

# Historical (Surface) Weather Observations for Global Reanalysis

Phil Jones

CRU, UEA, Norwich, UK

# Summary

- Surface Weather variables: pressure, temperature, humidity and wind direction
- What Time Period?
- ICOADS – marine data
- Terrestrial data – using ISPD, ISTI and HadISDH
- How much data are out there?
- Status of Historic Data Digitization
- Different uses and users for daily digitized data – Reanalysis input is just one of these
- Known Sources of Data – like SYNOP
- Data that can be used to verify/validate Reanalyses
- Homogeneity

# What Time Period?

- Reanalyses require SST/Sea-Ice and Forcing Datasets (GHGs, volcanoes, aerosols, land-use) as input
- The main constraint as to how far back we can go is the accuracy of these fields and reliable SST/Sea-Ice is the key determinant
- 20CR goes back to 1851, ERA-20C to 1901
- More SST and other marine observations are being digitized (within ICOADS), but the further back you go the more this will always be spatially limited
- So, probably about 1850 is the limit, but from 1900 will be much more reliable

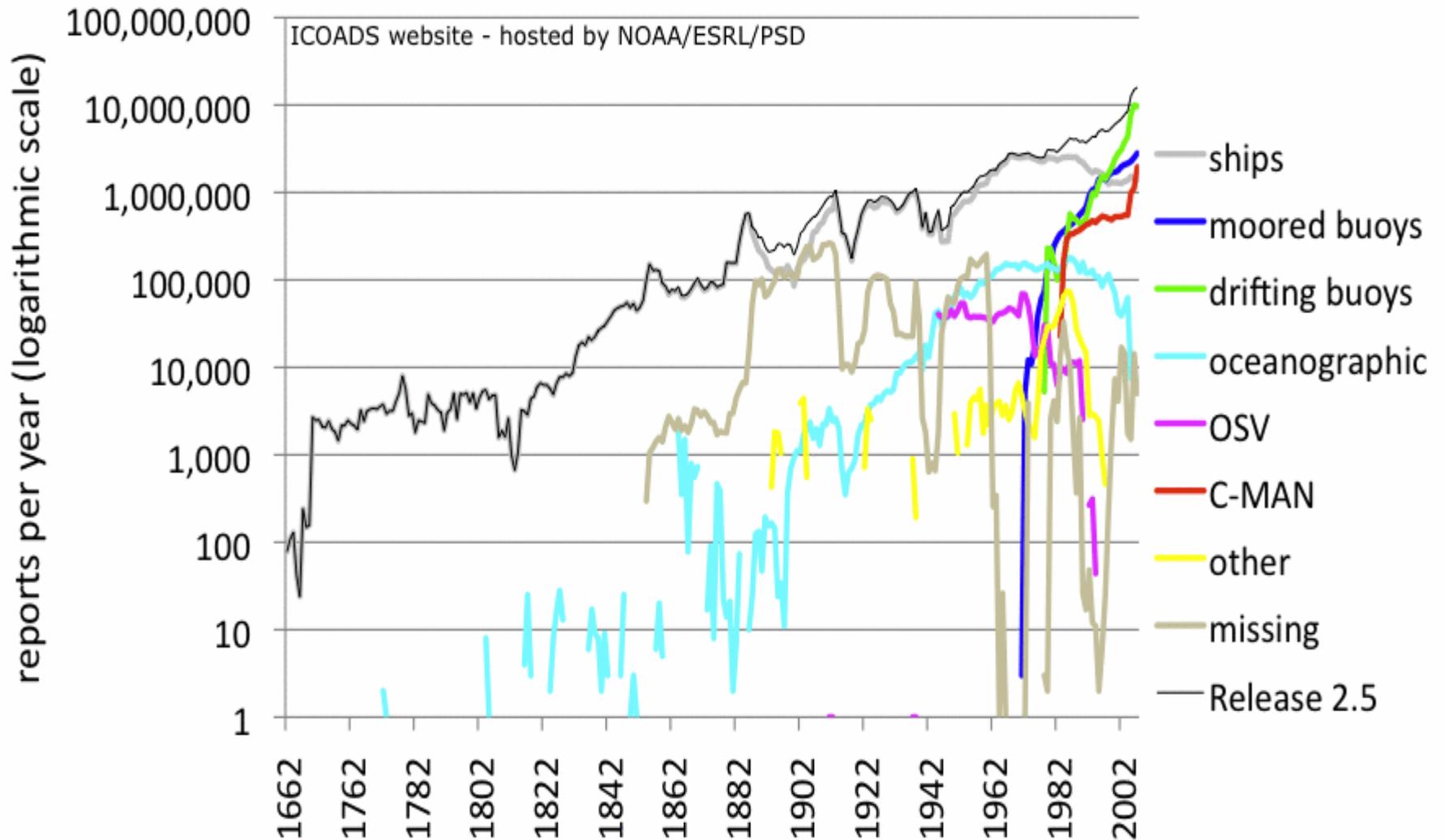
# What period is needed?

- This depends on the region
- Input data are more useful where they are scarce
- Even recently, more in Antarctica would help , as would more in Africa, South America and southern Asia, and other data sparse oceanic regions
- More likely that NMSs can supply additional data for more recent decades (but will they?), whereas data rescue is more likely for earlier decades in the 20th century
- The prospects of getting more is discussed for the oceans and then for terrestrial regions. For each it is necessary to know what we currently have and what the potential might be

# ICOADS (International Comprehensive Ocean Atmosphere Dataset)

- Main Marine Database of Ship-based measurements (<http://icoads.noaa.gov/> )
- Includes Pressure, Air and Sea Surface Temperature, Humidity, Wind Direction and Wind Speed (Beaufort estimate) associated with location and time
- ICOADS is the basic source dataset for HadSST3 and HadISST2 gridded datasets
- Being enhanced by digitizing of more logbook material (from UK ships, whaling logbooks etc)
- Digitizing using crowdsourcing through <http://www.oldweather.org>

# ICOADS - # of Reports/Year



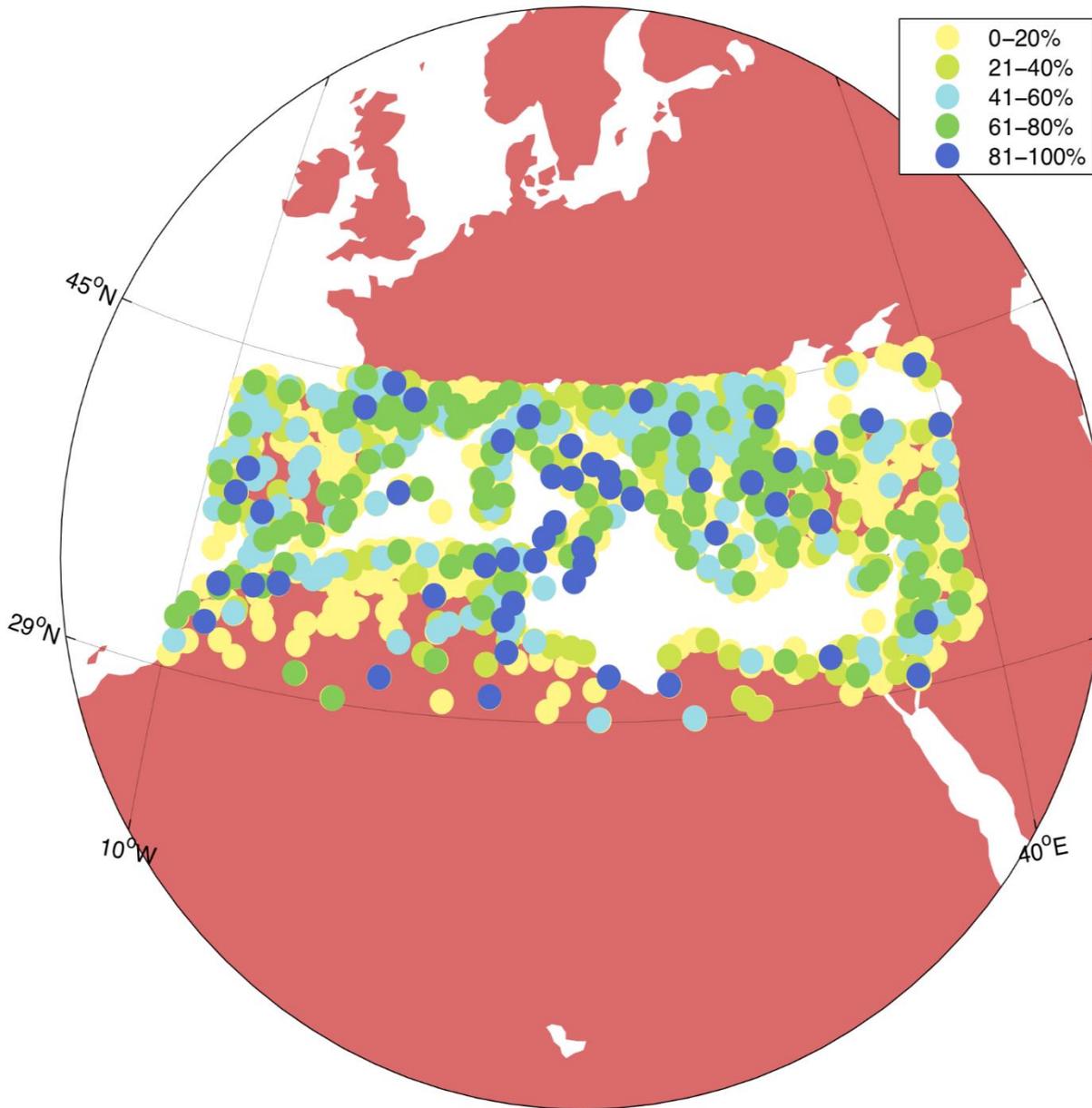
# Terrestrial Atmospheric Variables

- Pressure
- Air Temperature
- Humidity
- Wind Direction
- Currently there are separate archives (e.g. ISPD, ISTI and Humidity using ISDLite)

# Known Sources of Terrestrial Data

- SYNOP+METAR – in ECMWF's MARS Archive
- CLIMAT – monthly averages/totals
- NMSs – some make the data freely available. Most that do have been added. Many do not (e.g. Malaysia)
- Datasets like E-OBS, make many of the station series available
- HadEX2, again many of the daily station series are available – but sometimes just Indices
- Both these are daily Tx/Tn and Precipitation, so not for input to Reanalyses, but useful for validation. Not always clear how independent the data is though!

ecmwfstats.45N.29N.040E.010W.dat (SLP monthly coverage 1960–2010)



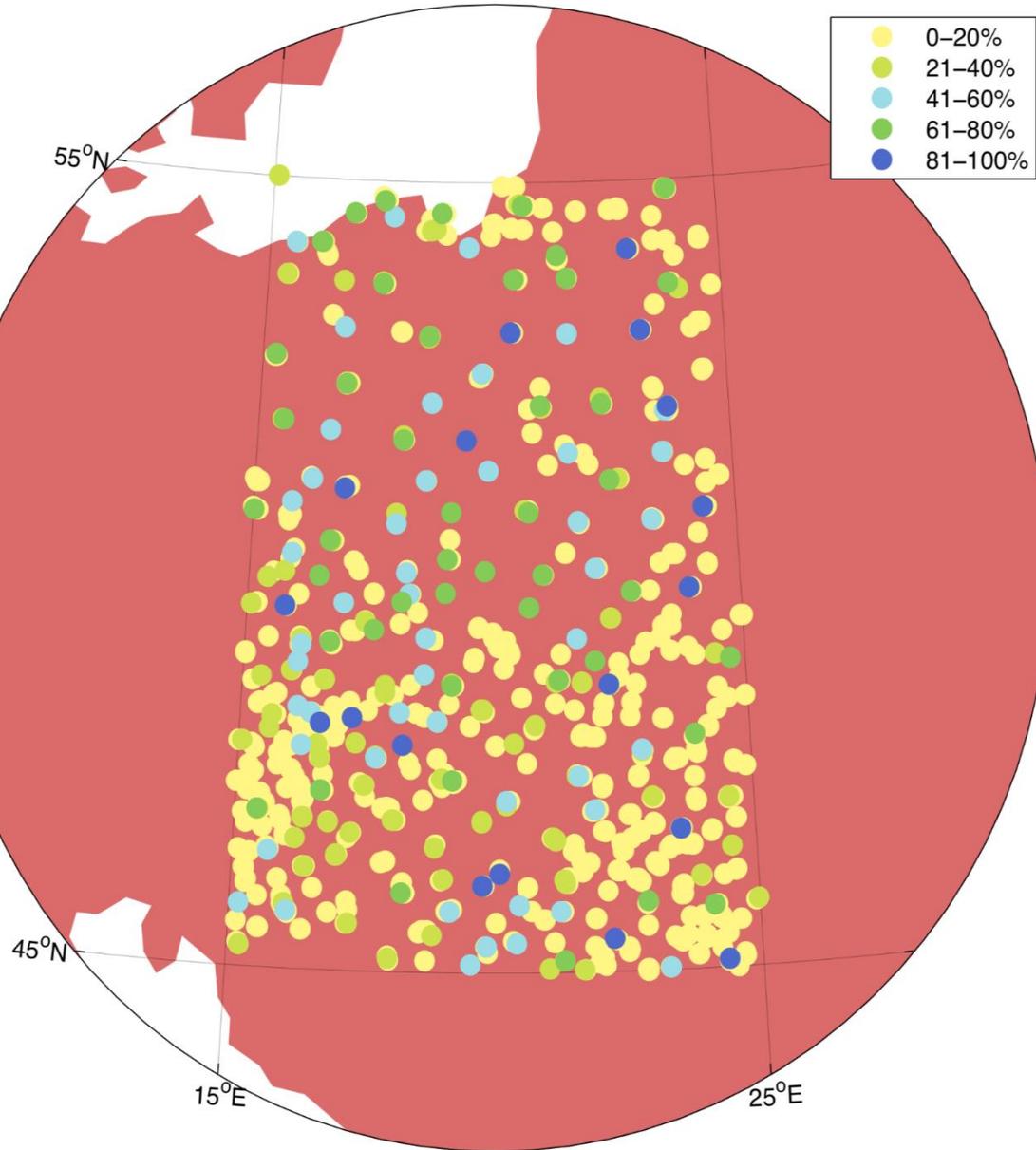
## What is the MARS Archive

Mediterranean. ECMWF archive appears to have more stations that are more complete than might be expected (compared to the other two European regions)

This map was produced as part of the UERRA project, where the contents of Archive for surface variables were assessed.

Later UERRA looked at the homogeneity of the record – see later

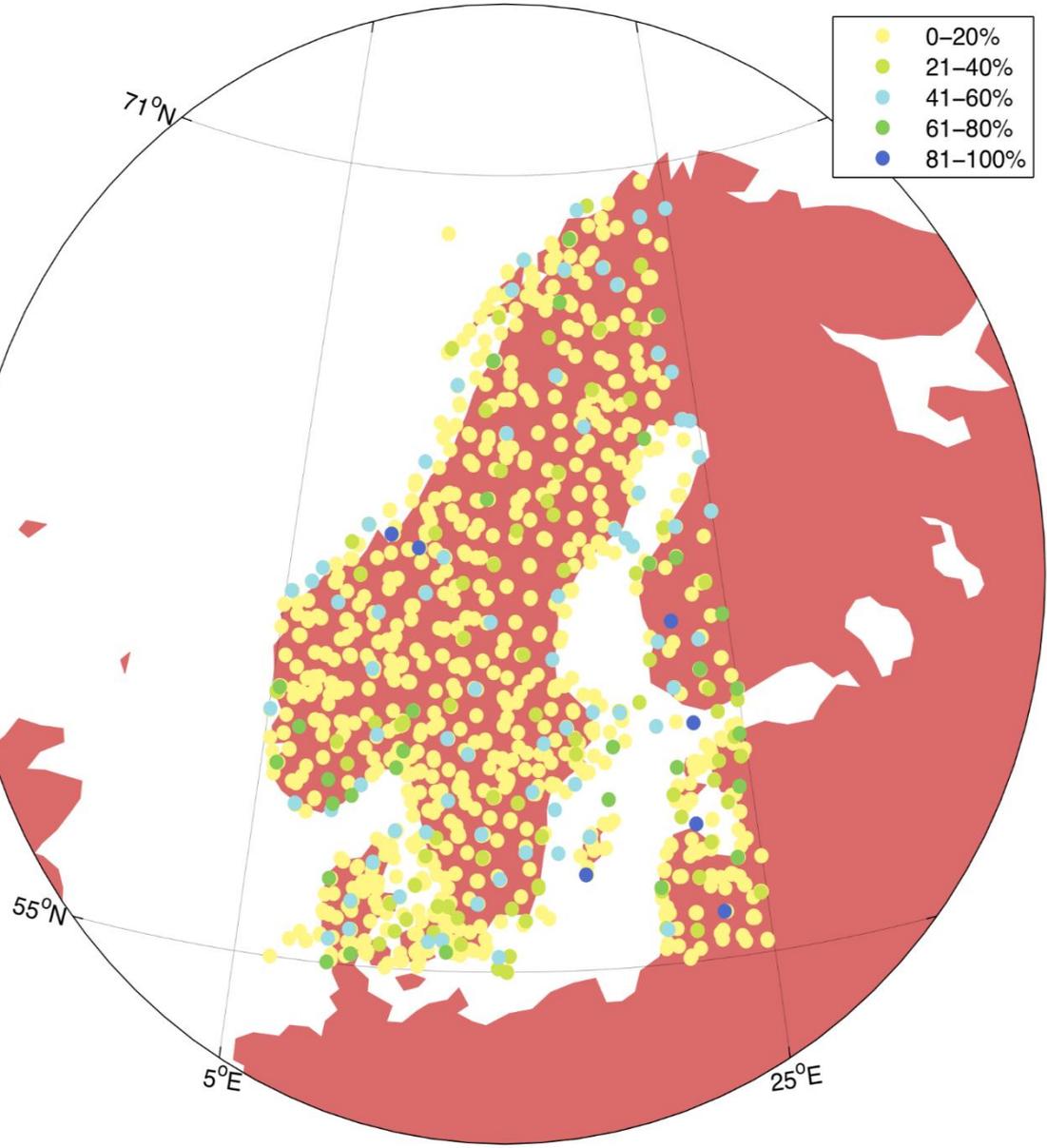
ecmwfstats.55N.45N.025E.015E.dat (SLP monthly coverage 1960–2010)



Eastern Europe – from Poland southwards

Archive appears to have less complete stations, but this is very susceptible to what each NMS releases. Poland, for example, releases very few series over the SYNOP network

ecmwfstats.71N.55N.025E.005E.dat (SLP monthly coverage 1960–2010)



## Scandinavia

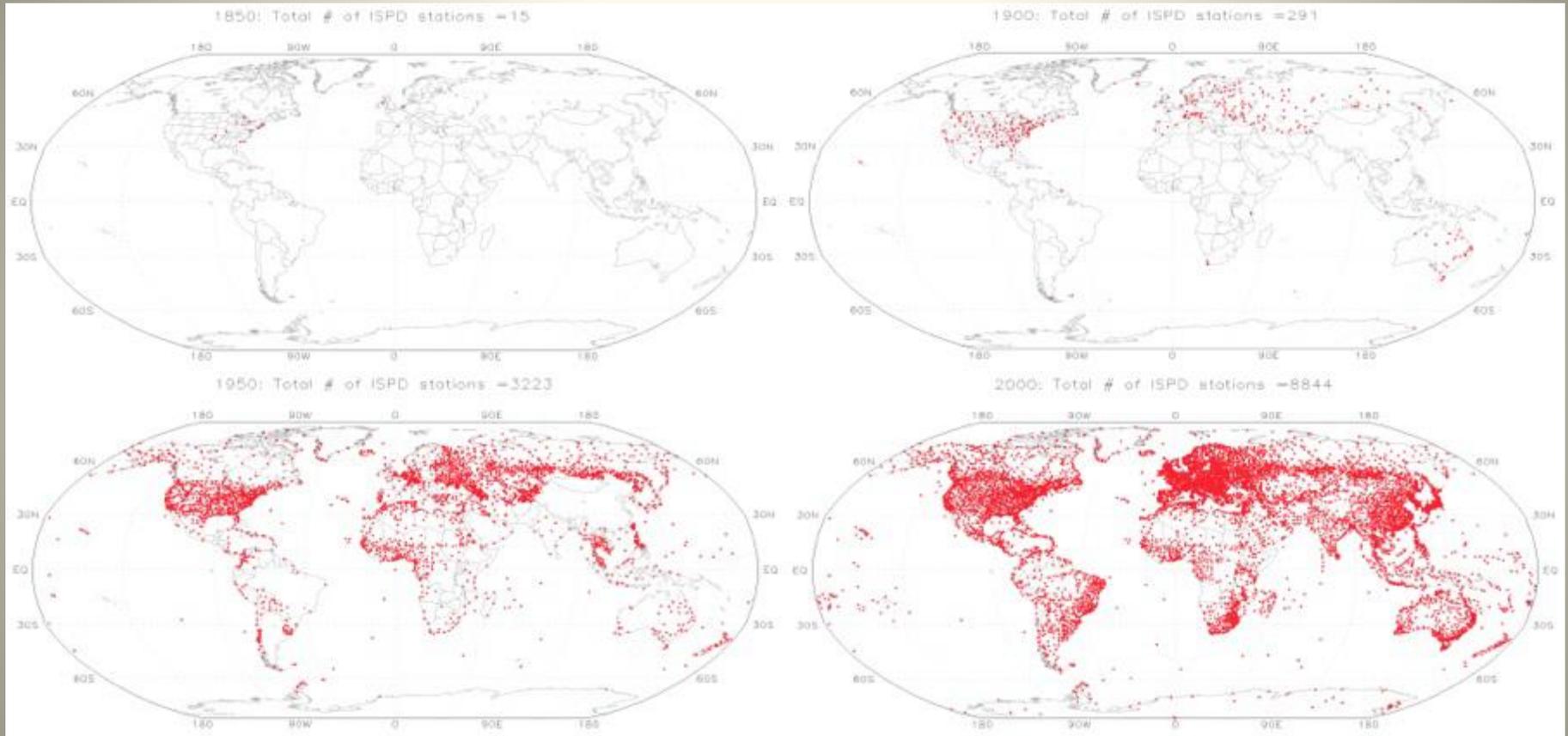
Here, there appears to be many stations, but most of these are recent series that are relatively short. So many of these are recently installed AWS data

Here the 1960s is often missing

# International Surface Pressure Databank (ISPD)

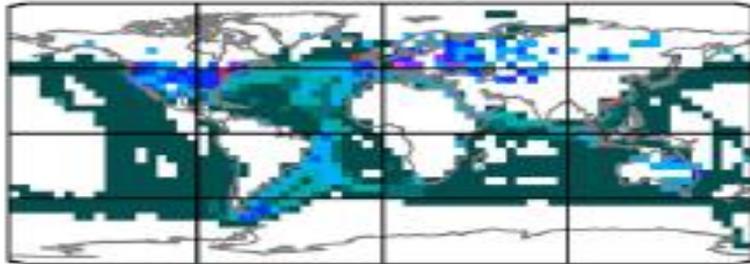
- Developed to document what is in the archive, so different Reanalyses have a numbering system for the version they have used
- Documented in a paper by Cram et al (2015)
- Able to document and show improvements in the archive
- This databank should include what is in the MARS Archive, or the US equivalent. Some plots highlight the terrestrial, others with ICOADS
- Cram, T. et al., 2015: The International Surface Pressure Databank version 2. *Geoscience Data Journal* (in press)

# ISPDv2

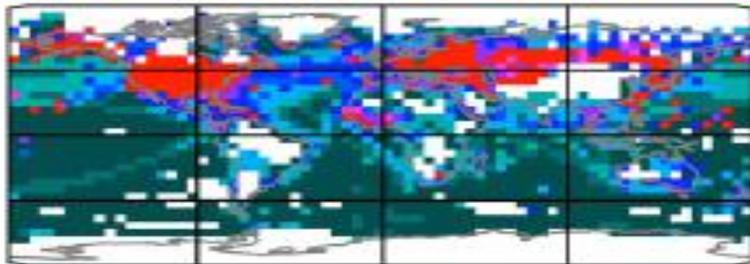


Stations in 1850, 1900, 1950 and 2000

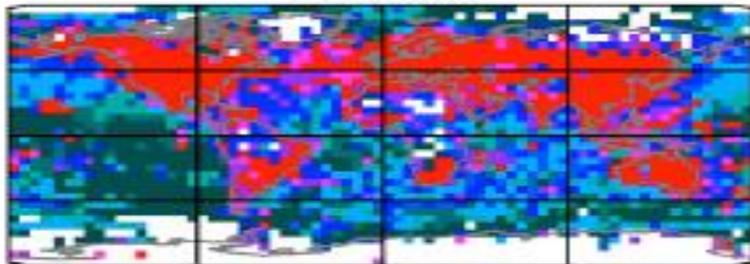
1900



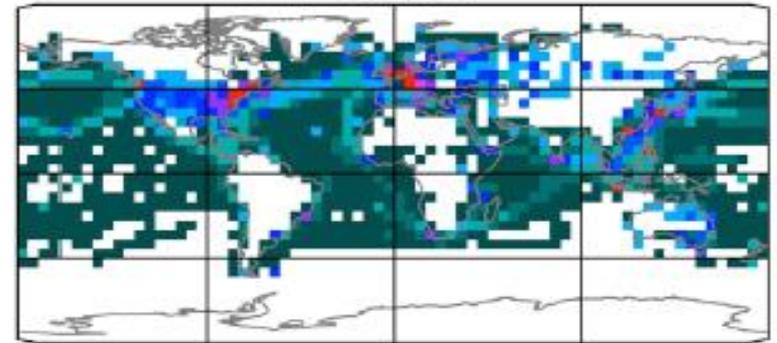
1950



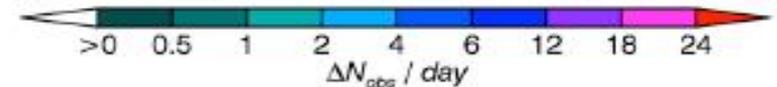
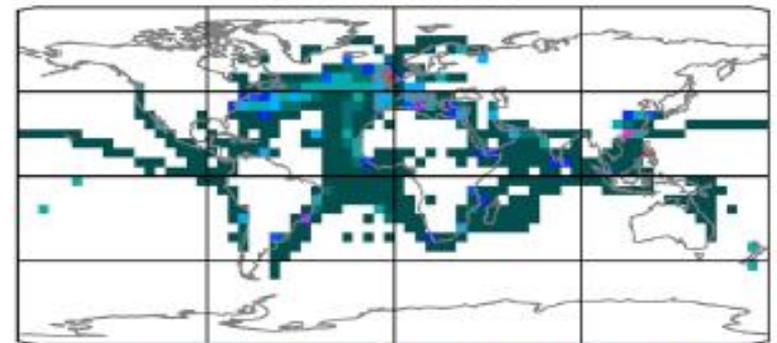
2000



1918

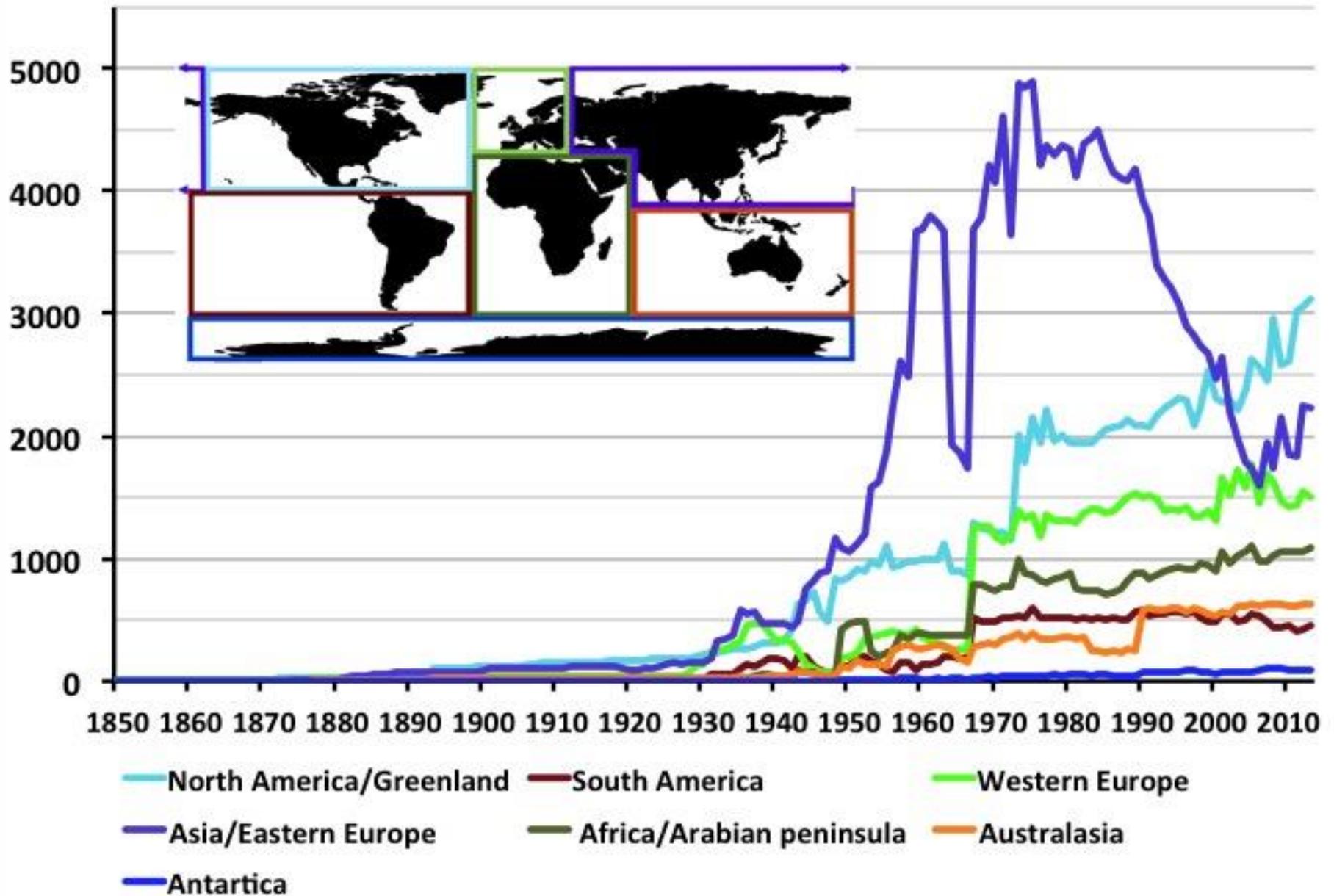


1918 V3-V2



Documenting what is in the archive (ISPD+ICOADS) and improvements

# Number of ISPD Stations by Region



# International Surface Temperature Initiative (ISTI)

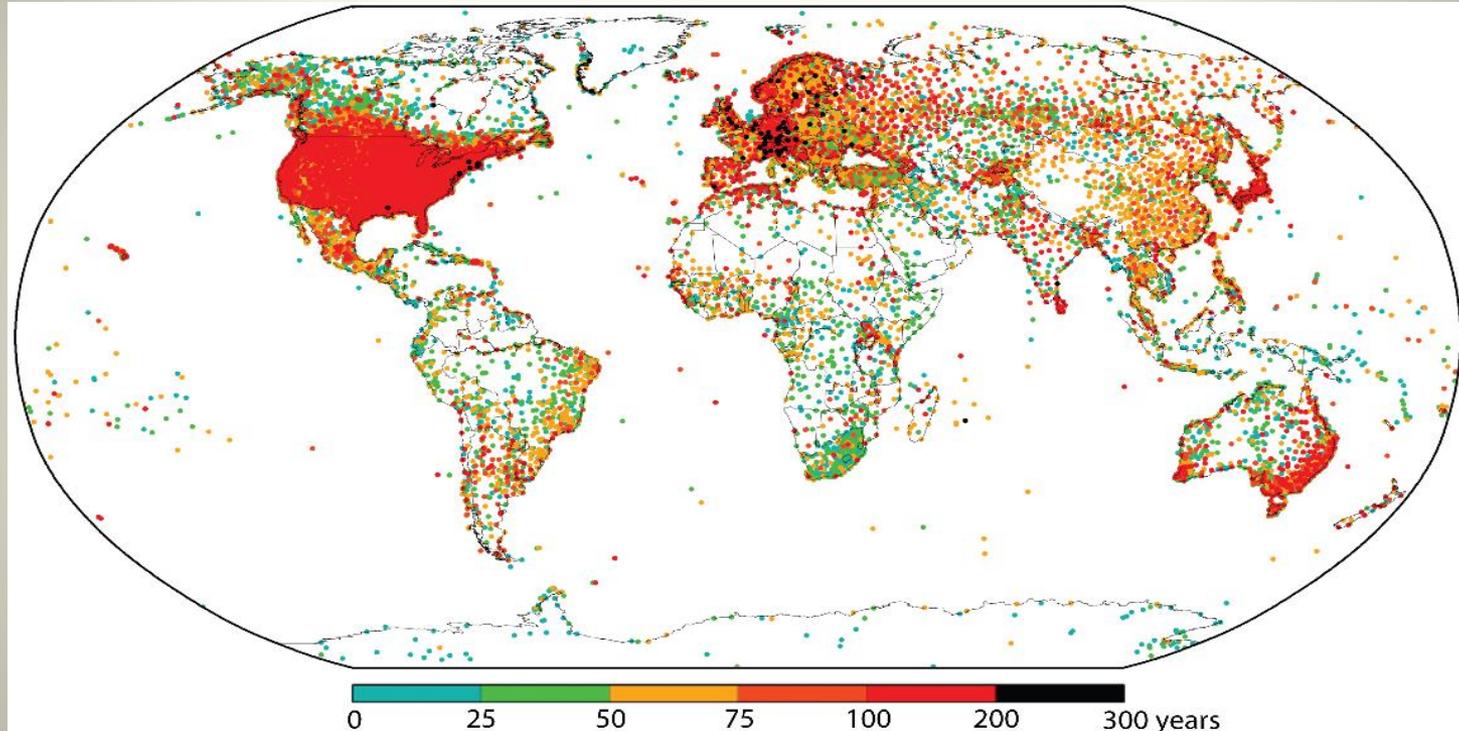
- <http://www.surface temperatures.org/>
- Set up in 2010 to co-ordinate homogeneity assessment of daily station temperature series
- Latest version of the data described by Rennie et al. (2014)
- Plot of the stations later
- ISTI also aims to homogenize the sub-daily series
- ISTI has many stations, but over the longer term is less than the new GHCNv4
- J. J. Rennie et al., 2014: The international surface temperature initiative global land surface databank: Monthly temperature data release description and methods. *Geosci. Data J.* **1**, 75–102.

# Terrestrial version of ICOADS

- This is a longer term aim (combining ISPD, ISTI and HadISDH) using NCDC's ISD), so keeping observations at the same place/time together
- ICOADS keeps all the original measurements together, for a single observation time
- Useful for checking the values, and for unit conversion
- Keeps the original units, and should allow unit conversion for humidity (dewpoint, RH, Vapour Pressure) to be more exact as this is non-linear when you move to daily and monthly. Also ensuring that temperature together with the pressure
- Would also include daily and/or sub-daily precipitation
- Keeping the variables together assists in homogeneity, as a problem with one variable is likely to affect others
- Data Rescue often restricted to a few variables

# How much data is out there?

- Monthly data possibly provide an upper limit?
- Daily and sub-daily data must have existed, but are they still available and have they been digitized and/or available?



- Monthly station length in the first release of the Global Land Meteorological Databank (Rennie et al., 2014)
- The above plot is deceptive, as it is just based on start/end years. It doesn't say which variable, or anything about completeness of the record!

# Why are data digitized?

- Most National Met Services (NMSs) have digitized most of their national data holdings, particularly since 1951 or 1961. Even though digitized, some NMSs don't make these series freely available (e.g. UK, Malaysia)
- Digitization efforts now have to have several usages and users. Difficult to resource as just being a good thing to do! Reanalysis is just one of these users. Putting recent events in a longer-timescale context is another, particularly with respect to extremes. Individual variables have specific market potential (e.g. wind/solar data). Some NMSs see this as a way to raise funds
- NMSs often have more data in hard copy form, but are occasionally unaware of its existence in their library/archive
- Also some NMSs don't realize that longer series exist for their country. Many think that the records began when the Met Service was created or the country became independent. Much resides in colonial archives
- All scanning of old data, with a view to digitization, should stipulate that the data are made available when digitized. This doesn't happen in Britain. Money found for scanning of UK Daily Weather Reports, but not their digitizing, and if it was then the digitized data would not be made freely available

# State of Historic Data Digitization

- Efforts such as ACRE (Atmospheric Circulations Reconstructions over the Earth) have co-ordinated much data digitization of both land and marine datasets
- Being part of an apparently larger product is more likely to get a national/regional project funded
- ACRE co-ordinates to avoid the same data being digitized twice, but can't keep track of all efforts around the world. ACRE is mainly Rob Allan!
- Co-ordination has been undertaken at the regional scale by the WMO/DARE Initiative over the Mediterranean Basin by EURO4M and UERRA
- The Expert Team on Data Rescue, part of WMO's Commission for Climatology (CCL) has this co-ordination as one its aims (<http://www.wmo.int/pages/prog/wcp/ccl/opace/opace1/ET-DARE-1-2.php>) and seeks to explore efforts for crowdsourcing as a data rescue strategy through an International Data Rescue Portal
- CCL also has a Task Team on Homogenization (<http://www.wmo.int/pages/prog/wcp/ccl/opace/opace2/TT-HOM-2-4.php> )
- CCL not well resourced, but most members of these two teams have long-term support
- Mostly, digitization efforts funded out of research projects

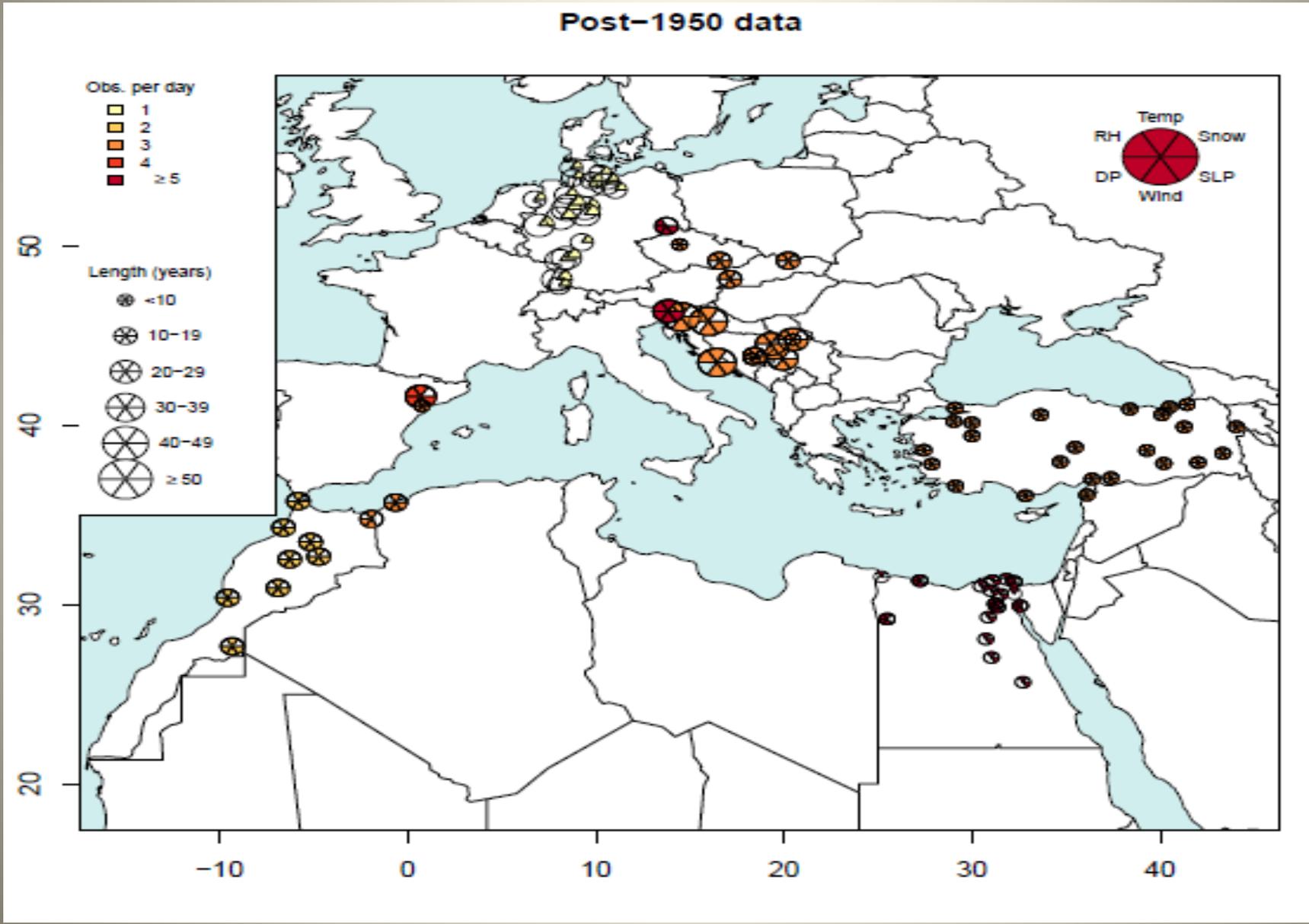
# Gridded Products (for Validation)

- Issues here are that while gridded products (not just E-OBS, but others like HadISD, HadEX2) can be used they are affected by issues of variable amounts of station data through time
- This has the greatest effect on extreme values
- For monthly products (like CRUTEM4) it is possible to allow for variable station numbers by adjusting the variance, but this is difficult for many gridded products that use Angular Distance Weights, Correlation Decay Distances and Kriging
- Many gridded products (both E-OBS and HadEX2) make use of more daily station series than they can release, so knowing how many and which stations used is not always possible

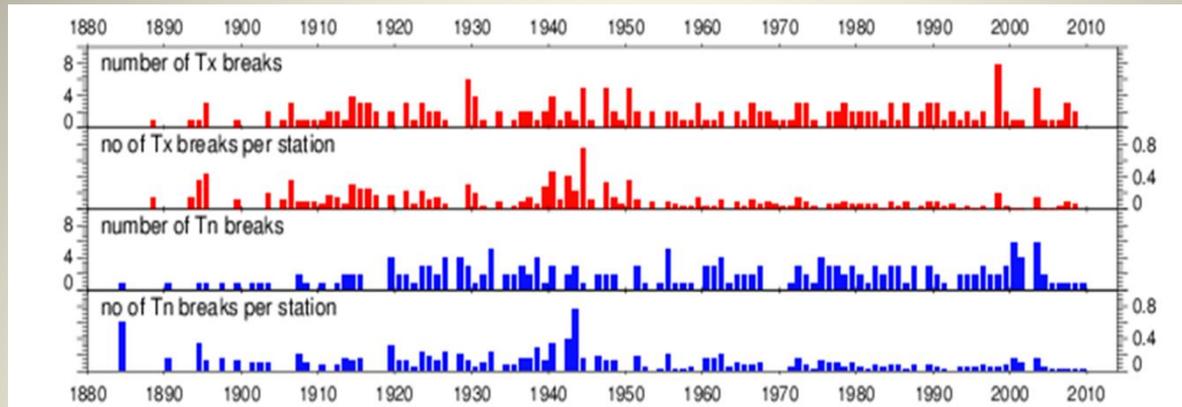
# Homogeneity

- Discussed by others
- Issue is that even after digitization it is necessary to check the series for long-term homogeneity and also for outliers and conversion from old units
- Methods exist for this and whilst the various developers want their method to be the best, but provided the method is reasonable, the results won't differ that much
- As more data are digitized homogeneity needs to be periodically reassessed
- Best done with monthly data. Adjustments can be applied to the daily and sub-daily timescales

# Digitizing efforts within UERRA –see Manola Brunet’s Poster



## Number of breakpoints found in the temperature series using ACMANT – the homogeneity checking software used



The merged Mediterranean climate dataset has been made publicly available through the ZENODO data centre ([doi:10.5281/zenodo.7531](https://doi.org/10.5281/zenodo.7531)) and provided to different databanks and repositories, such as the ECA&D portal (<http://eca.knmi.nl/>), to the CRU temperature datasets (<http://www.cru.uea.ac.uk/cru/data/temperature/>) and to the International Surface Temperature Initiative (ISTI) databank (<http://www.surface temperatures.org/databank>). In addition, the merged air pressure series have been also delivered to the International Surface Pressure Databank (<http://rda.ucar.edu/datasets/ds132.0/>), contributing to the increase of vital historical climate data availability and accessibility over the targeted Mediterranean region

Brunet, M., Gilabert, A. Jones, P.D. and Efthymiadis, D., 2014: A historical surface climate dataset from station observations in Mediterranean North Africa and Middle East areas. *Geosciences Data Journal* (in press).

Brunet, M., Jones, P.D., Jourdain, S., Efthymiadis, D., Kerrouche, M. and Boroneant, C., 2013: Data sources for rescuing the rich heritage of Mediterranean historical surface climate data. *Geoscience Data Journal*, doi: 10.1002/gdj3.4.

# Determining the Emphases

- Always said to go for the easiest (the low-hanging fruit)
- Cram et al (2015) highlight station number counts, but illustrate the large differences between continents
- As with monthly data (CRUTEM4) more data in data-sparse regions is better than more in Europe and North America, so Latin America, Africa and southern Asia
- These are the very regions where it is difficult to get access to the raw data, although international-twinning efforts and projects (e.g. CLIMANDES, PRASDES and CRC-SAS) are paving the way to enhance regional data availability and accessibility over South America, despite some countries wanting to charge for the data. Paying charges isn't the answer as these data cannot be made available to users
- With limited resources, should the focus be a few stations in many regions, or more in a few regions?
- This choice can never really be made, as users will often digitize what they can access and what variables they want

# Conclusions

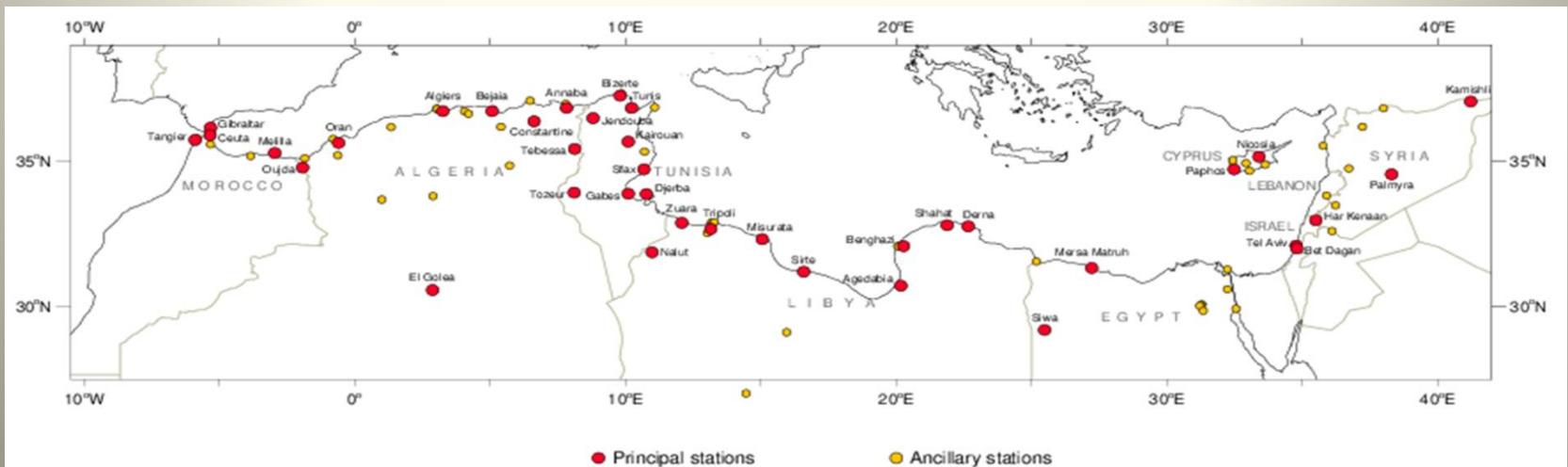
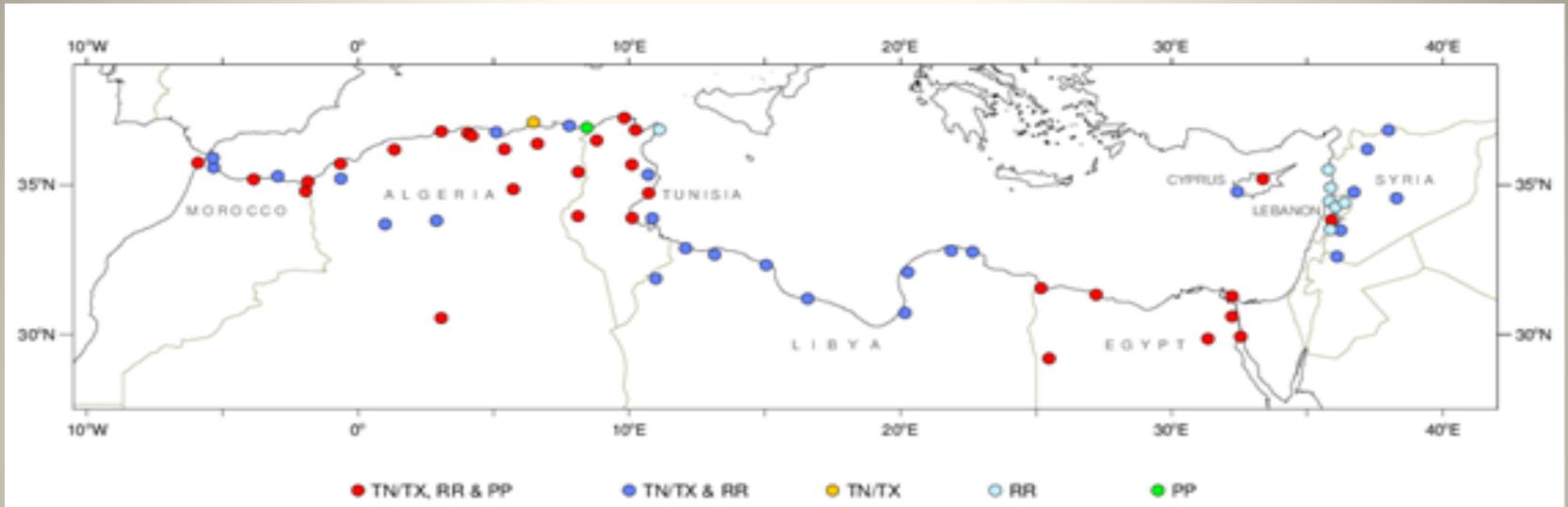
- Variety of potential sources
- Need to know what is currently in the Archive
- Need to maintain the archive, so better adding back breakpoints rather than adjusted data?
- Data digitization has a number of usages and users. Reanalysis just one of these
- Determining the focus for enhancing the Archive – should be more where there is currently sparse data, as opposed to just increasing data counts

# NMSs

- Few make their daily data freely available
- Need to continue to harass NMS' that don't – this includes many Met Services in Europe such as the UK and Spain.
- If we want to get others to release their data, we have to release ours

# URV D1.13 Final daily and sub-daily dataset for the Southern Mediterranean

Top (stations) and bottom (sites of homogenized temperature series)



# Total number of Adjustments – the whole Mars Archive

## Current Results

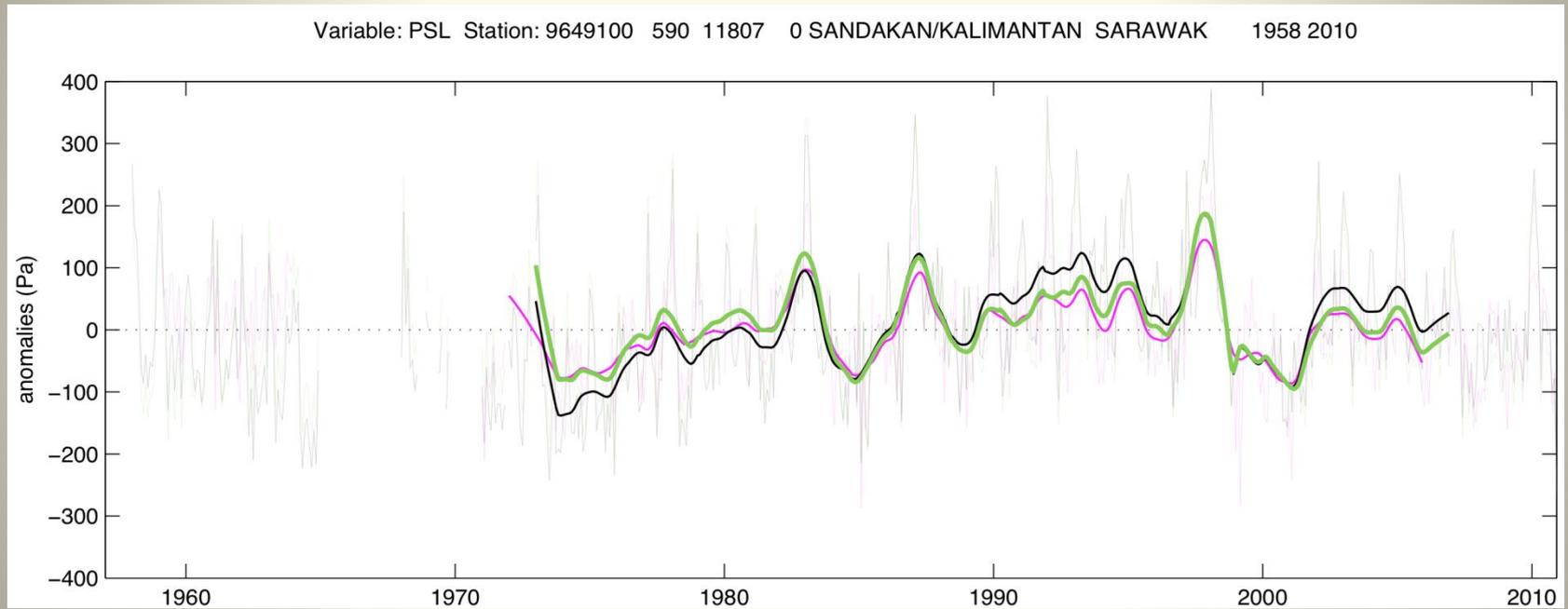
Variable	Total Stations	Total Tested	Total Adjusted	Total adjustments in each iteration			
				Iter 1	Iter 2	Iter 3	Iter 4
PSL	3053	2992	1320	2211	972	530	
RHM	3168	3075	2253	5004	2406	1281	653
WSP	3210	3119	2400	6273	3728	2345	1512
WDR	3211	3148	2070	4237	2250	1384	838
TMP	3214	3151	1004	1345	461	169	

Undertaken in the EURO4M project

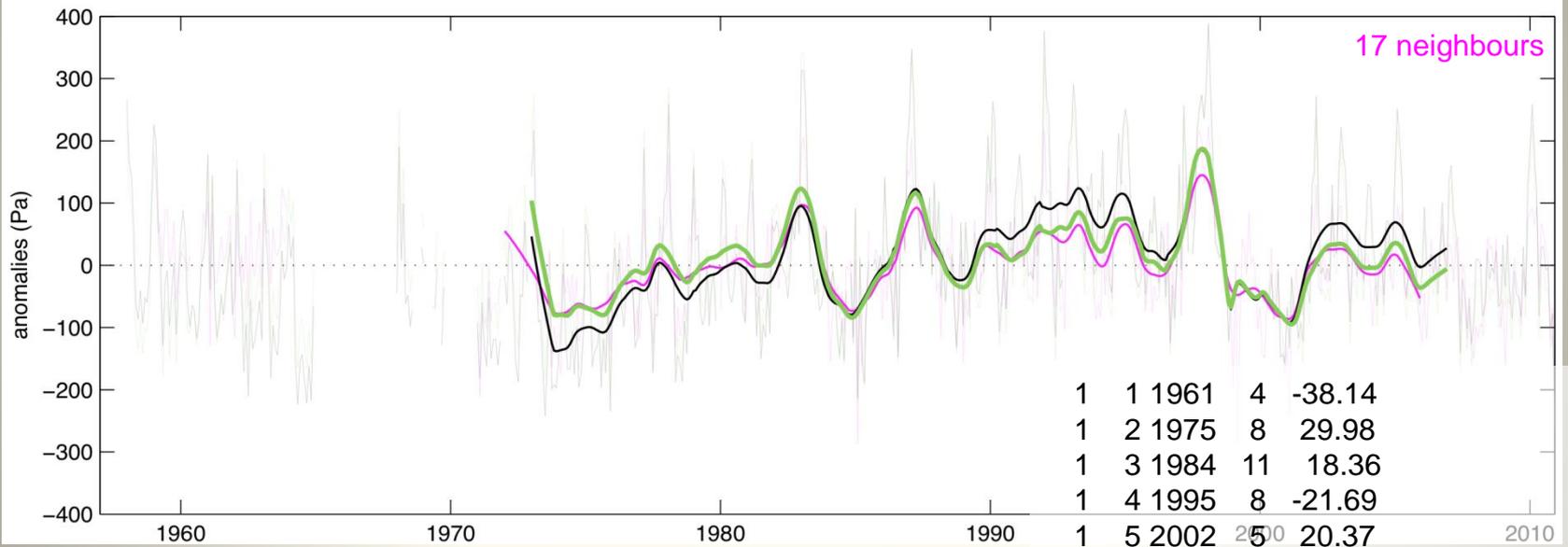
Some of this has fed back to the MARS Archive

Could just feedback the breakpoints?

# Sea Level Pressure



Variable: PSL Station: 9649100 590 11807 0 SANDAKAN/KALIMANTAN SARAWAK 1958 2010

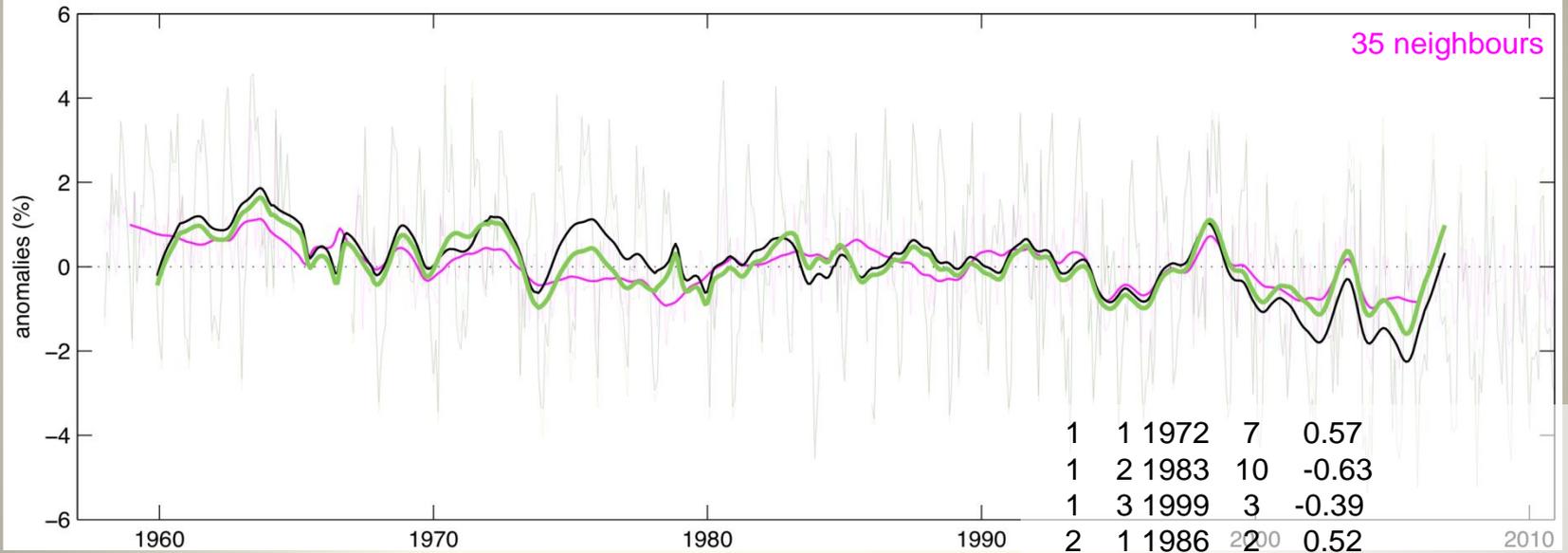


# RH

Variable: RHM Station: 4781500 3323 13162 0 OITA

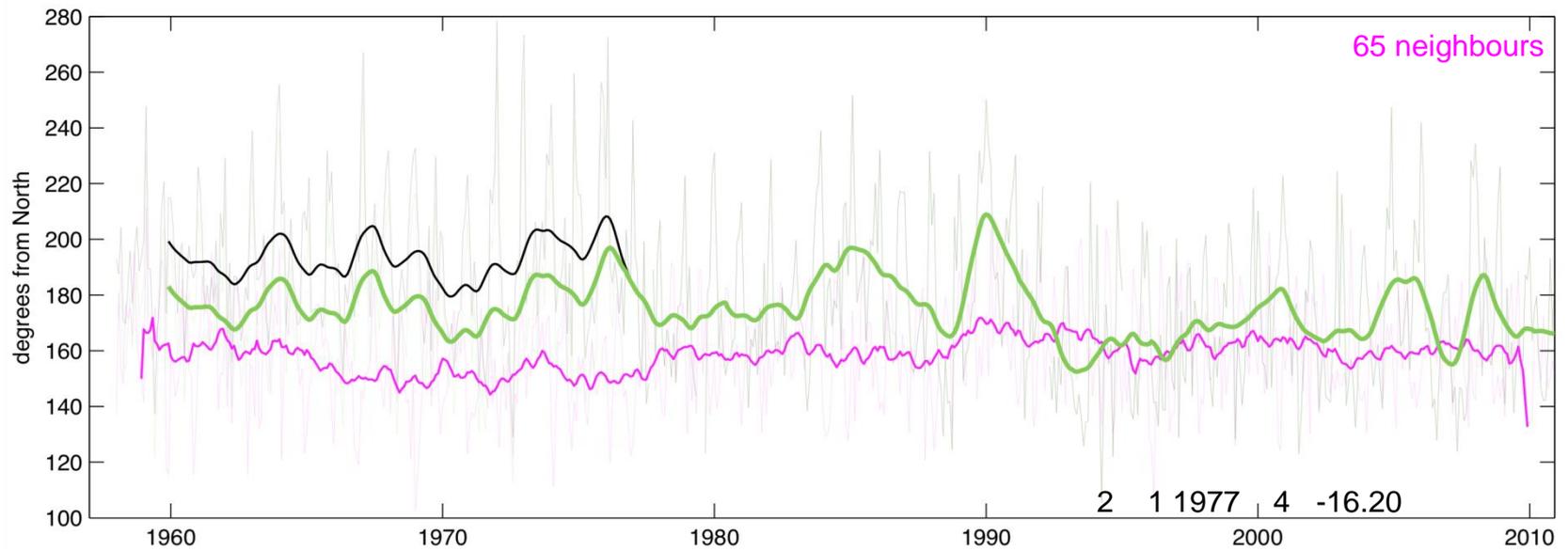
JAPAN

1958 2010





Variable: WDR Station: 3747200 4302 4743 0 MAHACKALA RUSSIA (EUR) 1958 2010



## WIND DIRECTION