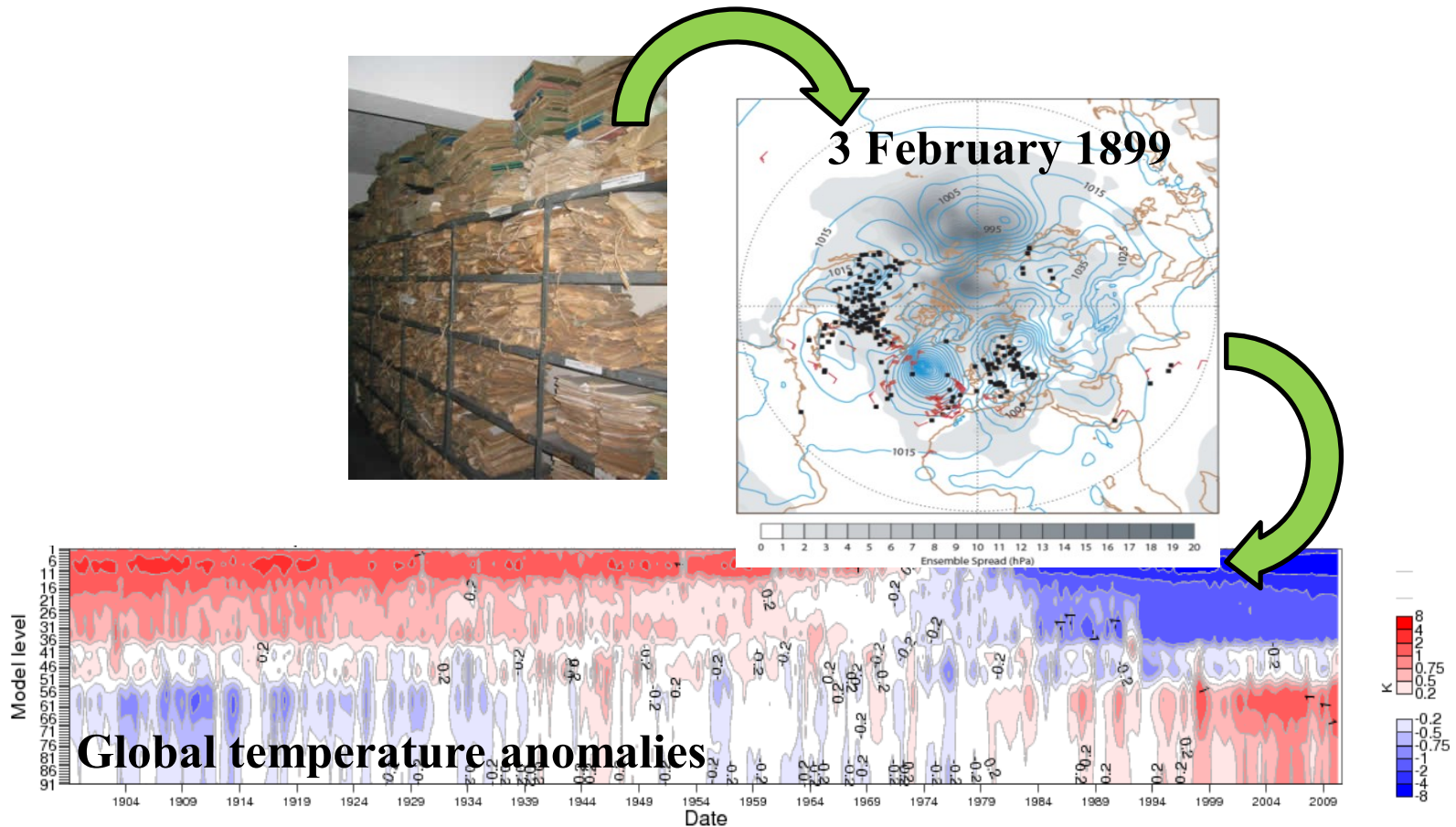


# The usage of feedback information in the ERA-20C reanalysis

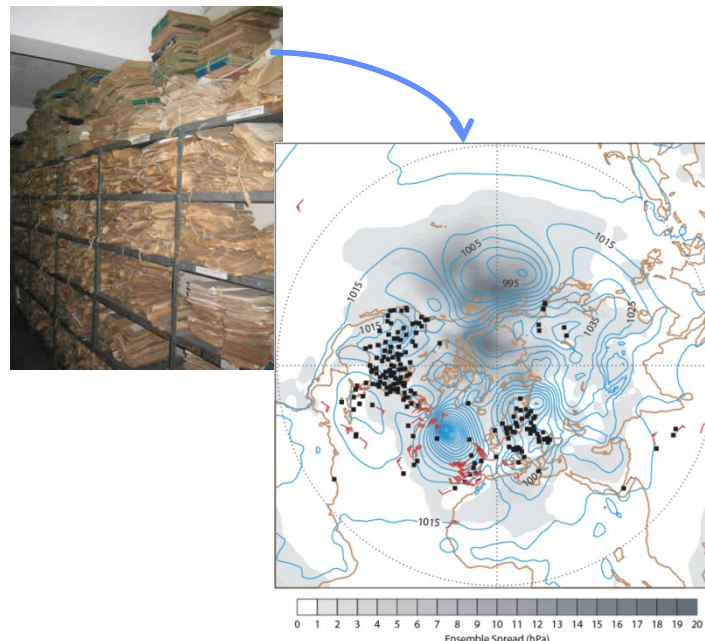
Hans Hersbach, Paul Poli, Dick Dee and reanalysis team



# The ERA-CLIM project

## EU FP7 funded

**Goal:** Preparing input observations, model data, and data assimilation systems for a global atmospheric reanalysis of the 20<sup>th</sup> century



## Main components:

1. **Data rescue** efforts (in-situ surface, upper-air)
2. **Atmospheric boundary conditions** (SST, sea ice), forcing data (solar, ozone, aerosols), satellite reprocessing
3. Conducting a number of **20<sup>th</sup> century pilot reanalyses**
4. Use of **reanalysis feedback** to improve the data record
5. Provide **access** to reanalysis data and observation quality information

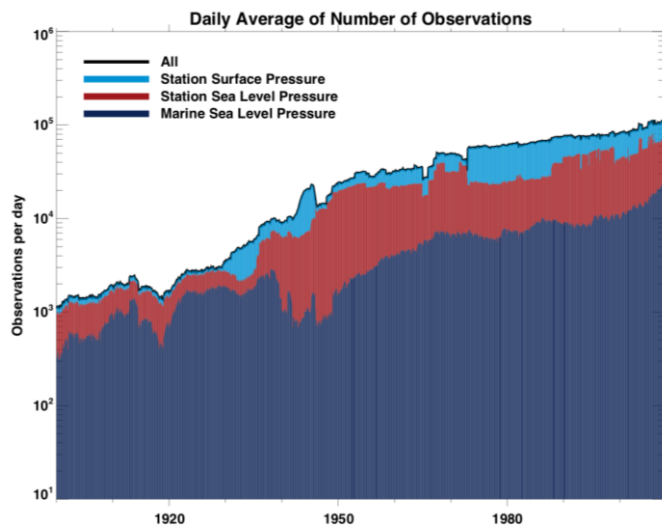
# *The ERA-20C pilot reanalysis*

Inspired on the NOAA/CIRES **20CR** reanalysis

- ✓ **1899 - 2010**
- ✓ **Usage of synoptic pressure and wind** from **ICOADS** and **ISPD** (~1.5 Billion obs)
- ✓ **Variational bias correction for pressure observations**
  - based on *break-point* analysis from **20CR** feedback information
  - minimize the dependency between reanalyses.
- ✓ CMIP5 forcing and HadISST2 SST and sea-ice forcing

# The ISPD data bank

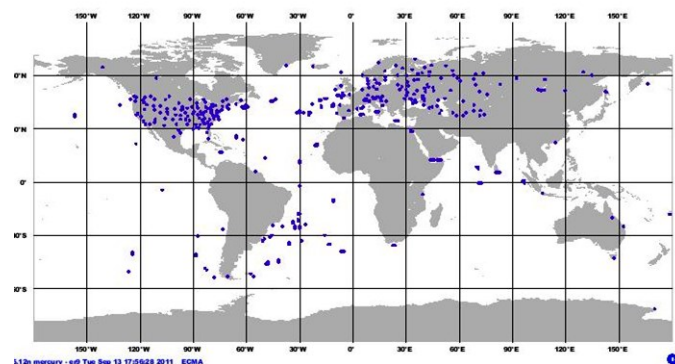
## *International Surface Pressure Data bank (Version 2.2)*



Courtesy :20<sup>th</sup> Century Reanalysis Project (NOAA/CIRES)

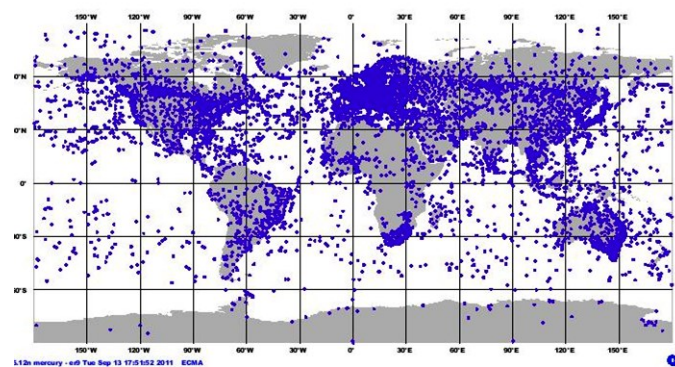
- surface pressure and MSLP
- 1.4 Billion observations from 1768 – 2008
- Many collections
- Contains feedback info from **20CR**
- *We used Version 3.2.6; created our feedback from off-line collocation 20CR*

ODB database : ECMA      Query: ispd1900010112  
No. of data points      585



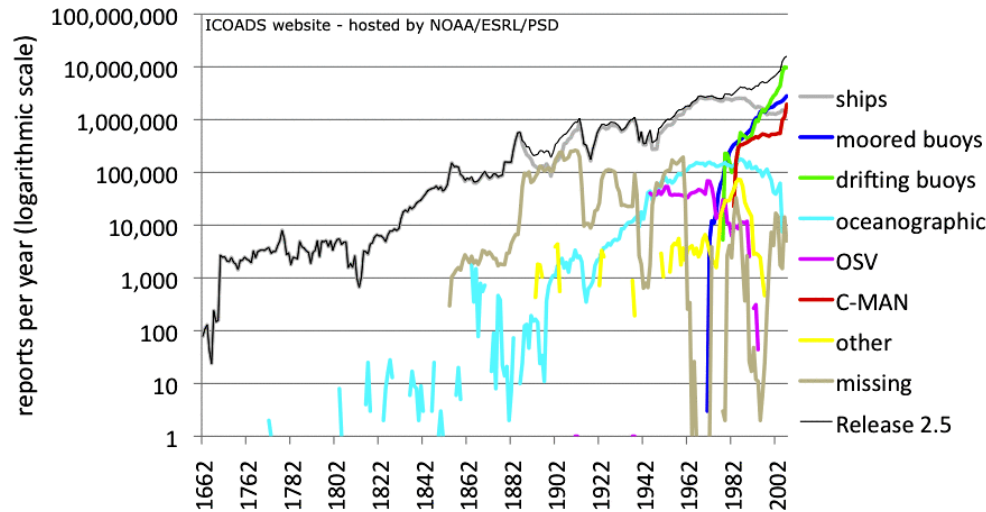
*The observation network has evolved quite a bit*

ODB database : ECMA      Query: ispd2007010112  
No. of data points      18951



# The ICOADS data set (R2.5.1)

## *International Comprehensive Ocean-Atmosphere Data Set*



Courtesy: **ICOADS**, Scott Woodruff

- 350 million reports, 1662-2011;
- MSLP, Pressure tendency, wind, T2m, WBT, Dew point T, SST, ocean wave height, period, present/past weather, visibility, clouds, ...
- MSLP already in ISPD
- Many collections (DECKS)
- No feedback information in ICOADS; but we perform an offline collocation with **20CR**, instead.

# Importing of ISPD and ICOADS into ODB-2 and MARS

**From native format (hdf5, IMMA) using dedicated programs.**

Mapping of information is the most non-trivial part.

**Introduction of some new ODB-2 columns to facilitate stratification.**

- **Source:** ISPD 3.2.6, ICOADS 2.5.1
- **Collection identifier:** each source assembled data records from various origins. For example, collection number 761 in ICOADS 2.5.1 contains “Japanese Whaling Ship Data (CDMP/MIT digitization)”
- **Unique identifier:** each source has its own way of defining this variable. It is usually a number which needs to be used along with something else (observation date and time for example)
- **Timeseries\_index:** added by us, allows later to pull out easily a single station time-series, across the entire observation feedback archive

# Need for platform tracking

**In situ network: bias depends on history of a specific platform/station.**

- ISPD and ICOADS contain more than **1 Million** platforms (mostly SHIPS)
- distinguished by *station id* or *call sign*:

The top 4 consists of **ICOADS** data for which no information on call sign is available

statid	collection_id	Nobs	First date, lat, lon				Last date, lat, lon			
????????	000104	24,227,426	19520101	219900	57.90	345.80	20021216	129900	11.50	81.49
????????	000105	9,857,296	18000204	169900	-51.50	302.50	19741211	099900	59.60	1.50
SHIP	000104	3,389,093	19700101	009900	71.20	35.10	20061231	189900	-35.50	138.20
MASKSTID	000105	1,331,515	20071201	009900	83.00	97.60	20081231	239900	60.60	3.70

**Need to find a way to split such platforms/stations: timeseries index**

- ✓ For each call signs, start with one subgroup
- ✓ Create new subgroup if an observation had to move by more than 50 m/s
- ✓ However, allow for minimal distance of 200 km.



# Variational BC using break-point analysis

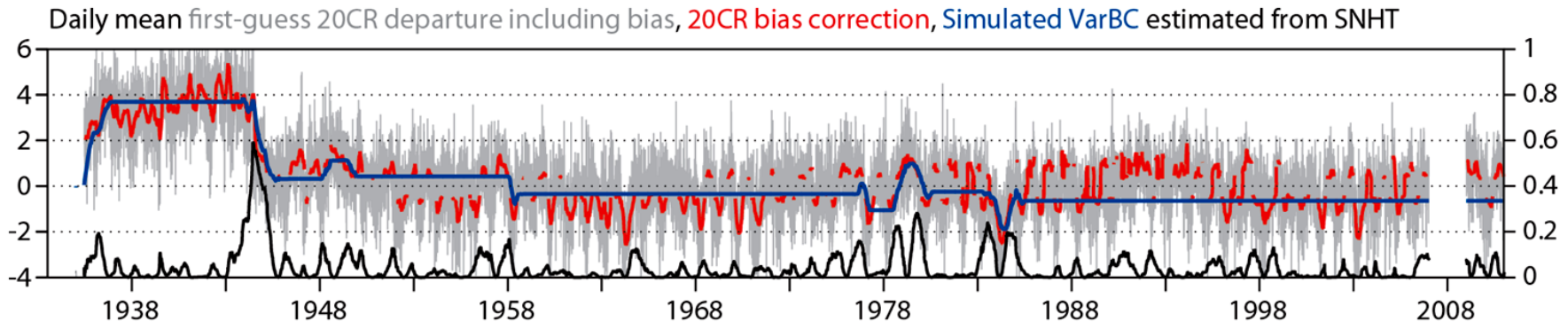
## Perform VarBC on platform level

- Has to be rigid; otherwise the signal is taken away from the observations

## Build on feedback information from 20CR (ISPD and ICOADS)

- detect break points in advance, create quantity: *bias volatility* (=normalized SNHT)

Time Series Index: 000004, Stat id: 723060HU, Surface Pressure Ps (hPa), all data Manual Land SYNOP,  
01000: Global Land Surface Observations (Federal Climate Complex Integrated Surface Database) 1901-2008



## Standard Normal Homogeneity Test (SNHT)

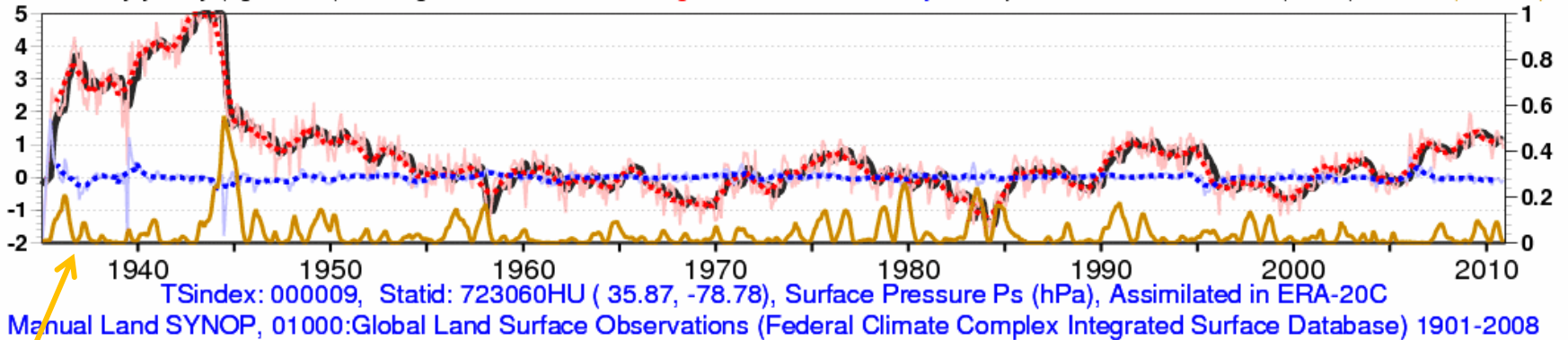
- Used e.g., by *Haimberger (2005)* for homogenization of radiosonde data
- Expresses the difference in long-term average departures *before* and *after* an observation.
- Let the response time of **VarBC** depend on the history of the *bias volatility*



# Applied VarBC in ERA-20C

## Raleigh-Durham International Airport (NC)

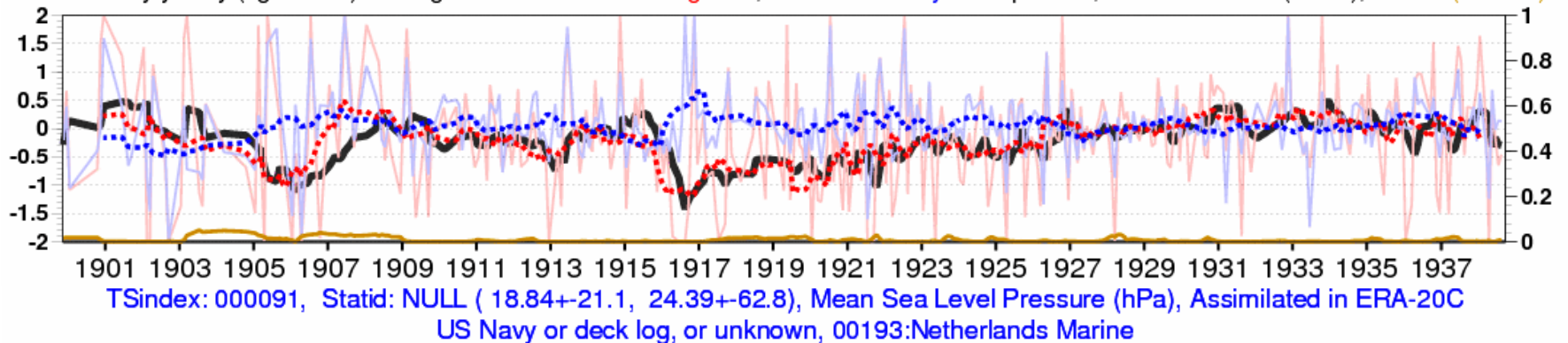
Monthly/yearly (light/dark) average of **uncorrected first-guess**, **corrected analysis** departure, bias correction (black), **SNHT** (bottom)



From 20CR

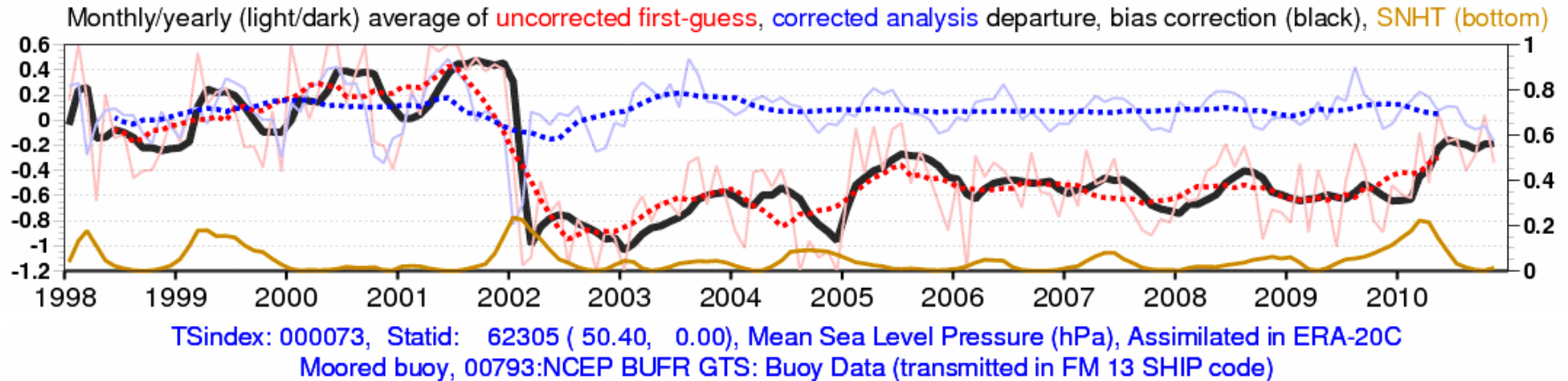
## Unknown SHIP

Monthly/yearly (light/dark) average of **uncorrected first-guess**, **corrected analysis** departure, bias correction (black), **SNHT** (bottom)

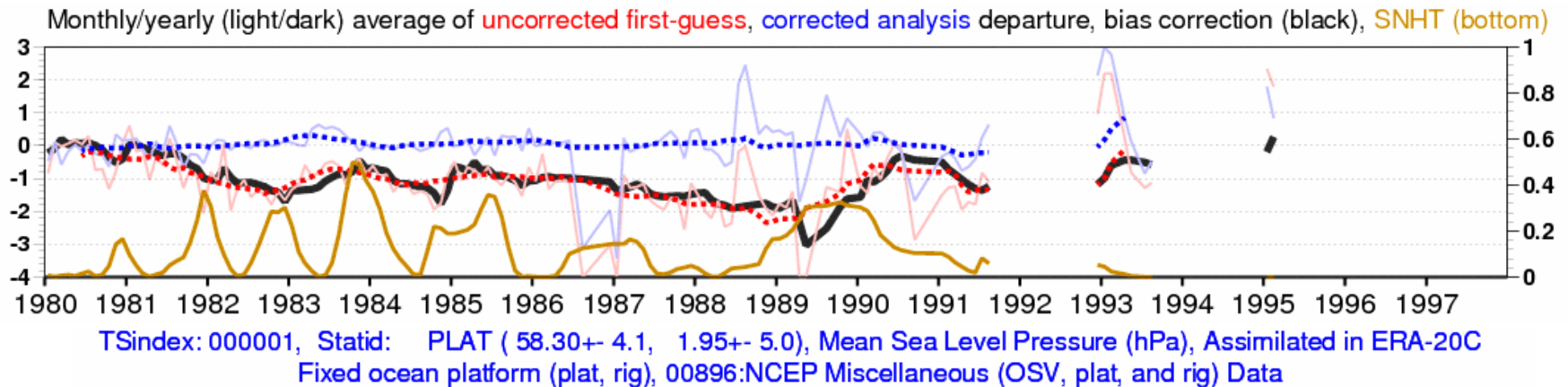


# Applied VarBC in ERA-20C

## Moored Buoy



## Platforms and Rigs

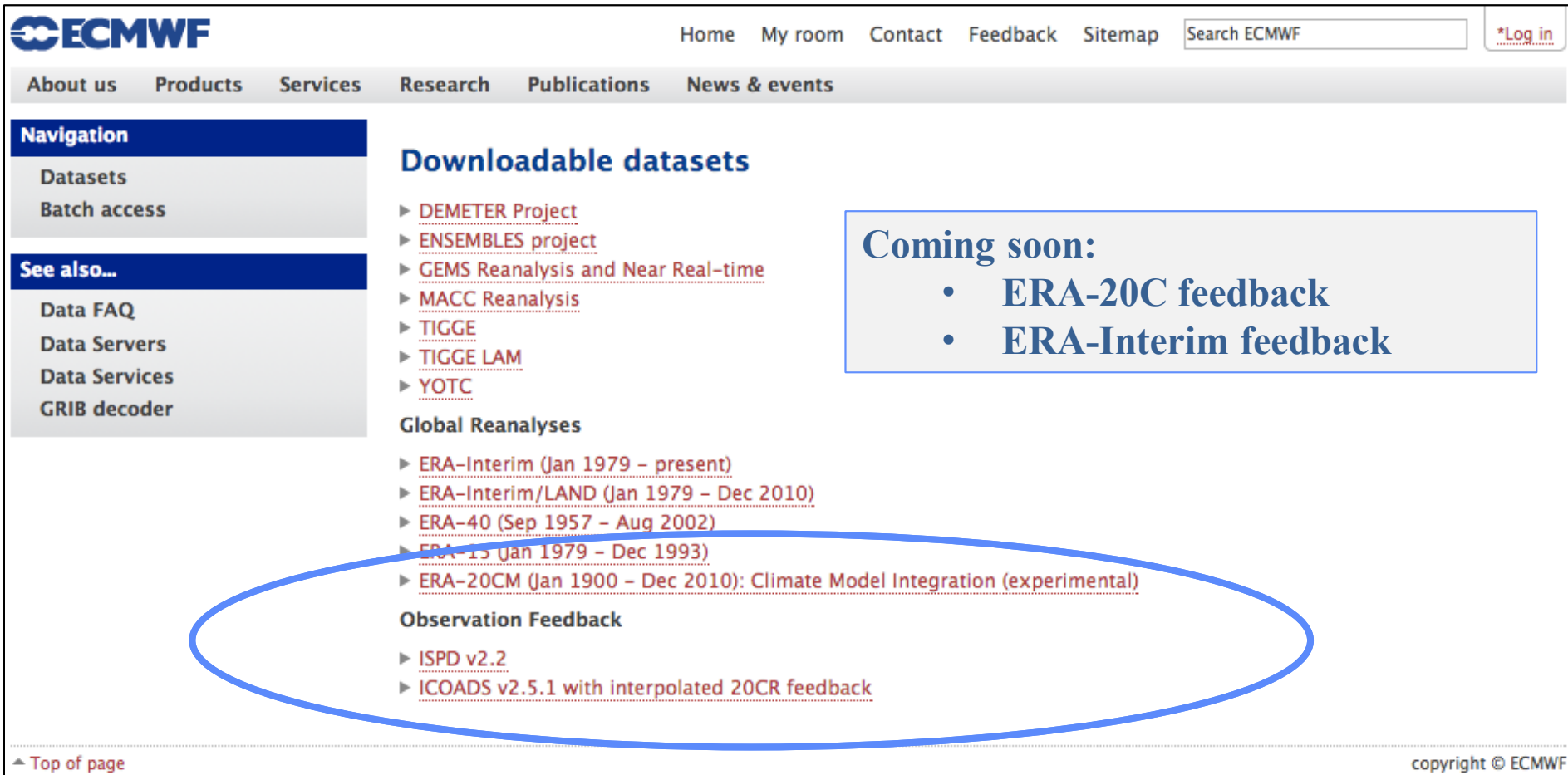


# *Access to reanalysis data and observation quality information*

## **Observation feedback archive (OFA)**

- Based on data format used at ECMWF: **ODB**
  - ✓ One row per observation, not per report
  - ✓ Is, therefore, flexible
- On ECMWF data server
- Contains *valuable feedback* information:  
Quality control, model departures, bias estimates, traceability,...

Data server at <http://apps.ecmwf.int/datasets/>



The screenshot shows the ECMWF Data server website. The header includes the ECMWF logo, navigation links (Home, My room, Contact, Feedback, Sitemap), a search bar, and a login link. A secondary navigation bar lists categories like About us, Products, Services, Research, Publications, and News & events. On the left, a 'Navigation' sidebar contains links to Datasets and Batch access, and a 'See also...' section with links to Data FAQ, Data Servers, Data Services, and GRIB decoder. The main content area is titled 'Downloadable datasets' and lists several projects: DEMETER Project, ENSEMBLES project, GEMS Reanalysis and Near Real-time, MACC Reanalysis, TIGGE, TIGGE LAM, and YOTC. Below this is a section for 'Global Reanalyses' listing ERA-Interim (Jan 1979 - present), ERA-Interim/LAND (Jan 1979 - Dec 2010), ERA-40 (Sep 1957 - Aug 2002), ERA-15 (Jan 1979 - Dec 1993), and ERA-20CM (Jan 1900 - Dec 2010): Climate Model Integration (experimental). The 'Observation Feedback' section lists ISPD v2.2 and ICOADS v2.5.1 with interpolated 20CR feedback. A blue oval highlights the ERA-20CM and Observation Feedback sections. A 'Coming soon:' box on the right lists ERA-20C feedback and ERA-Interim feedback. The footer includes a 'Top of page' link and a copyright notice for ECMWF.

ECMWF

Home My room Contact Feedback Sitemap Search ECMWF \*Log in

About us Products Services Research Publications News & events

**Navigation**

- Datasets
- Batch access

**See also...**

- Data FAQ
- Data Servers
- Data Services
- GRIB decoder

**Downloadable datasets**

- ▶ [DEMETER Project](#)
- ▶ [ENSEMBLES project](#)
- ▶ [GEMS Reanalysis and Near Real-time](#)
- ▶ [MACC Reanalysis](#)
- ▶ [TIGGE](#)
- ▶ [TIGGE LAM](#)
- ▶ [YOTC](#)

**Global Reanalyses**

- ▶ [ERA-Interim \(Jan 1979 - present\)](#)
- ▶ [ERA-Interim/LAND \(Jan 1979 - Dec 2010\)](#)
- ▶ [ERA-40 \(Sep 1957 - Aug 2002\)](#)
- ▶ [ERA-15 \(Jan 1979 - Dec 1993\)](#)
- ▶ [ERA-20CM \(Jan 1900 - Dec 2010\): Climate Model Integration \(experimental\)](#)

**Observation Feedback**

- ▶ [ISPD v2.2](#)
- ▶ [ICOADS v2.5.1 with interpolated 20CR feedback](#)

**Coming soon:**

- ERA-20C feedback
- ERA-Interim feedback

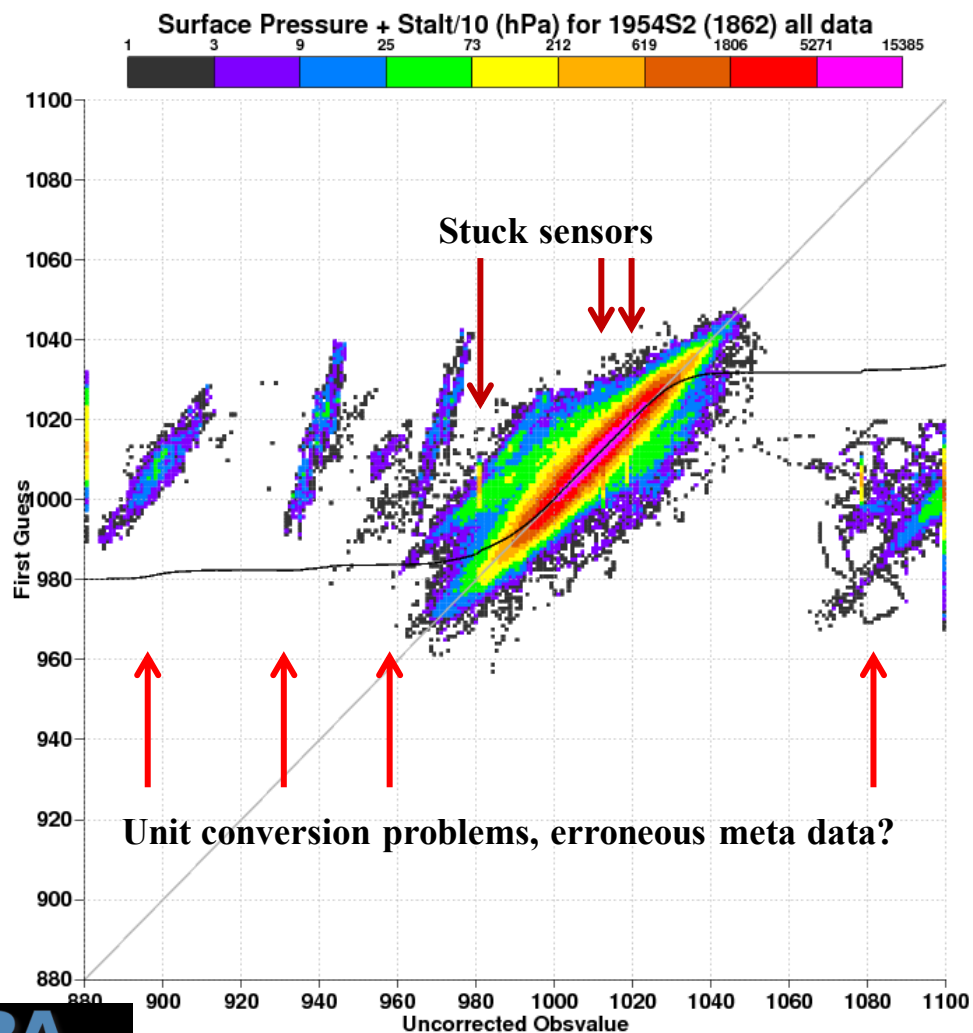
▲ Top of page

copyright © ECMWF



# Feedback: Quality control

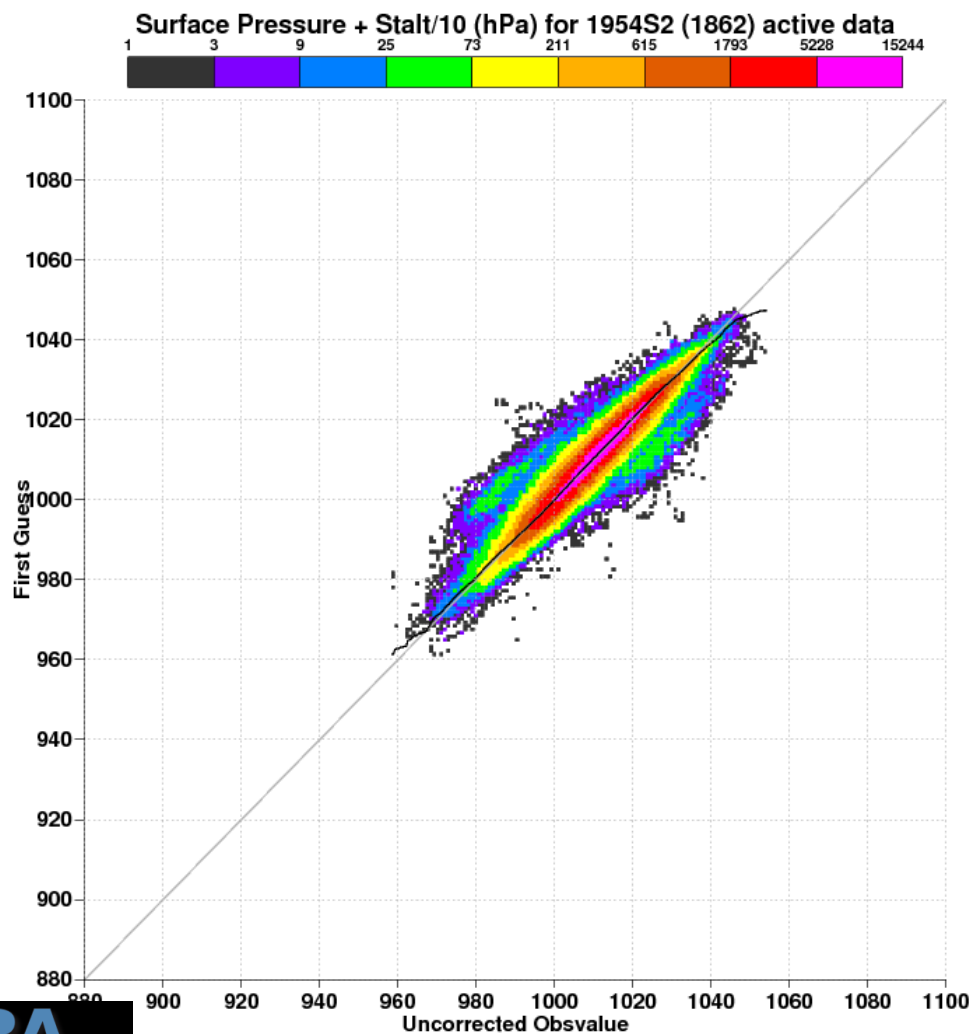
ERA-20C *all data*



Number of collocations: 1989418  
Y-X Mean, stdv: -0.003 15.675  
X Mean, stdv: 1010.623 17.344  
Y Mean, stdv: 1010.620 9.379

# Feedback: Quality control

## ERA-20C *after screening*



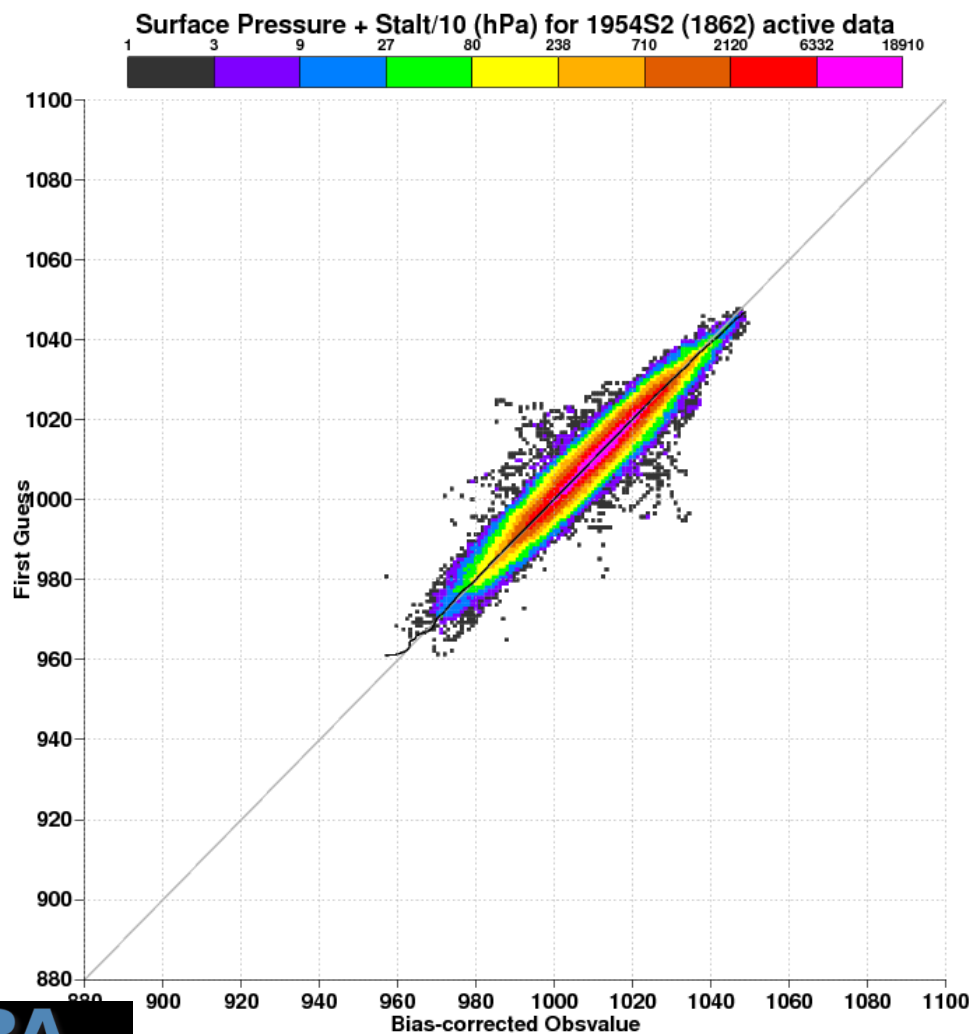
Number of collocations: 1910905  
Y-X Mean, stdv: 0.261 2.652  
X Mean, stdv: 1010.466 9.303  
Y Mean, stdv: 1010.728 9.364





# Feedback: Quality control

## ERA-20C *Bias Correction*

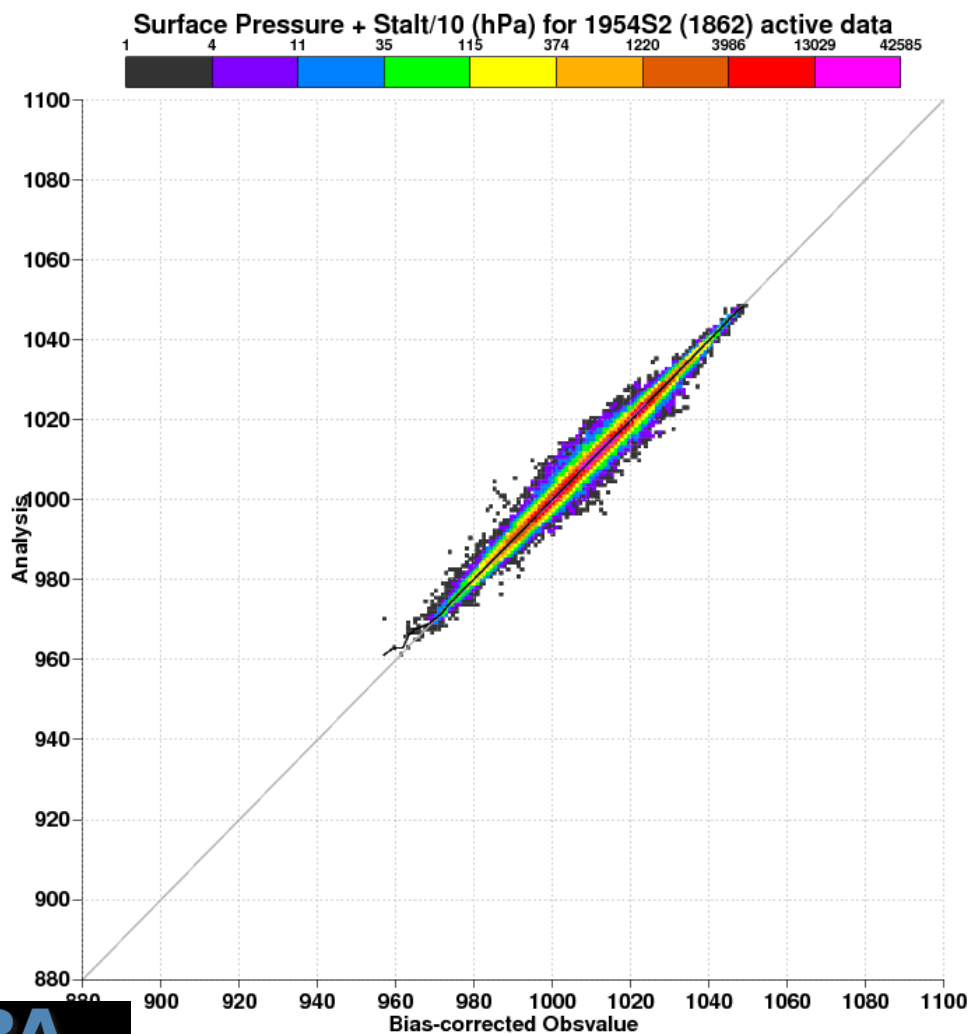


Number of collocations: 1910905  
Y-X Mean, stdv: 0.172 2.001  
X Mean, stdv: 1010.556 9.388  
Y Mean, stdv: 1010.728 9.364



# Feedback: Quality control

ERA-20C *after assimilation*

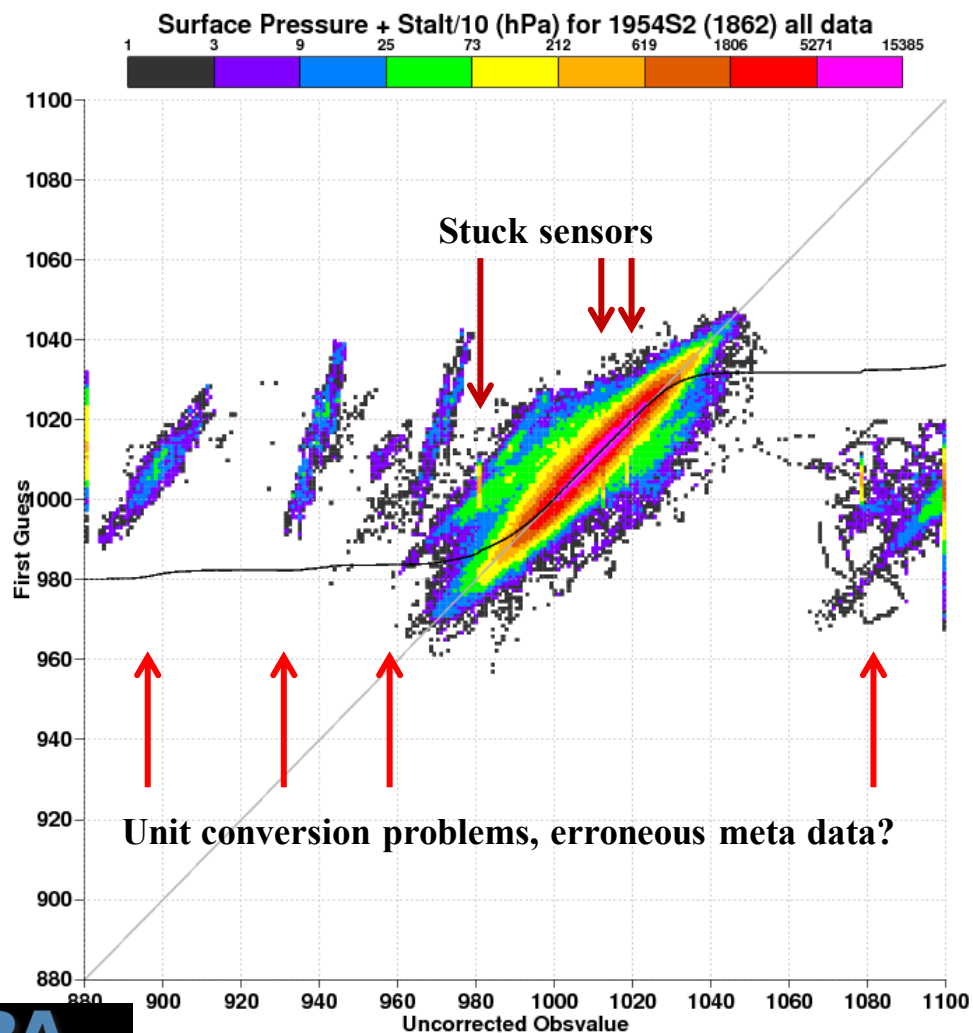


Number of collocations: 1910905  
Y-X Mean, stdv: -0.008 0.752  
X Mean, stdv: 1010.556 9.388  
Y Mean, stdv: 1010.547 9.329



# Feedback: Quality control

ERA-20C *all data*

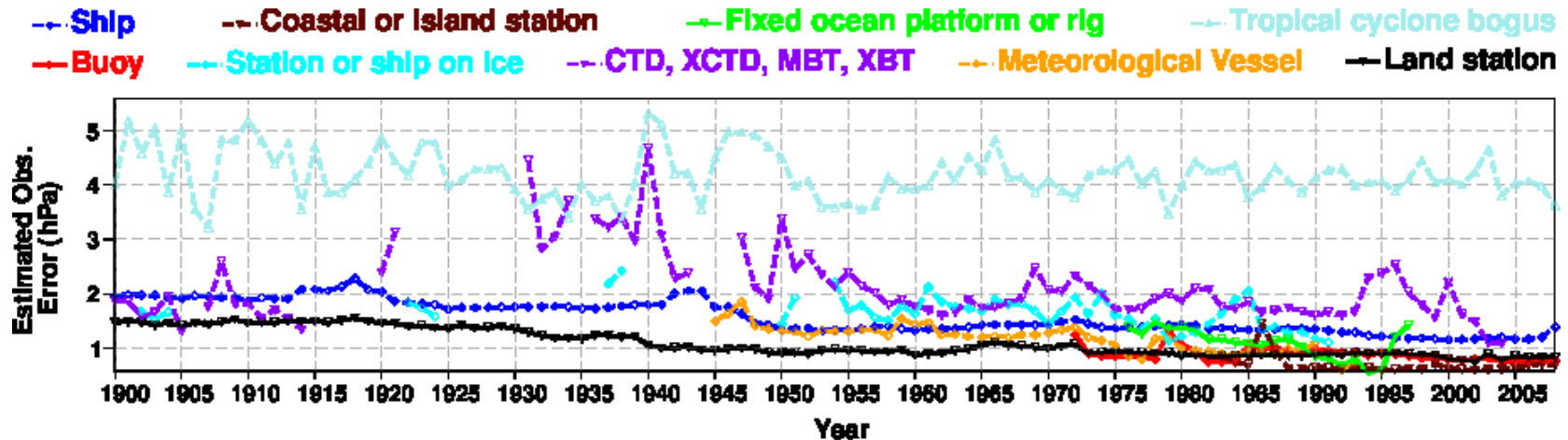


Number of collocations: 1989418  
Y-X Mean, stdv: -0.003 15.675  
X Mean, stdv: 1010.623 17.344  
Y Mean, stdv: 1010.620 9.379

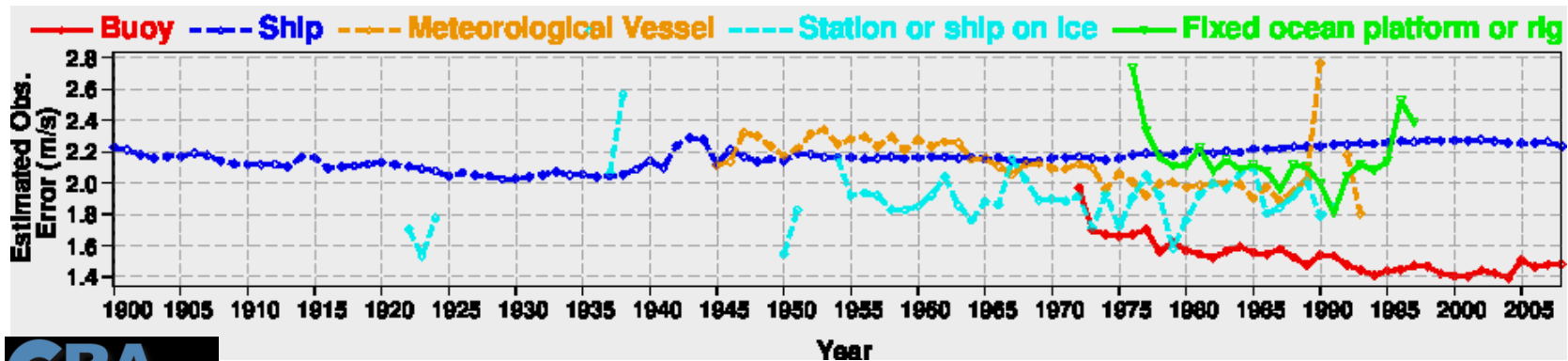


# Feedback: estimates on observation error (*Desrosier et. al*) from (model-obs) departure statistics

## Pressure



## Wind



# Final remarks

Method of VarBC using Break point analysis appears rather robust

- the dependency on previous reanalysis (20CR) is minimized

Feedback information contains important information on the observations:

- Detection of erroneous data or meta data
- Bias and random error estimates

During ERA-CLIM some odb-2 (diagnostic) tools have been created:

- command-line utility for **density plot** between any columns from any ODB-2 file (as shown in this presentation)
- Command-line utility for **animations** from any set of ODB-2 files