The International Surface Pressure Databank version 2


1National Center for Atmospheric Research, Boulder, CO, USA
2University of Colorado, CIRES, Boulder, CO, USA
3NOAA Earth System Research Laboratory, Physical Sciences Division, Boulder, CO, USA
4ERT, Inc., Asheville, NC, USA
5Atmospheric Circulation Reconstructions over the Earth Initiative, Met Office Hadley Centre, Exeter, UK
6Met Office Hadley Centre, Exeter, UK
7NOAA National Centers for Environmental Information, Asheville, NC, USA
8School of Earth Sciences, University of Melbourne, Melbourne, Vic., Australia
9Institute of Geography and Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland
10Météo-France, Toulouse, France
11Royal Netherlands Meteorological Institute (KNMI), De Bilt, The Netherlands
12Centre for Climate Change, Universitat Rovira i Virgili, Tarragona, Spain
13International Environmental Data Rescue Organization (IEDRO), Deale, MD, USA
14University Corporation for Atmospheric Research, Boulder, CO, USA
15P. P. Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia
16European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, UK
17Climatic Research Unit, University of East Anglia, Norwich, UK
18Center of Excellence for Climate Change Research/Department of Meteorology, King Abdulaziz University, Jeddah, Saudi Arabia
19Icelandic Met Office, Reykjavik, Iceland
20Météo-France, Direction de la Climatologie, Toulouse, France
21South African Weather Service, Pretoria, South Africa
22Department of Coupled Ocean–Atmosphere–Land Processes Research, Japan Agency for Marine–Earth Science and Technology, Yokosuka, Japan
23Dipartimento di Fisica, Università degli Studi di Milano, Milan, Italy
24National Institute of Water and Atmospheric Research (NIWA), Auckland, New Zealand
25Department of Geography, Climatology, Climate Dynamics and Climate Change, Justus Liebig University Gießen, Gießen, Germany
26British Antarctic Survey, Natural Environment Research Council, Cambridge, UK
27Department of Geography, University of South Carolina, Columbia, SC, USA
28Hong Kong Observatory, Hong Kong, China
29Norwegian Meteorological Institute, Oslo, Norway
30Meteological and Hydrological Service, Zagreb, Croatia
31Instituto Dom Luiz, Laboratório Associado, Universidade de Lisboa, Lisbon, Portugal
32Nicolaus Copernicus University, Torun, Poland
33Climate Research Division, Science and Technology Branch, Environment Canada, Toronto, ON, Canada
34Midwestern Regional Climate Center, Illinois State Water Survey, University of Illinois at Urbana-Champaign, Urbana, IL, USA
35I.M. Systems Group, Inc., Rockville, MD, USA
36NOAA National Centers for Environmental Prediction, Environmental Modeling Center, Camp Springs, MD, USA

*Correspondence: Thomas Cram, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000, USA, E-mail: tcram@ucar.edu

© 2015 The Authors. Geoscience Data Journal published by Royal Meteorological Society and John Wiley & Sons Ltd. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
The International Surface Pressure Databank (ISPD) is the world’s largest collection of global surface and sea-level pressure observations. It was developed by extracting observations from established international archives, through international cooperation with data recovery facilitated by the Atmospheric Circulation Reconstructions over the Earth (ACRE) initiative, and directly by contributing universities, organizations, and countries. The dataset period is currently 1768–2012 and consists of three data components: observations from land stations, marine observing systems, and tropical cyclone best track pressure reports. Version 2 of the ISPD (ISPDv2) was created to be observational input for the Twentieth Century Reanalysis Project (20CR) and contains the quality control and assimilation feedback metadata from the 20CR. Since then, it has been used for various general climate and weather studies, and an updated version 3 (ISPDv3) has been used in the ERA-20C reanalysis in connection with the European Reanalysis of Global Climate Observations project (ERA-CLIM). The focus of this paper is on the ISPDv2 and the inclusion of the 20CR feedback metadata. The Research Data Archive at the National Center for Atmospheric Research provides data collection and access for the ISPDv2, and will provide access to future versions.

Received: 15 May 2014, revised: 3 June 2015, accepted: 8 June 2015

Key words: surface pressure, sea level pressure, observations, stations, ships

Dataset

Identifier: doi:http://dx.doi.org/10.5065/D6SQ8XDW
Title: International Surface Pressure Databank (ISPDv2)
Publisher: Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory, Boulder, Colorado, USA
Publication year: 2010
Resource type: Dataset
Version: 2.0

Introduction

The International Surface Pressure Databank (ISPD) is the world’s largest collection of global surface and sea-level pressure observations. Since its inception in 2002, its development has been facilitated by cooperative efforts between the Global Climate Observing System (GCOS)/World Climate Research Program (WCRP) Working Group on Observational Data Sets for Reanalysis [2007 to 2011], and the ongoing efforts of the GCOS Working Group on Surface Pressure and the international Atmospheric Circulation Reconstructions over the Earth initiative (ACRE, Allan et al., 2011). The full observational record is extracted and assembled from established international archives and newly available collections from more than 60 different contributing organizations, shown in Tables 1 and 2.

The ISPD version 2 (ISPDv2) covers the period 1768–2012. It merges data from three input components: observations from land stations, marine observing systems, and tropical cyclone best track pressure reports. The land station component was extracted from many national and international collections of sea-level pressure and surface pressure. The largest contributor to this component was the Integrated Surface Database (Smith et al., 2011), which consists of global hourly and synoptic surface observations collected from many sources. The land stations were merged following a two-step algorithm to first remove duplicates within a collection and then remove duplicates between collections (Yin et al., 2008).

The marine component consists of sea-level pressure observations extracted from the International Comprehensive Ocean–Atmosphere Data Set (ICOADS, Woodruff et al., 1998; Parker et al., 2004; Woodruff et al., 2005; Worley et al., 2005; Woodruff et al., 2011), a global ocean marine meteorological and surface ocean dataset that is considered the most complete of its kind. It is comprised of measurements and observations extracted from many international data sources, including ship reports, moored and drifting buoys, coastal stations, and other marine platforms. ICOADS release 2.4 was used in the ISPDv2 for the period 1952–2008 and ICOADS release 2.5 (Woodruff et al., 2011) was used for the period 1785–1951 and 2009–2012.

The tropical cyclone component was taken from the International Best Track Archive for Climate Stewardship (IBTrACS; Knapp et al., 2010). The IBTrACS dataset consists of global tropical cyclone best track position and intensity observations, and reports collected from each of the World Meteorological Organization (WMO) Regional Specialized Meteorological
Table 1. Contributing organizations to the International Surface Pressure Databank.

1. All-Russian Research Institute of Hydrometeorological Information World Data Center
2. Atmospheric Circulation Reconstructions over the Earth (ACRE)
3. Australian Bureau of Meteorology
4. Australian Meteorological Association, Todd Project team
5. British Antarctic Survey
6. Cook Islands Meteorological Service
7. Danish Meteorological Institute
8. Deutscher Wetterdienst (DWD)
9. Environment Canada, Climate Research Division
10. ETH Zurich, Switzerland
11. European and North Atlantic Daily to Multidecadal Climate Variability (EMULATE)
12. European Reanalysis and Observations for Monitoring (EURO4M)/The WMO MEDiterranean Data REscue Initiative (MEDARE)
13. European Reanalysis of Global Climate Observations (ERA-CLIM)
14. GCOS Atmospheric Observation and Ocean Observation Panels for Climate Working Group on Surface Pressure
15. GCOS/WCRP Working Group on Observational Data Sets for Reanalysis
16. Hong Kong Observatory
17. Icelandic Meteorological Office (IMO)
18. Instituto Geofísico do Universidade do Porto, Portugal
19. International Best Track Archive for Climate Stewardship (IBTrACS)
20. International Comprehensive Ocean-Atmosphere Data Set (ICOADS)
21. International Environmental Data Rescue Organization (IEDRO)
22. Japan Agency for Marine-Earth Science and Technology (JAMSTEC)
23. Japan Meteorological Agency
24. Jersey Met Department
25. Koninklijk Nederlands Meteorologisch Instituut (KNMI; Royal Netherlands Meteorological Institute)
26. Lamont-Doherty Earth Observatory of Columbia University
27. McGill University, Canada
28. Met Office Hadley Centre, UK
29. Météo-France
30. Météo-France – Division of Climate
31. Meteorological and Hydrological Service, Croatia
32. National Center for Atmospheric Research (NCAR), USA
33. National Institute for Water and Atmospheric Research (NIWA), New Zealand
34. Nicolaus Copernicus University – Department of Meteorology and Climatology, Poland
35. Niue Meteorological Service
36. NOAA Climate Database Modernization Program (CDMP), USA

Table 1. (continued)

37. NOAA Earth System Research Laboratory (ESRL), Physical Sciences Division, USA
38. NOAA Midwest Regional Climate Center at UIUC, USA
39. NOAA National Centers for Environmental Prediction (NCEP), USA
40. NOAA National Climatic Data Center (NCDC), USA
41. NOAA Northeast Regional Climate Center at Cornell University, USA
42. NOAA Pacific Marine Environmental Laboratory, USA
43. Norwegian Meteorological Institute
44. Ohio State University – Byrd Polar Research Center, USA
45. Oldweather.org
46. Portuguese Institute of Sea and Atmosphere (IPMA), Portugal
47. Proudman Oceanographic Laboratory, UK
48. Signatures of environmental change in the observations of the Geophysical Institutes (SIGN)
49. South African Weather Service
50. South Eastern Australian Recent Climate History (SEARCH) project, The University of Melbourne
51. Tanzania Meteorological Agency
52. University of Aberdeen, Scotland, UK
53. University of Bern, Switzerland
54. University of Colorado – Climate DAta REscue (MEDARE)
55. University of East Anglia – Climatic Research Unit, UK
56. University of Giessen – Department of Geography, Germany
57. University of Lisbon – Instituto Dom Luiz, Portugal
58. University of Milan – Department of Physics, Italy
59. University of Porto-Instituto Geofísico, Portugal
60. University Rovira i Virgili – Centre for Climate Change (C3), Spain
61. University of South Carolina, USA
62. University of Toronto – Department of Physics, Canada
63. University of Washington, USA
64. World Meteorological Organization – MEDiterranean climate DAta REscue (MEDARE)
65. Zentralanstalt für Meteorologie und Geodynamik (ZAMG; Austrian Weather Service)

Contributors to version 2 (cf. Compo et al., 2011, Table II).

Centers, Tropical Cyclone Warning Centers, and other national agencies. The IBTrACS Beta version was used for the years 1952–2006, version v01r01 for 1871–1883 and 1886–1951, version v02r01 for 1884–1885 and 2007–2008, version v03r02 for 2009–2010, and version v03r05 for 2011–2012.

The inclusion of various versions of both ICOADS and IBTrACS in ISPDv2 was the result of using the most up-to-date data that were available during the assimilation into the Twentieth Century Reanalysis (20CR; Compo et al., 2011). This rationale is elaborated upon in Section 2.1.
<table>
<thead>
<tr>
<th>ISPD ID</th>
<th>Name</th>
<th>Description</th>
<th>Period</th>
<th>No. stations/records</th>
<th>NCDC reference</th>
<th>NCAR reference</th>
<th>Reference publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>0105</td>
<td>ICOADS Release 2.5</td>
<td>Global Marine Surface Observations</td>
<td>1784–2007</td>
<td>~261 million</td>
<td>DSI-1173</td>
<td>ds540.0, ds540.1</td>
<td>Woodruff et al. (2011)</td>
</tr>
<tr>
<td>0200</td>
<td>ICOADS Auxiliary Kobe</td>
<td>Global Marine Surface Observations</td>
<td>1889–1943</td>
<td>~3135</td>
<td>DSI-1170</td>
<td>ds530.0</td>
<td></td>
</tr>
<tr>
<td>0400</td>
<td>ICOADS Auxiliary Russian</td>
<td>Global Marine Surface Observations</td>
<td>1950–2000</td>
<td>~1789</td>
<td>DSI-1170</td>
<td>ds530.0</td>
<td></td>
</tr>
<tr>
<td>0500</td>
<td>ICOADS Auxiliary Russian</td>
<td>Global Marine Surface Observations</td>
<td>1889–2000</td>
<td>~7873</td>
<td>DSI-1170</td>
<td>ds530.0</td>
<td></td>
</tr>
<tr>
<td>0700</td>
<td>ICOADS Auxiliary Challenger</td>
<td>Global Marine Surface Observations</td>
<td>1872–1876</td>
<td>(n/a)</td>
<td>DSI-1170</td>
<td>ds530.0</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>Federal Climate Complex</td>
<td>Integrated Surface Database</td>
<td>1901–2008</td>
<td>23 363/1.2 billion</td>
<td>DSI-3505</td>
<td>ds463.3</td>
<td>Smith et al. (2011)</td>
</tr>
<tr>
<td>0004</td>
<td>Air Weather Service TD13</td>
<td>Global Land Surface Observations</td>
<td>1901–1973</td>
<td>32/55 748</td>
<td>DSI-1170</td>
<td>ds530.0</td>
<td></td>
</tr>
<tr>
<td>1005</td>
<td>Hadley Centre</td>
<td>Individual Stations from Hadley Centre</td>
<td>1833–present</td>
<td>11/520 594</td>
<td>DSI-1170</td>
<td>ds530.0</td>
<td>Allan et al. (2011)</td>
</tr>
<tr>
<td>1006</td>
<td>CDMP-International</td>
<td>Chile, Mexico, Uruguay</td>
<td>1800s–1980</td>
<td>14/326 175</td>
<td>DSI-1170</td>
<td>ds530.0</td>
<td>Dupigny-Giroux et al. (2007)</td>
</tr>
<tr>
<td>1011</td>
<td>Royal Netherlands Meteorological Institute (KNMI)</td>
<td>KNMI stations</td>
<td>1911–2006</td>
<td>49/629 491</td>
<td>DSI-1170</td>
<td>ds530.0</td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td>NCEP Operational BUFR Archive</td>
<td>Global Observations</td>
<td>1928–1948</td>
<td>1053/2 940 516</td>
<td>DSI-1170</td>
<td>ds530.0</td>
<td></td>
</tr>
<tr>
<td>3004</td>
<td>Environmental Canada Pressure Observations</td>
<td>Environmental Canada</td>
<td>1842–2004</td>
<td>746/25 277 856</td>
<td>DSI-1170</td>
<td>ds530.0</td>
<td></td>
</tr>
<tr>
<td>3005</td>
<td>West African Synoptic observations</td>
<td>West African Land Surface</td>
<td>1850–1980</td>
<td>123/3 022 702</td>
<td>DSI-1170</td>
<td>ds530.0</td>
<td></td>
</tr>
<tr>
<td>3006</td>
<td>The Australian Bureau of Meteorology Station</td>
<td>Australian Land stations</td>
<td>1900–1956</td>
<td>63/1 418 093</td>
<td>DSI-1170</td>
<td>ds530.0</td>
<td></td>
</tr>
</tbody>
</table>

(continued)
<table>
<thead>
<tr>
<th>ISPD ID</th>
<th>Name</th>
<th>Description</th>
<th>Period</th>
<th>No. stations/records</th>
<th>NCDC reference</th>
<th>NCAR reference</th>
<th>Reference publication</th>
</tr>
</thead>
<tbody>
<tr>
<td>3009</td>
<td>Spanish Hourly Pressure Observations from EMULATE</td>
<td>Hourly Spanish Land Stations</td>
<td>1850–2003</td>
<td>5/190 899</td>
<td></td>
<td></td>
<td>Ansell et al. (2006)</td>
</tr>
<tr>
<td>3014</td>
<td>Norway</td>
<td>Land Stations</td>
<td>1863–2007</td>
<td>16/225 687</td>
<td></td>
<td></td>
<td>Ansell et al. (2006)</td>
</tr>
<tr>
<td>3015</td>
<td>Croatia Meteorological and Hydrological Service Land Stations</td>
<td></td>
<td>1858–2005</td>
<td>4/444 398</td>
<td></td>
<td></td>
<td>Ansell et al. (2006)</td>
</tr>
<tr>
<td>4000</td>
<td>Hong Kong Hourly Pressure Observations</td>
<td>Hong Kong Observatory</td>
<td>1885–1939</td>
<td>1/430 920</td>
<td></td>
<td></td>
<td>Valente et al. (2008)</td>
</tr>
<tr>
<td>4001</td>
<td>Jakarta/Batavia Pressure Observations</td>
<td>Dutch Royal Observatory</td>
<td>1866–1944</td>
<td>1/604 518</td>
<td>ds490.0</td>
<td></td>
<td>Valente et al. (2008)</td>
</tr>
<tr>
<td>4002</td>
<td>Liverpool</td>
<td>Proudmans Ocean. Lab stations</td>
<td>1768–1793</td>
<td>1/9994</td>
<td></td>
<td></td>
<td>Valente et al. (2008)</td>
</tr>
<tr>
<td>4003</td>
<td>Jersey Channel Island Pressure Observations</td>
<td>Channel Island Stations</td>
<td>1864–1913</td>
<td>7/159 781</td>
<td></td>
<td></td>
<td>Valente et al. (2008)</td>
</tr>
<tr>
<td>5000</td>
<td>Antarctic Expedition keyed by Hadley Centre</td>
<td></td>
<td>1899–1941</td>
<td>11/119 991</td>
<td></td>
<td>Allan et al. (2011)</td>
<td></td>
</tr>
<tr>
<td>8001</td>
<td>International Best Track Archive for Climate Stewardship (IBTRACS)</td>
<td>National Climatic Data Center</td>
<td>1948–present</td>
<td></td>
<td>DSI-9637</td>
<td>Knapp et al. (2010)</td>
<td></td>
</tr>
<tr>
<td>010000-019999</td>
<td>NCAR Upper Air Stations</td>
<td>National Center for Atmospheric Research</td>
<td>1943–1998</td>
<td>1081/1 275 272</td>
<td>ds370.1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The columns denote, respectively, the ISPDv2 collection ID, collection name, description, yearly period of record, total number of stations and observational records in the collection, reference number for the corresponding National Climatic Data Center (NCDC) dataset, reference number for the corresponding dataset archived in the Research Data Archive (RDA) at the National Center for Atmospheric Research (NCAR), and reference publication.

*The Air Weather Service TD13 observational records originating from collection ID 1004 are incorrectly labeled with collection ID 2001 in the native HDF5 tables in both ISPDv2 and ISPD version 3.*
1. Dataset content and coverage

The ISPDv2 dataset is stored in the Hierarchical Data Format version 5 (HDF5: http://www.hdfgroup.org), which allows for efficient archiving of large, complex data structures and a diverse set of metadata. Each HDF5 file contains 13 tables and 15 directory subgroups that define how to decode the individual records, and each observation record includes metadata defining the data collection source and observation type (e.g. station observation, marine observation, surface pressure observation from a radiosonde sounding).

Because the ISPDv2 was assimilated as observational input into the 20CR, each record also includes the 20CR quality control and assimilation feedback metadata. These quality control and feedback metadata include the results of the five-step quality control procedure described in Appendix B of Compo et al. (2011) and all the relevant statistical quantities returned by the 20CR ensemble data assimilation system. The format of the HDF5 metadata, therefore, is designed to allow traceability of observations from their original source to the ISPDv2, and to permit direct feedback from the 20CR data assimilation back to the original source archives.

Each observation at any single observation time is assigned a unique number that when combined with the time stamp (year, month, date, hour, and minute) forms a unique identifier within the complete dataset. Thus, every observation has an unambiguous reference within the full ISPDv2 dataset, which allows for precise identification and traceability. This data management feature is critical for accurately adding additional metadata from other future reanalyses, and evolving the ISPDv2 to new versions while maintaining the usage provenance of each record.

The 20CR required hourly files from the ISPDv2 data for the data assimilation system, thus each data file in the ISPDv2 collection contains one hour of global observations. Based on this file organization and data coverage, a maximum of 24 data files per day and 8760 data files per (non-leap) year exist in the ISPDv2 collection. The total dataset volume is 491.32 GB. The HDF5 data are available for download as either individual hourly data files, or as yearly tar files, which are tarballs of one entire year of HDF5 data files.

The ISPDv2 is also available in NetCDF and ASCII column data formats, both of which are derived from the native HDF5 format. The data content within these two formats is not a complete reproduction of the tabular content and metadata that are contained within the HDF5 data. It does contain, however, essential information, including the 20CR quality control feedback information, for most users to carry out their research projects. The ASCII data are available as monthly tar files for direct download from the National Center for Atmospheric Research (NCAR) Research Data Archive (RDA) website (http://dx.doi.org/10.5065/D6SQ8XDW), while the NetCDF data are produced on demand through the data subsetting service described in Section 3.1.

The data availability varies considerably over the period of record. Maps of annual land station distribution (Figure 1) highlight the steady growth in the number of reporting stations, and hence data coverage, since the year 1850 (station coverage maps for every year are available at http://www.esrl.noaa.gov/psd/data/ISPD/v2.0/). The sources of land station

![Figure 1. Annual distribution of ISPDv2 land station locations for the years 1850, 1900, 1950, and 2000. Station coverage maps for every year in the ISPDv2 are available at http://www.esrl.noaa.gov/psd/data/ISPD/v2.0/.](image-url)
observations in the 19th and early 20th centuries are located primarily in Europe and North America. The data coverage increases dramatically across most of the rest of the globe from 1950 to the present.

Figure 2 shows time series of the total number of land observing stations in each indicated continental region by year from 1850 to 2010. The notable decline in stations during the mid-1960s—primarily in the Asia/Eastern Europe time series—is attributed to the reduction in the Integrated Surface Database records during this period, and is explained in Smith et al. (2011) as the result of the transition from keying of observational data to digital transmission and receipt of data.

The Asia time series also exhibits a considerable decline in station coverage starting in the mid-1980s. A listing of the number of stations originating from the Integrated Surface Database for the years 1963–1964 and selected years between 1980 and 2001 is shown in Table 3, which reveals these reductions are primarily associated with the observations originating from the Russian Federation and other former Soviet Union countries. These reductions can be explained by the deterioration of funding of the meteorological observation network in the final years of the Soviet Union existence and thereafter in the 1990s and 2000s. The observation network health during this post-Soviet period was a function of the economic situation in each of the 15 newly independent nations (and, to some extent, a function of foreign sponsorship). The China, North and Central America, and South-West Pacific regions also exhibit notable decreases between 1963 and 1964.

The average number of pressure observations per day contained in the 20CR feedback records from all ISPDv2 components (station, marine, and tropical cyclone) is illustrated by maps for the years 1900, 1950, and 2000 in Figure 3. The preponderance of observations in the Northern Hemisphere in earlier years is evident, as is the growth of marine observations over time (see Woodruff et al., 2011 for a complete description).

The total number of observations per year in the ISPDv2 increases steadily from approximately 100 000 records in 1870 to 53 million in 2010 (Figure 4). A local increase and subsequent decrease in this trend during the 1878–1894 period is attributed to the addition of 1.8 million records from the US Marine Meteorological Journals and digitized for ICOADS Release 2.5 (Woodruff et al., 2011). A decrease during the World War I years is evident, while an increase during the World War II years is attributed to the inclusion of 1.6 million ICOADS records digitized from the UK Royal Navy Ships’ logbooks (Brohan et al., 2009; Woodruff et al., 2011). A decrease during the mid-1960s is associated with the aforementioned decline in station records in the Integrated Surface Database (Smith et al., 2011). A decrease in 2007 is most likely due to a delay in processing, and therefore incomplete records, in the Integrated Surface Database. Other small-scale variations in the curve likely can be correlated with increased scientific interest (local peaks) and research budget reductions (local troughs; Fleming, 2000).

The dataset contains reports from more than 10 000 land stations, which, in addition to the marine observing systems reports, comprise over 1.5 billion observations in total. Between 1860 and 1917, the total number of marine and surface station observations is comparable, but in later years the total count

![Figure 2](image-url). Time series of the total number of ISPDv2 land observing stations in each indicated continental region by year from 1850 to 2010. The inset panel indicates the boundaries used for the continental regions.
is dominated by land surface station reports (not shown). This prevalence of land station observations is primarily due to the input data compiled for the National Centers for Environmental Prediction (NCEP)/NCAR Reanalysis project (1948–2003; Kalnay et al., 1996; Kistler et al., 2001).

2. Data usage and application

2.1. Reanalysis

The primary motivation for the creation and development of the ISPD was to facilitate progress on research and understanding of long-term trends and variations in global surface pressure. In particular, a long-term historical dataset, such as the ISPD, becomes essential to providing an observational underpinning to retrospective climate analysis datasets, commonly known as reanalyses. Reanalysis products are used extensively in climate research, applications, and services, including for monitoring and comparing current climate condi-

Figure 3. Maps showing the average number of pressure observations per day contained in the 20CRv2 feedback records from all ISPDv2 components: station, marine, and tropical cyclone for the years 1900 (top), 1950 (middle), and 2000 (bottom). Counts of observations are made in five-degree grid boxes.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>47</td>
<td>28</td>
<td>877</td>
<td>1007</td>
<td>925</td>
<td>927</td>
<td>869</td>
<td>1058</td>
</tr>
<tr>
<td>1</td>
<td>184</td>
<td>166</td>
<td>717</td>
<td>813</td>
<td>740</td>
<td>748</td>
<td>777</td>
<td>935</td>
</tr>
<tr>
<td>2</td>
<td>914</td>
<td>223</td>
<td>800</td>
<td>695</td>
<td>683</td>
<td>577</td>
<td>410</td>
<td>462</td>
</tr>
<tr>
<td>3</td>
<td>795</td>
<td>124</td>
<td>1031</td>
<td>878</td>
<td>835</td>
<td>751</td>
<td>607</td>
<td>584</td>
</tr>
<tr>
<td>4</td>
<td>466</td>
<td>353</td>
<td>916</td>
<td>932</td>
<td>864</td>
<td>876</td>
<td>867</td>
<td>1049</td>
</tr>
<tr>
<td>5</td>
<td>518</td>
<td>347</td>
<td>727</td>
<td>598</td>
<td>592</td>
<td>587</td>
<td>558</td>
<td>638</td>
</tr>
<tr>
<td>6</td>
<td>211</td>
<td>222</td>
<td>689</td>
<td>768</td>
<td>758</td>
<td>776</td>
<td>825</td>
<td>887</td>
</tr>
<tr>
<td>7</td>
<td>484</td>
<td>280</td>
<td>1524</td>
<td>1594</td>
<td>1671</td>
<td>1697</td>
<td>1560</td>
<td>1770</td>
</tr>
<tr>
<td>8</td>
<td>132</td>
<td>132</td>
<td>550</td>
<td>695</td>
<td>661</td>
<td>684</td>
<td>630</td>
<td>671</td>
</tr>
<tr>
<td>9</td>
<td>689</td>
<td>626</td>
<td>578</td>
<td>865</td>
<td>874</td>
<td>898</td>
<td>860</td>
<td>841</td>
</tr>
</tbody>
</table>

The columns represent the station count for the years listed and organized by WMO regional block number (0 and 1 = Europe; 2 and 3 = Russian Federation and other former Soviet Union countries; 4 = Asia; 5 = China; 6 = Africa; 7 = North America, Central America, and the Caribbean; 8 = South America and Antarctica; and 9 = South-West Pacific).

Figure 4. Time series of the number of pressure observations per year in version 2 of the International Surface Pressure Databank (ISPDv2) from 1870 to 2010. Note the logarithmic scale along the y-axis. Inset panel: time series during the same period showing the number of observations in the Northern Hemisphere (blue curve) and Southern Hemisphere (red curve).
tions with those of the past, identifying the causes of climate variations and change, and preparing climate predictions. Information derived from reanalyses is also being used increasingly in commercial and business applications in sectors such as energy, agriculture, water resources, and insurance.

Prior to completion of the ISPD, most reanalysis products (including those from the NCEP/NCAR Reanalysis Project (Kalnay et al., 1996; Kistler et al., 2001) and the European Centre for Medium-Range Weather Forecasts [ECMWF; Uppala et al., 2005]) could only extend back to about 1950 since they relied on assimilating a set of observations that included upper-level atmospheric information. The primary input data source for the NCEP/NCAR Reanalysis, for example, are the global rawinsonde observations, which contain a period of record substantial enough to produce that reanalysis back to 1948. The limited time range of these reanalyses restricts their usefulness for many climate research applications. The ISPD adds value by providing a much longer period of record and thus enables the development of reanalyses with longer time spans.

The first reanalysis to make use of the ISPD is the 20CR Project (Compo et al., 2011), which assimilates only the ISPDv2 surface and sea-level pressure observations and prescribes observed monthly sea-surface temperature and sea-ice distributions from HadISST1.1 (Rayner et al., 2003) as boundary conditions. This global reanalysis dataset spans the late 19th century, the entire 20th century, and the early 21st century (1871–2012). It estimates the state of the atmosphere – the ‘analysis’ – by combining the hourly and synoptic ISPDv2 observations in a 6-h time window with a dynamically generated 9-h first guess forecast initialized from the previous analysis. Cycling this procedure with overlapping 6-h analyses and 9-h forecasts has the effect of spreading the observational information from the ISPD three-dimensionally in space and in time.

As the 20CR was produced, the early period ISPDv2 ‘beta’ version was revised to include newly digitized additions and improve the overall coverage of stations. As ICOADS Release 2.5 was then available, it was also included to increase the number of marine observations in the early period. Similarly for IBIrrACS, newly digitized records from the most recent versions of IBIrrACS were added to ISPDv2 as it evolved and improved over time. The final ISPDv2, therefore, contains multiple source versions of ICOADS and IBIrrACS. The newer versions of ICOADS and IBIrrACS did not change appreciably except to expand their data inputs and coverage, therefore we do not anticipate that the use of multiple versions introduces inconsistencies in the pressure dataset.

2.2. Twentieth century reanalysis quality control and data assimilation feedback

During the data assimilation processing of the 20CR, the ISPDv2 input data were interrogated by a five-step quality control procedure that included, among other steps, checking the observations for meteorological plausibility, comparing them with neighbouring data values, and performing a bias correction on the land component (see Compo et al., 2011 for more information on this procedure). The 20CR then applied an Ensemble Kalman Filter data assimilation method which used background first-guess fields supplied by an ensemble of 56 forecasts from an experimental version of the NCEP Global Forecast System (GFS) numerical weather prediction model (Kanamitsu et al., 1991; Moorthi et al., 2001; Saha et al., 2006). These quality control and data assimilation results, which include the first-guess fields, analysis departures, bias estimates, and observation errors, were then written back into the ISPDv2 so that each observational record contains this feedback information. Future users of the ISPDv2, and those who contributed the original observational data, therefore can utilize this information to make an informed decision on the quality and usefulness of each observational record during the 20CR time period.

One project that utilized the 20CR feedback metadata is the ERA-20C reanalysis (Poli et al., 2013), which is the first reanalysis produced under the European Reanalysis of Global Climate Observations database project (ERA-CLIM: www-era-clim.eu). The ERA-20C is a global reanalysis that spans the 20th century (1900–2010) assimilating only surface and sea-level pressure observations from ISPD and ICOADS and marine surface winds from ICOADS. Feedback information from the 20CR was used in the bias correction scheme for ISPD data that were assimilated in the ERA-20C reanalysis; this was based on a break-point analysis using the 20CR first-guess departures (Hersbach et al., 2015). In locations where breakpoints were suspected (e.g. instances of instrument errors or a change in station location), the ERA-20C bias correction scheme assigned less confidence to the 20CR first-guess value, and thus was allowed to be more adaptive compared to cases where no irregularities in long-term departures were identified.

In another study, Wang et al. (2014) used the 20CR feedback metadata to show that the strong extratropical cyclone events in the 20CR agree well with the geostrophic wind extremes derived from in situ pressure observations. This study illustrates how the 20CR quality control procedure performed well in identifying errors in the observational record. For example, the 20CR quality control system identified and rejected 143 of 146 erroneous observation values from the Aberdeen, Scotland records for the period 1871–1921. This quality control feedback information, therefore, is valuable in identifying observation errors which might otherwise be overlooked and, coupled with continued data rescue efforts such as those facilitated by ACRE (Allan et al., 2011), serves to strengthen and reduce uncertainty in the observational record.
3. Data access and user services

3.1. Research data archive data location and accessibility

The primary repository for ISPDv2 is in the RDA (http://rda.ucar.edu) at NCAR in Boulder, Colorado. The RDA is a free and open data collection where data discovery can be achieved through faceted searches based on Global Change Master Directory (GCMD: http://gcmd.nasa.gov) metadata keywords, free text queries, and lists highlighting the most used datasets. Data access is free, but requires each user to register through a simple online process that validates the submitted email address.

The ISPDv2 data access is organized by year and month on the RDA website (http://dx.doi.org/10.5065/D6SQ8XDW), and users may browse through the pages to locate specific data of interest. Data files are organized in default groups by observation year and month. They can be downloaded directly from the RDA web interface using server-supplied ‘wget’ scripts. Users may select a collection of files to download from default lists by using the ‘Web File Listing’ option under the ‘Data Access’ tab.

A more refined option is provided through the data subset request form, which may be accessed via the ‘Get a subset’ link under the ‘Data Access’ tab. This produces a customized data subset of the HDF5 data based on user-provided constraints. Users can specify the temporal limits, spatial domain, observation type (s), and ASCII or NetCDF data output formats through the data subset request form (Figure 5). The output file compression can also be requested on this form (not shown). Observations from individual observing stations may also be requested, which is a useful feature for users who wish to procure time series for specific locations or regions. Once submitted, the data subset requests are produced by a delayed mode data processing procedure. Users are notified and directed to a web download location when their data request output files are accessible.

3.2. Dataset citation

The RDA also provides a data citation service. The dataset homepage provides citation syntax in the standard forms recommended by the Federation of Earth System Information Partners (ESIP), American Meteorological Society (AMS), American Geophysical Union (AGU), DataCite, and the Geoscience Data Journal. This dataset citation is also available in Research Information System (RIS) format so that users may import the citation for this dataset directly into their citation reference management software (e.g. EndNote, Zotero, etc.). One key element in the dataset citation is the data access date (‘Accessed dd mmm yyyy’). Leveraging the fact that RDA data users are registered, a customized data citation can be prepared for the user by using the ‘Get a customized data citation’ link on the dataset home page. The dates that users access the data are recorded and can be retrieved on demand at later times.

3.3. File content metadata

File content metadata are collected as part of the data processing and preparation of the ISPDv2 for the RDA. The metadata supports additional information services for interested users. The overall ISPDv2 metadata summary is provided from the dataset home page under the ‘More Details’ link. Here, tabulations of the observing station location, observation type, platform identification, and maps of the global distribution for the full ISPDv2 can be viewed (Figure 6).

In addition, users may view the content metadata for the yearly tar file archives by clicking on a looking glass icon adjacent to the tar file listing, accessible under the ‘Data Access’ tab. This service enables users to query the ISPDv2 via an interactive map of land station locations, and more efficiently determine the observational data available at any particular time and region of interest (Figure 7).

3.4. Supporting software and documentation

Source code developed at the University of Colorado CIRES and NOAA Earth System Research Laboratory and written in C language to read and decode the HDF5 data into ASCII column output is provided on the RDA website under the ‘Software’ tab on the dataset home page. Documentation describing the HDF5 tables and ASCII output is also available for download under the ‘Documentation’ tab on the dataset home page.

3.5. ECMWF data location and accessibility

A subset of the ISPDv2 is also archived at ECMWF (http://apps.ecmwf.int/datasets). Here, users may extract temporal subsets from the ISPDv2 archive stored at ECMWF and refine their selection by observation platform. The output from these requests is written in ASCII column format and contains the date, location, observed value, report type, unique identifier which allows traceability to other ISPD versions, and several feedback parameters from the 20CR, including the 20CR bias estimate, the 20CR ensemble mean first guess pressure minus the ‘modified’ observed pressure, and the ensemble mean analysis pressure minus the ‘modified’ observed pressure. The ‘modified’ pressure is the pressure value after the 20CR assimilating model. This is similar to a reduction to sea level but instead is an adjustment to the 20CR orography in the assimilating model. This value is then subtracted from the pressure at the observing station location, which could be higher or lower than the station elevation. In addition to the ASCII output, users may produce observation count maps for their data selection on the ECMWF web interface.
4. Future plans

Since the release of ISPDv2, over 22 additional organizations have contributed to the station and marine components of the ISPD (Table 1) and the period of record now extends back to 1755. These new contributions have been incorporated into a newer version 3 of the ISPD (ISPDv3), which will be made available in the near future through the NCAR RDA. The observation feedback archive (OFA) version of this dataset as used in the ERA-20C reanalysis is available from ECMWF (Hersbach et al., 2015).

The gain in observations for ISPDv3 is illustrated by the time series shown in Figure 8. Most of the increase occurs in the Northern Hemisphere, and there are visible increases in the late 19th century and during the World War I years. The 2007 decrease in merged records from the Integrated Surface Database into ISPDv2 (cf. Figure 4) is recovered in ISPDv3. These and increases in other years come from diverse station and marine collections contributed to the ISPD under the auspices of GCOS, WCRP, and the ACRE initiative. This includes the initial efforts of the OldWeather.org citizen science project that focused on UK

Figure 5. The data subsetting interface found on the ISPDv2 Research Data Archive webpage after user authentication. Users choose temporal range, spatial selection (choices are interactive map functions, manual latitude/longitude entry, or station identifier entry), and also can select one or many observation types, and ASCII or NetCDF output data formats.
### Summary for platform observations in HDF5 format

<table>
<thead>
<tr>
<th>Observing platform</th>
<th>Observation type</th>
<th>Average frequency of data (varies by individual platform ID and data type)</th>
<th>Temporal/geographical coverage (each dot represents a 3° box containing one or more observations)</th>
</tr>
</thead>
</table>
| CMAN station       | Surface         | • Assimilation feedback parameters  
                    • Observed sea level pressure (hPa)  
                    • 1 to 6 per min  
                    • 1 to 51 per h  
                    • 1 to 20 per day | 1983-01-01 to 2011-12-31                                                                 |
| Coastal station    | Surface         | • Assimilation feedback parameters  
                    • Observed sea level pressure (hPa)  
                    • 1 per day  
                    • 5 per week | 1986-01-01 to 1989-12-31                                                                 |
| Drifting buoy      | Surface         | • Assimilation feedback parameters  
                    • Observed sea level pressure (hPa)  
                    • Observed surface pressure (hPa)  
                    • 1 to 10 per min  
                    • 1 to 50 per h  
                    • 1 to 20 per day  
                    • 1 to 6 per week  
                    • 1 to 4 per month  
                    • 1 to 10 per year | 1899-01-01 to 2011-12-31                                                                 |
| Fixed ocean platform| Surface         | • Assimilation feedback parameters  
                    • Observed sea level pressure (hPa) | 1971-01-01 to 1997-12-31                                                                 |

**Figure 6.** Example of the content metadata interface found on the ISPDv2 Research Data Archive webpage. Here, users can view and investigate tabulations of the location, observation type, platform identification, and maps of the global distribution of observations for the full dataset.

Royal Navy observations during 1914–1923, the citizen-led Todd Project team (http://www.charlestodd.net) that focused on recovering observations assembled by Sir Charles Todd, Meteorologist of the Colony of South Australia during 1879–1909, and the efforts of many other university, national, and international organizations (Table 1). Another notable contribution are the North African locations, which arose from the joint effort between the European Union-funded EUropean Reanalysis and Observations for Monitoring (EURO4M: http://www.euro4m.eu) project in connection with the WMO/MEditerranean climate DAta REscue (MEDARE) initiative (http://www.omm.urv.cat/MEDARE; Brunet et al., 2014). The merging of these contributions results in increases that are particularly dramatic when viewed as a map, as shown for the example year of 1918 (Figure 9). The increases over eastern South America, eastern Africa, China, New Zealand, and selected central Pacific islands are especially evident.

ISPDv3 (specifically, version 3.2.6) was implemented into the European Reanalysis of Global Climate Observations database (ERA-CLIM: www.era-clim.eu), with a goal of using these and additional observations...
in an ECMWF data assimilation system to generate new global climate reanalyses of the 20th century. The ERA-20C global reanalysis, the first reanalysis under this project, spans the 20th century (1900–2010) assimilating only surface and sea-level pressure observations from ISPDv3 and marine wind observations from ICOADS 2.5 using a 4D-Var assimilation system (Poli et al., 2013).

In addition, ISPD version 3.2.9 is being assimilated into the next generation of the 20CR (version 2c). 20CRv2 uses the same NCEP atmosphere/land coupled model at the same spectral resolution of total wavenumber 62 (about 200 km by 200 km) as 20CRv2 but has different prescribed boundary conditions. Daily sea surface temperatures come from a new Simple Ocean Data Assimilation with sparse input version 2 (SODAsi.2, B.S. Giese, H.F. Seidel, G.P. Compo and P.D. Sardeshmukh, in review). Prescribed monthly averaged sea ice concentrations from the new COBE-SST2 (Hirahara et al., 2014) correct the known misspecification in 20CRv2 (Compo et al., 2011; Brönnimann et al., 2012). This activity is currently underway and will be reported on in due course.

The ISPD also will be used in the planned centennial coupled reanalysis being developed under the Japa-
Comments and questions on the ISPD can be directed to the RDA data specialist or left at the reanalyses.org webpage for this dataset (http://reanalyses.org/observations/international-surface-pressure-databank).

Acknowledgements

Construction of the ISPD is supported by the U.S. Department of Energy Office of Science (Biological and Environmental Research program) and the NOAA Climate Program Office. The Twentieth Century Reanalysis Project and the International Surface Pressure Databank used resources of the National Energy Research Scientific Computing Center and of the Oak Ridge Leadership Computing Facility at Oak Ridge National Laboratory, which are supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231 and Contract No. DE-AC05-00OR22725, respectively. Manola Brunet and Philip D. Jones are grateful to the European Union-funded projects EUropean Reanalysis and Observations for Monitoring (EURO4M, FP7-SPACE-2009-1 Project No. 242093) and Uncertainties in Ensembles of Regional Reanalyses (UERRA, FP7-SPACE-2013-1 Project No. 607193). Pavel Groisman was partially supported by Grant 14.B25.31.0026 of the Russian Ministry of Education and Science. Renate Auchmann was supported by the Swiss National Science Foundation (SNF; Project TWIST, 200021_146599/1).

The authors thank the following individuals and organizations for their contributions: (1) Rajmund Przybylak of the Nikolaus Copernicus University, Department of Meteorology and Climatology, Poland. (2) Frank Le Blancq and many others, Jersey Met Service, for Channel Island data. (3) The IT group at CIERES and NOAA/ESRL Physical Sciences Division. (4) Prashant Sardeshmukh of University of Colorado/CIRES for valuable guidance and input. (5) Paul Dellamarta and MeteoSwiss colleagues who originally provided data to EMULATE. (6) Ingeborg Auer of Zentralanstalt für Meteorologie und Geodynamik (ZAMG) who originally provided Austrian data to EMULATE. (7) Clive Wilkinson of NCDC and Climatic Research Unit, University of East Anglia, Norwich, UK, and Eric Freeman of NCDC for World War II UK Royal Navy Data. (8) David Tse of Hong Kong Observatory for Hong Kong data. (9) Volker Wagner for a subset of the Deutscher Wetterdienst HISTOR marine data. (10) Società MeteoItaliana (SMI) for the Torino record. (11) Mac Benoy and the Australian Meteorological Association, Todd Project team. (12) Dick Dee, ECMWF, and the ERA-CLIM project. (13) Manuel Barros, Instituto Geofísico da Universidade do Porto, University of Porto, Vila Nova de Gaia, Portugal. (14) Jonathan Burroughs, Joseph Elms (retired) and Thomas C. Peterson, NOAA National Climatic Data Center, Asheville, NC, USA. (15) Karen Andsager (retired), formerly of Midwestern Regional Climate Center, Champaign-Urbana, IL, USA.
The international surface pressure databank

References


Hersbach H, Poli P, Dee D. 2015. The observation feedback archive for the ICOADS and ISPD data sets. ERA Report Series, 18, ECMWF, UK.


