, D.: CERA-20C: A coupled reanalysis of the twentieth century, J. Adv. Model. Earth Sy., 10, 1172–1195, https://doi.org/10.1029/2018MS001273, 2018.

Lassa, J. A.: Disaster policy change in Indonesia 1930–2010: from government to governance?, Int. J. Mass Emerg. Disasters, 31, 130–159, https://doi.org/10.1177/028072701303100202, 2013.

Lassa, J. A., Lai, A. Y. H., and Goh, T.: Climate extremes: an observation and projection of its impacts on food production in ASEAN, Nat. Hazards, 84, 19–33, https://doi.org/10.1007/s11069-015-2081-3, 2016.

Leiden University Libraries: Digital Collections, Leiden University, Leiden, the Netherlands, available at: https://digitalcollections.universiteitleiden.nl/ (last access: 22 July 2026).

Lorrey, A. M., et al.: Meteorological data rescue: citizen science lessons learned from Southern Weather Discovery, Patterns, 3, 100495, 2022.

Nash, D. J. and Hannaford, M.: Historical climatology in Africa: a state of the art, Past Global Changes Magazine, 28, 42–43, https://doi.org/10.22498/pages.28.2.42, 2020.

Nash, D. J., Adamson, G. C. D., Ashcroft, L., Bauch, M., Camenisch, C., Degroot, D., Gergis, J., Jusopović, A., Labbé, T., Lin, K.-H. E., Nicholson, S. D., Pei, Q., del Rosario Prieto, M., Rack, U., Rojas, F., and White, S.: Climate indices in historical climate reconstructions: a global state of the art, Clim. Past, 17, 1273–1314, https://doi.org/10.5194/cp-17-1273-2021, 2021.

NOAA Central Library: National Oceanic and Atmospheric Administration Central Library, available at: https://library.noaa.gov/ (last access: 22 July 2026).

National Library Board: National Library Board Singapore, Singapore, available at: https://www.nlb.gov.sg/ (last access: 22 July 2026).

NUS Libraries: NUS Libraries, National University of Singapore, Singapore, available at: https://nus.edu.sg/nuslibraries (last access: 22 July 2026).

PAGASA: About Us — History, Philippine Atmospheric, Geophysical and Astronomical Services Administration, available at: https://www.pagasa.dost.gov.ph/about-us (last access: 22 July 2026).

Pfister, L., Hupfer, F., Brugnara, Y., Munz, L., Villiger, L., Meyer, L., Schwander, M., Isotta, F. A., Rohr, C., and Brönnimann, S.: Early instrumental meteorological measurements in Switzerland, Clim. Past, 15, 1345–1361, https://doi.org/10.5194/cp-15-1345-2019, 2019.

Ren, G., Chan, J. C. L., Kubota, H., et al.: Historical and recent change in extreme climate over East Asia, Climatic Change, 168, 22, https://doi.org/10.1007/s10584-021-03227-5, 2021.

Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., Allan, R., Yin, X., Vose, R., Titchner, H., Kennedy, J., Spencer, L. J., Ashcroft, L., Brönnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Crouthamel, R., Domínguez-Castro, F., Freeman, J. E., Gergis, J., Hawkins, E., Jones, P. D., Jourdain, S., Kaplan, A., Kubota, H., Le Blancq, F., Lee, T.-C., Lorrey, A., Luterbacher, J., Maugeri, M., Mock, C. J., Moore, G. W. K., Przybylak, R., Pudmenzky, C., Reason, C., Slonosky, V. C., Smith, C. A., Tinz, B., Trewin, B., Valente, M. A., Wang, X. L., Wilkinson, C., Wood, K., and Wyszyński, P.: Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system, Q. J. Roy. Meteor. Soc., 145, 2876–2908, https://doi.org/10.1002/qj.3598, 2019.

Slivinski, L. C., Compo, G. P., Sardeshmukh, P. D., Whitaker, J. S., McColl, C., Allan, R. J., Brohan, P., Yin, X., Smith, C. A., Spencer, L. J., Vose, R. S., Rohrer, M., Conroy, R. P., Schuster, D. C., Kennedy, J. J., Ashcroft, L., Brönnimann, S., Brunet, M., Camuffo, D., Cornes, R., Cram, T. A., Domínguez-Castro, F., Freeman, J. E., Gergis, J., Hawkins, E., Jones, P. D., Kubota, H., Lee, T. C., Lorrey, A. M., Luterbacher, J., Mock, C. J., Przybylak, R. K., Pudmenzky, C., Slonosky, V. C., Tinz, B., Trewin, B., Wang, X. L., Wilkinson, C., Wood, K., and Wyszyński, P.: An evaluation of the performance of the Twentieth Century Reanalysis version 3, J. Climate, 34, 1417–1438, https://doi.org/10.1175/JCLI-D-20-0505.1, 2021.

Su, F., Gu, M., and Zhai, Y.: Weathering the storm: impacts of extreme climate events on food security in Southeast Asia and pathways to resilience, Food Secur., https://doi.org/10.1007/s12571-026-01670-9, 2026.

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

Teleti, P., Hawkins, E., and Wilkinson, C.: Weather Rescue at Sea: recovering historical weather observations from 19th century British naval ships, Geosci. Data J., 13, e70056, https://doi.org/10.1002/gdj3.70056, 2026.

Teleti, P. and Williamson, F.: SEA-HISTIN: an inventory of historical weather observations of Southeast Asia, Zenodo [data set], https://doi.org/10.5281/zenodo.20663521, 2026.

Teleti, P.: SEA-HISTIN: an inventory of historical weather observations of Southeast Asia [code], GitHub, available at: https://github.com/prt31/SEA-HISTIN (last access: 22 July 2026), 2026.

Thomas, F., Ngo-Duc, T., and Nguyen-Ngoc-Minh, T.: The Climate Archives of Indochina: insights for understanding climate change in Vietnam and Southeast Asia, AFD Research Papers No. 394, Agence Française de Développement, Paris, France, 44 pp., 2025.

Tsukahara, T., et al.: Archive research into Dutch naval log books, and preliminary open access of its digitisation, Itinerario, 49, 1–13, https://doi.org/10.1017/S0165115325100247, 2025.

van den Besselaar, E. J. M., Klein Tank, A. M. G., van der Schrier, G., Abass, M. S., Baddour, O., van Engelen, A. F. G., Freire, A., Hechler, P., Laksono, B. I., Jones, R. J. B., Kruk, M. C., Nordli, Ø., Zhang, F., Adamovic, M., Gómez-Delgado, F., Klok, E. J., Lekamge, M. L. D., Maskamu, S. M., and Nieto, J. J.: International Climate Assessment & Dataset: Climate Services Across Borders, B. Am. Meteorol. Soc., 96, 16–21, https://doi.org/10.1175/BAMS-D-13-00249.1, 2015.

van den Besselaar, E. J. M., van der Schrier, G., Cornes, R. C., Iqbal, A. S., and Klein Tank, A. M. G.: SA-OBS: A daily gridded surface temperature and precipitation dataset for Southeast Asia, J. Climate, 30, 5151–5165, https://doi.org/10.1175/JCLI-D-16-0575.1, 2017.

Wilkinson, C., Woodruff, S. D., Brohan, P., Claesson, S., Freeman, E., Koek, F., Lubker, S. J., Marzin, C., and Wheeler, D.: Recovery of logbooks and international marine data: the RECLAIM project, Int. J. Climatol., 31, 968–979, https://doi.org/10.1002/joc.2102, 2011.

Wilkinson, C., Brönnimann, S., Jourdain, S., Roucaute, E., Crouthamel, R., Brohan, P., Valente, A., Brugnara, Y., Brunet, M., Compo, G. P., et al.: Best Practice Guidelines for Climate Data Rescue, v1, Copernicus Climate Change Service Data Rescue Service, ECMWF, Reading, UK, 2019.

Wilkinson, C. and Freeman, E.: Best Practice Guidelines for Keying Data from Historic Marine Documents, Copernicus Climate Change Service Data Rescue Service, ECMWF, Reading, UK, 2021.

https://doi.org/10.5194/egusphere-2026-3459
Preprint. Discussion started: 24 July 2026

Williamson, F., Allan, R., Switzer, A. D., Chan, J. C. L., Wasson, R. J., D'Arrigo, R., and Gartner, R.: New directions in hydro-climatic histories: observational data recovery, proxy records and the atmospheric circulation reconstructions over the earth (ACRE) initiative in Southeast Asia, Geosci. Lett., 2, 2, https://doi.org/10.1186/s40562-015-0018-z, 2015.

Williamson, F., Allan, R., Ren, G., Yamada, S., Zaiki, M., Tsukahara, T., and Wilkinson, C.: Collating historic weather observations for the East Asian region: challenges, solutions, and reanalyses, Adv. Atmos. Sci., 35, 899–904, https://doi.org/10.1007/s00376-017-7259-z, 2018.

Williamson, F., Switzer, A., and Wasson, R.: Building a long-time series for weather and extreme weather in the Straits Settlements: a multi-disciplinary approach to the archives of societies, Clim. Past, 17, 791–803, https://doi.org/10.5194/cp-17-791-2021, 2021.

Wycherley, P. R.: Rainfall Probability Tables for Malaysia, Planting Manual No. 12, Rubber Research Institute of Malaya, Kuala Lumpur, 85 pp., 1967.