Data Acquisition and Processing

Eduardo B. M. Barbosa, Bianca A. Souza,
Luciana S. M. Carvalho, Luciana M. C. Mira

Instituto Nacional de Pesquisas Espaciais - Inpe
Rod. Presidente Dutra, Km. 40 - Cach. Paulista, SP - 12630-000
{eduardo.barbosa, bianca.antunes, luciana.carvalho, luciana.mira}@inpe.br

Introduction

- **Introduction**
- **Data Flow**
  - Data Acquistion
- **Pre-processing**
  - Screeening
  - Statistics
  - Quality Control
  - Data Coverage
- **Database**
  - Overview
  - Quality Control

Introduction

GTS

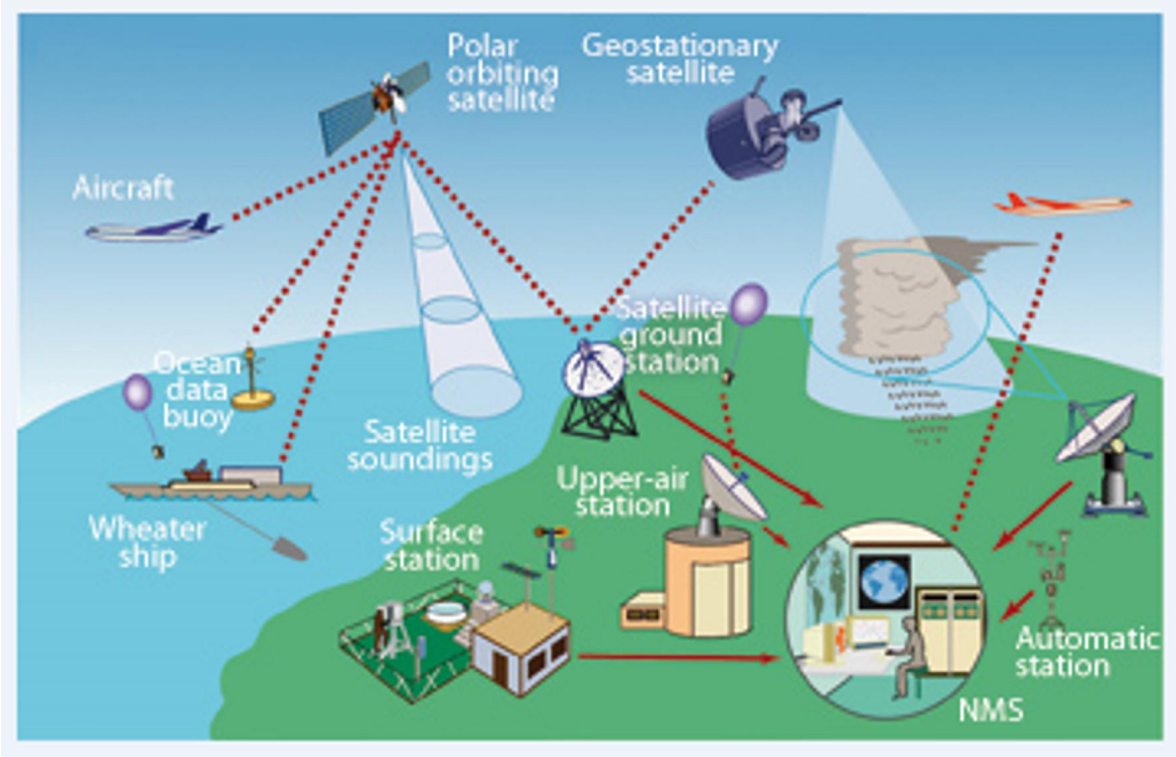


Fig. 1. The GTS system.

- Direct measurements (*in situ*)
- Visual observations
- Indirect measurements (sensors)

Data Flow

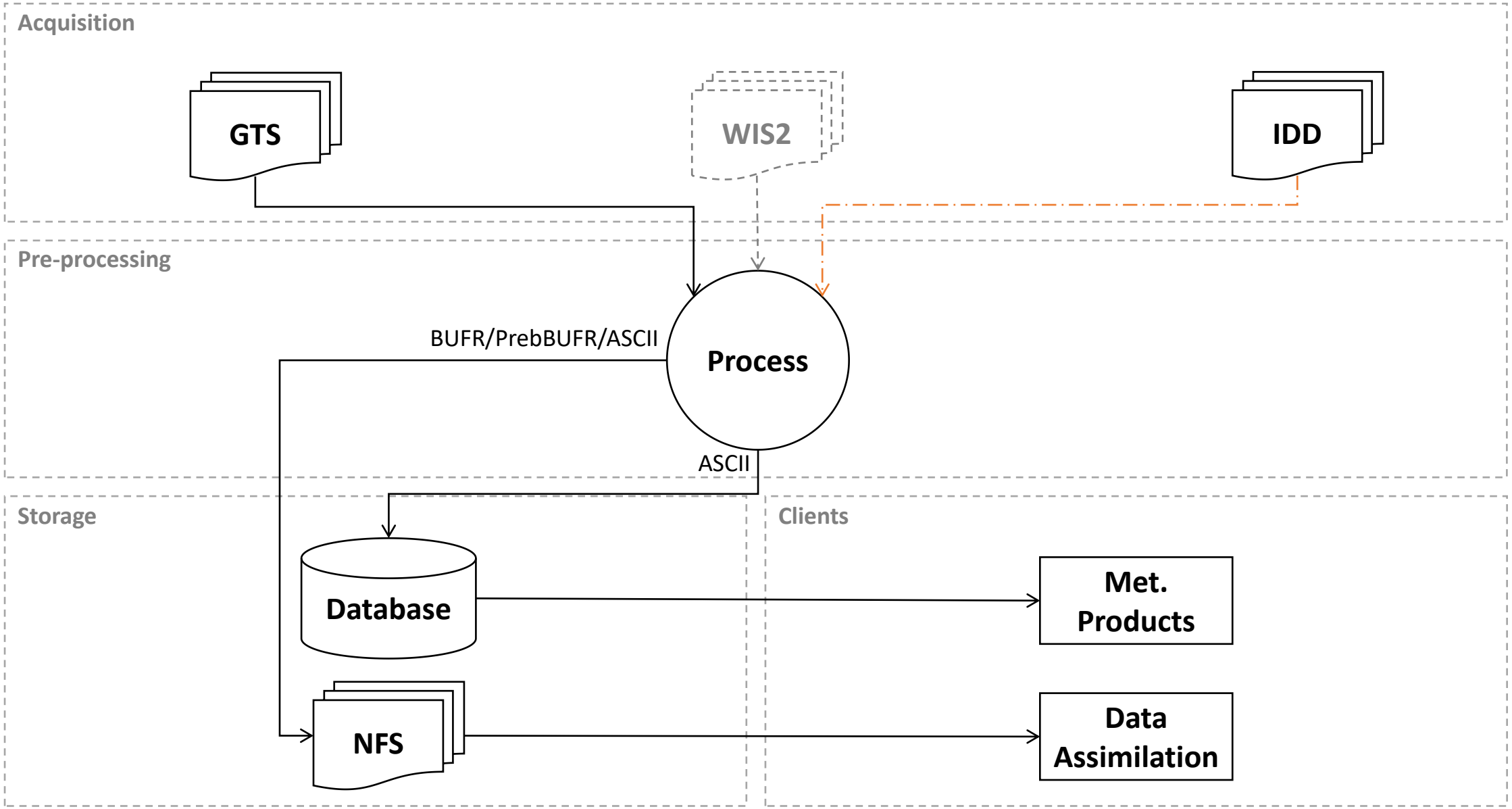


Fig. 2. Data flow.

Data Flow

- Data acquisition
  - ecFlow suite
  - Shell scripting (transfer)

```
GET
     URL    = http://nomads.ncep.noaa.gov
     DIR    = /pub/data/nccf/com/obsproc/prod/gdas.<YYYY><mm><dd>
     FNAME  = gdas.t<HH>z.prepbufr.nr
     DATE   = -1:0
     OUTPUT = /data/dboper/da
     OPTS   = parallel
```

Fig. 3. Parameter file used in the data acquisition.

- Other built in house scripts to the data acquisiton in different formats (e.g.: json, xml, etc.)

Pre-processing

Screening

```
SMBZ06 SBBR 151200
AAXX 15124
82022 32970 00708 10276 20222 30021 40117 333 20246 59011 60004=
82193 22570 50611 10282 20234 30104 40124 85200 333 20245 59012 69924=
82281 32570 40717 10291 20210 30054 40116 84200 333 20263 5//// 60004=
82332 32570 50803 10330 20247 30019 40111 81000 333 20241 59016 60004=
...
```

```
SABZ30 SBBR 151200
METAR SBAT 151200Z 24002KT 9999 FEW012 29/24 Q1013=
METAR SBBE 151200Z 06011KT 9999 BKN025 28/23 Q1012=
METAR SBBV 151200Z 07008KT CAVOK 28/22 Q1012=
METAR SBCC 151200Z 04005KT CAVOK 28/21 Q1014=
METAR SBCJ 151200Z 02006KT 9000 SCT010 24/20 Q1016=
...
```

Fig. 4. Example of meteorological messages in (old) TAC format.

- Abbreviated headings (T_1T_2 A_1A_2 II CCCC YYGGcc)
  - Shell scripting
  - ecCodes

Pre-processing

Screening

```
$ gtsinsp BIN511_20251115123927_00547253
#       Rec TT AA II CCCC ddHHMM NNN
          1 IN QI 07 SABM 151236
          2 IS MN 02 DEMS 151200 RRK
          3 IU AC 01 RKSL 151242
          4 IS MA 25 CWAO 151200 RRA
...
```

Fig. 5. Shell script used to read the messages header.

```
$ bufr_filter -o ${fname_out} - ${fname} <<@@
if (count == ${rec})
 if (dataCategory == 0 && (dataSubCategory == 6 || dataSubCategory == 7)) {
   append;
 }
}
@@
```

Fig. 6. ecCodes used to split the messages.

Pre-processing

Statistics

Tab. 1. Last days statistics of the synoptic acquisition in BUFR format. (Ref.: November, 2025)

	10/11	11/11	12/11	13/11	14/11	15/11
SABM	968	949	910	953	917	911
SLLP	14	14	15	14	13	13
SBBR	167	153	157	162	156	141
SCSC	54	50	50	48	54	64
SKBO	2	2	2	2	2	2
SEQU	4	4	4	4	2	2
SOCA	75	80	70	71	76	82
SYTM	0	0	0	0	0	0
SGAS	17	13	14	26	13	17
SPIM	6	6	3	4	5	4
SMZY	0	0	0	0	0	0
SUMU	48	48	48	48	48	48
SVMR	0	0	0	0	0	0

Pre-processing

Statistics

Tab. 2. Last days statistics of the radiosonda acquisition in BUFR format. (Ref.: November, 2025)

	10/11	11/11	12/11	13/11	14/11	15/11
SABM	13	12	12	14	10	14
SLLP	0	0	0	0	0	0
SBBR	6	5	6	6	7	6
SCSC	1	1	1	1	1	1
SKBO	0	0	0	0	0	0
SEQU	0	0	0	0	0	0
SOCA	4	4	4	2	4	2
SYTM	0	0	0	0	0	0
SGAS	0	0	0	0	0	0
SPIM	4	4	4	4	4	4
SMZY	0	0	0	0	0	0
SUMU	0	0	0	0	0	0
SVMR	0	0	0	0	0	0

Pre-processing

Quality Control

- In the pre-processing context, only gross controls

```
# Range
if d < 0 or d > 360:
 ...
if t < -56 or t > 56:
 ...

# Physical consistency
if d == 0 and f > 3:
 ...
if tn > t:
 ...
```

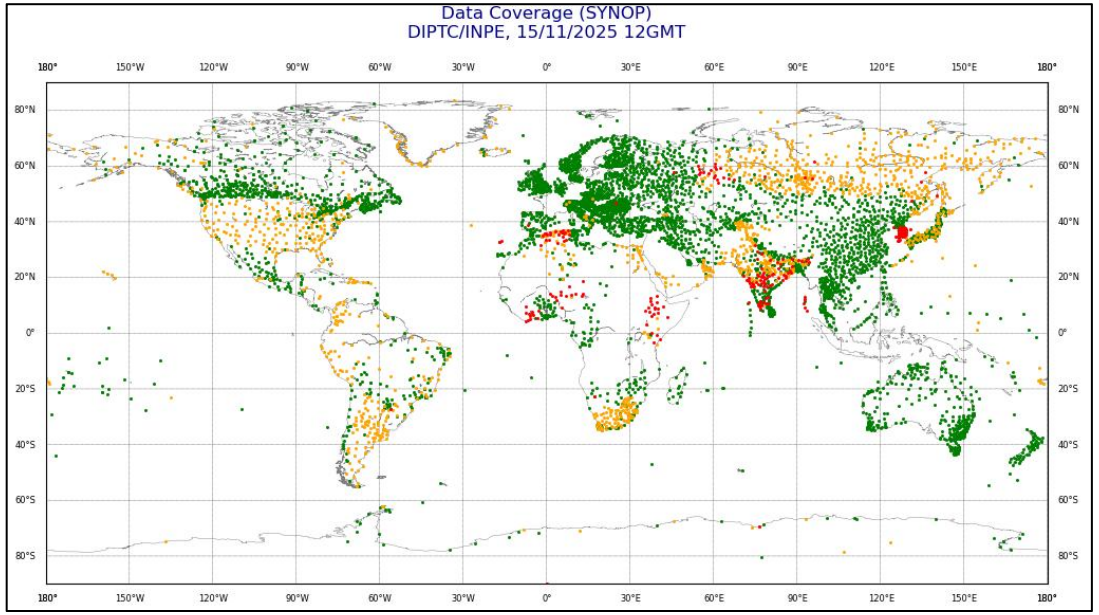
Fig. 7. Gross quality control checks.

- However, there are other checks at the database level
  - Step and neighborhood checks

Pre-processing

Quality Control

(a) No duplucate checks



(b) Duplicate checks

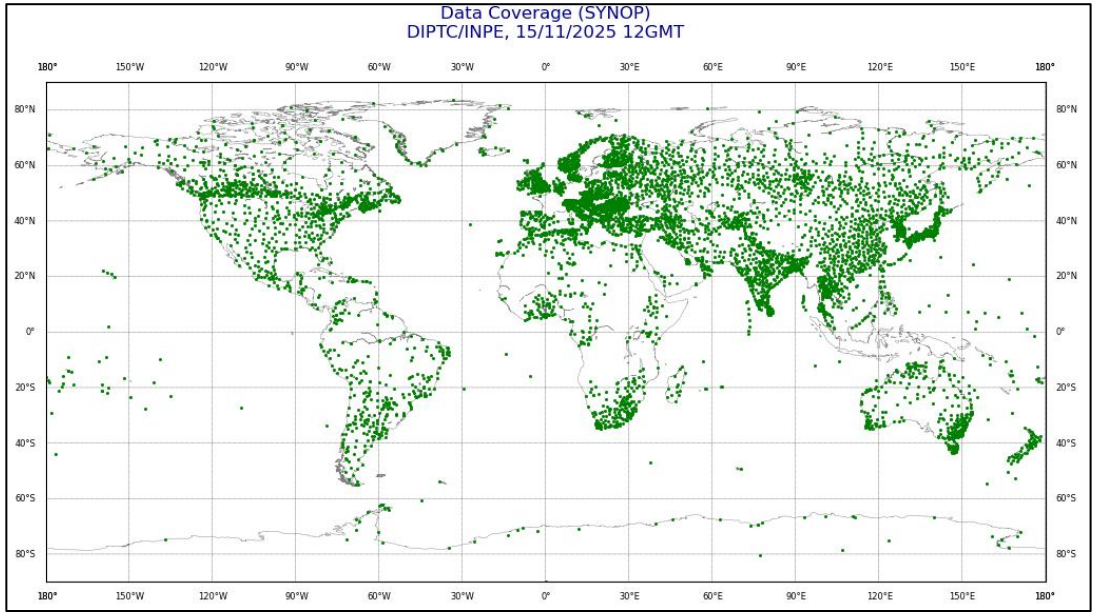


Fig. 8. Duplicate check.

- Why it occurs?
  - Joint of BUFR and TAC flow!
  - Same code, coords., station type, obs. date and centre

Pre-processing

Data Coverage

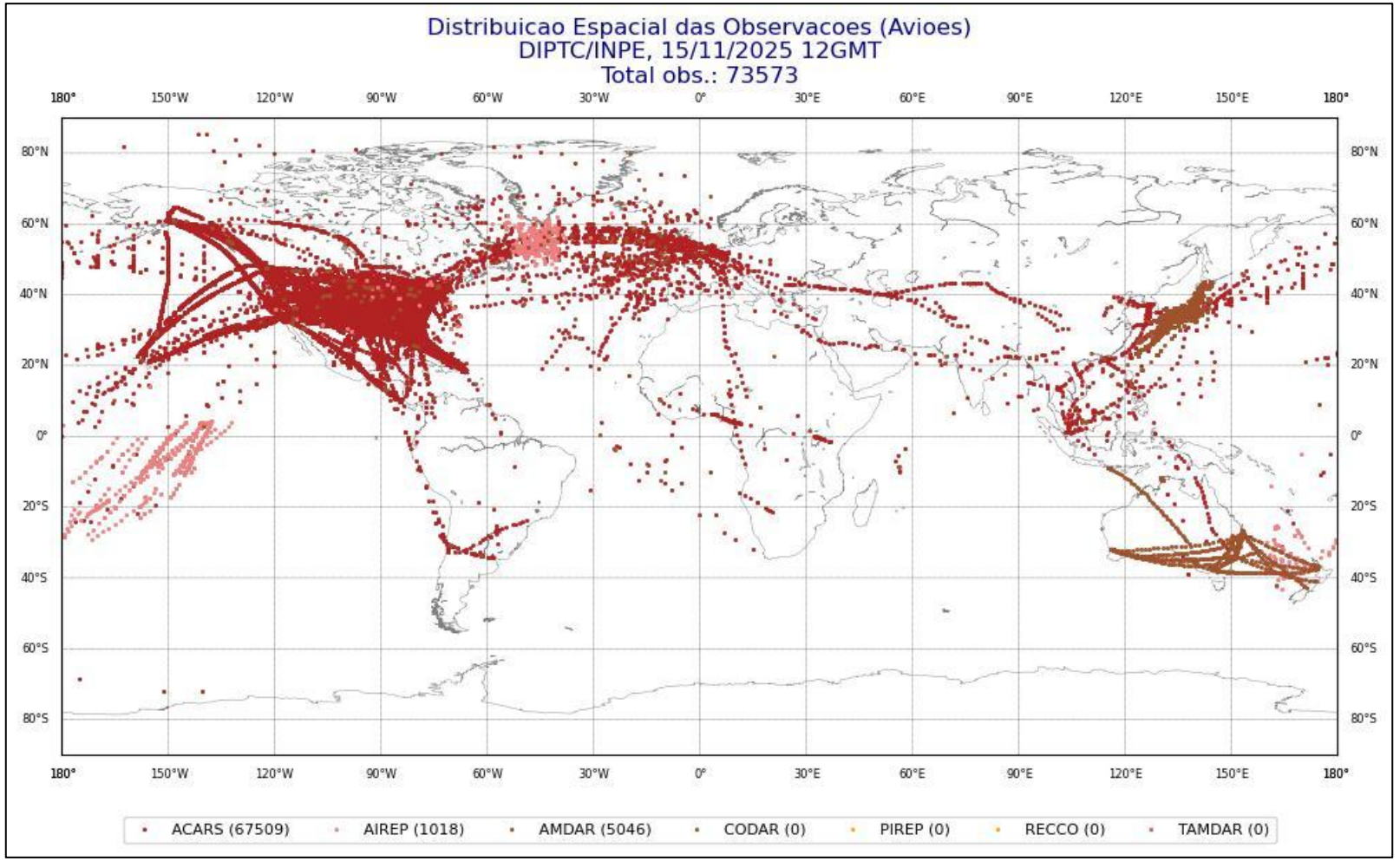


Fig. 9. Data coverage for aircraft observations.

Pre-processing

Data Coverage

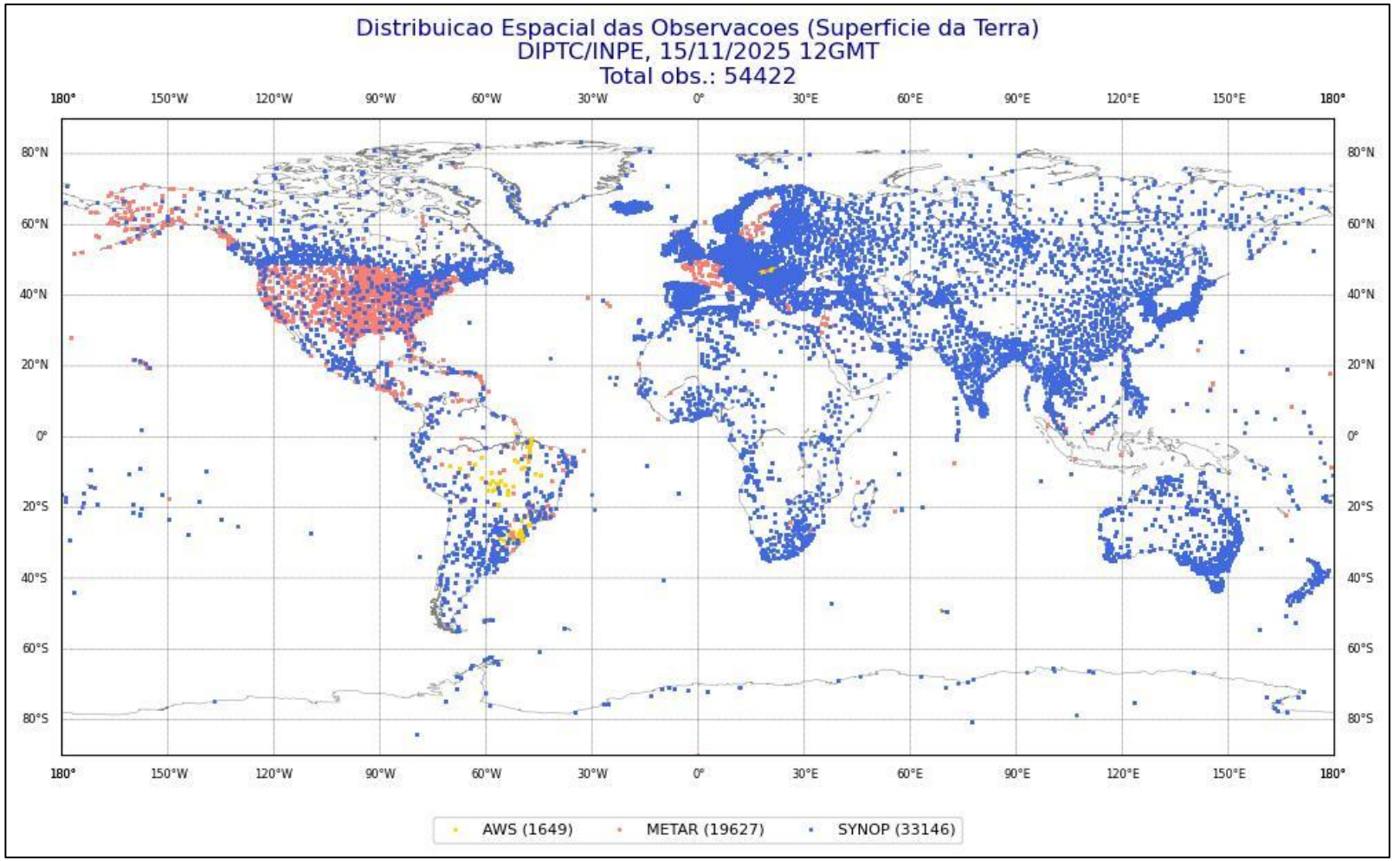


Fig. 10. Data coverage for Earth surface observations.

Pre-processing

Data Coverage

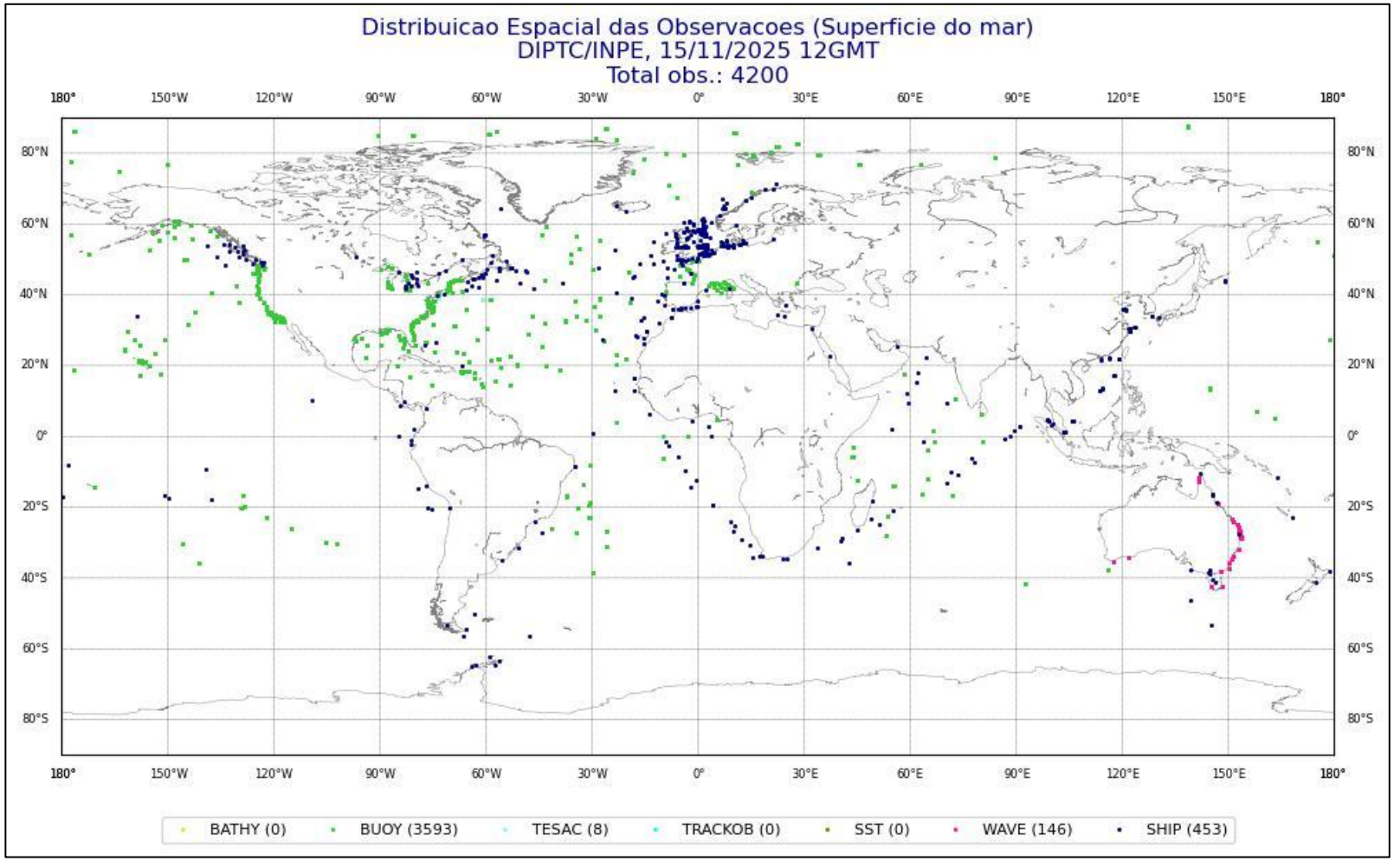


Fig. 11. Data coverage for buoy and ship observations.

Pre-processing

Data Coverage

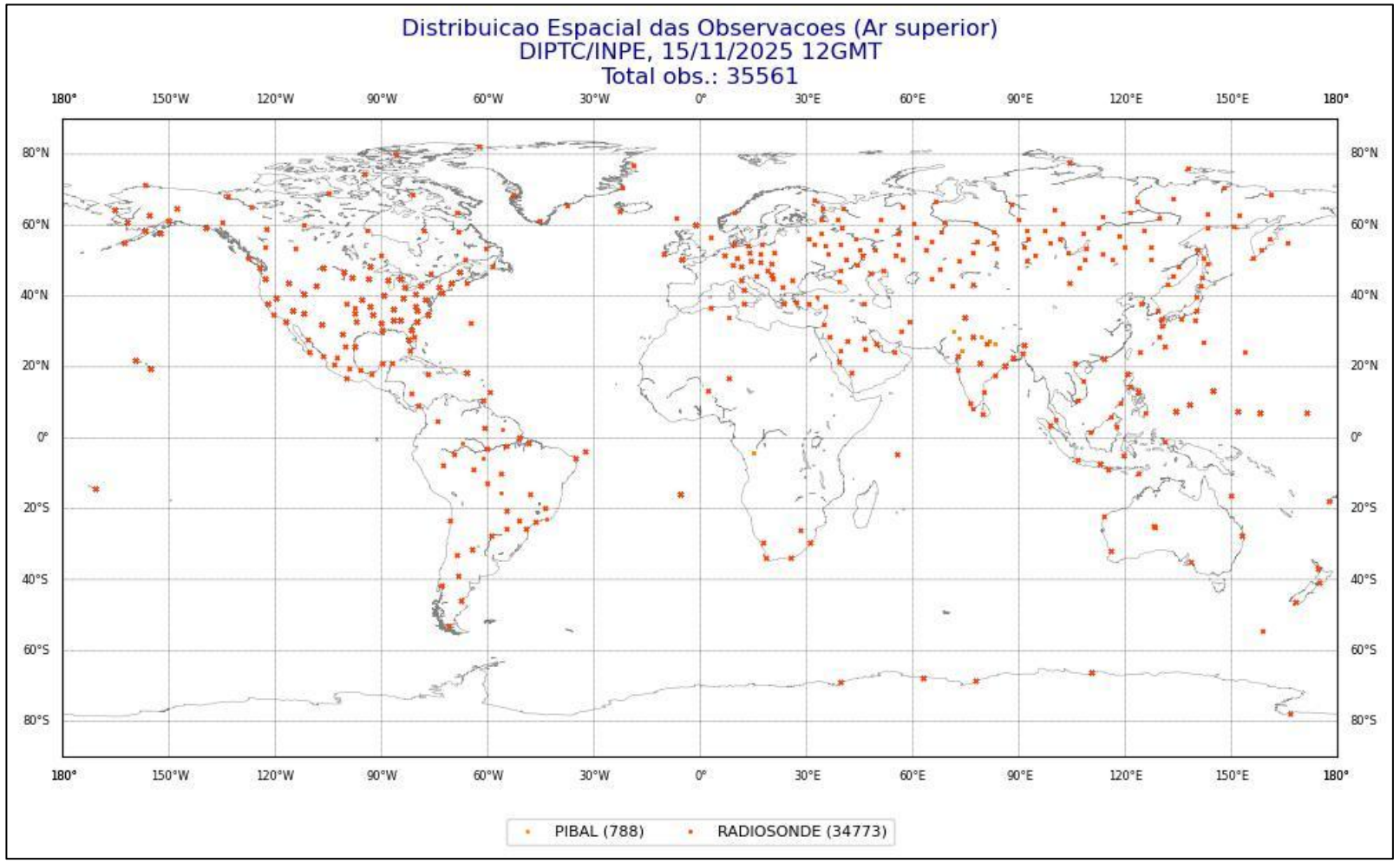


Fig. 12. Data coverage for ballon and radiosonde observations.

Database

Overview

- Raw (approximately) hourly data
  - Since, 1995
  - Data types: AIRCRAFT, AWS, BUOY, METAR, PILOT, SHIP, SYNOP and TEMP
- Derived daily data
  - Precipitation, since 1888
  - Temperatures, since 1920
  - Other parameters, since 1995
- Technology
  - PostgreSQL 14
  - Volume of obs ~ 1TB

Database

Overview

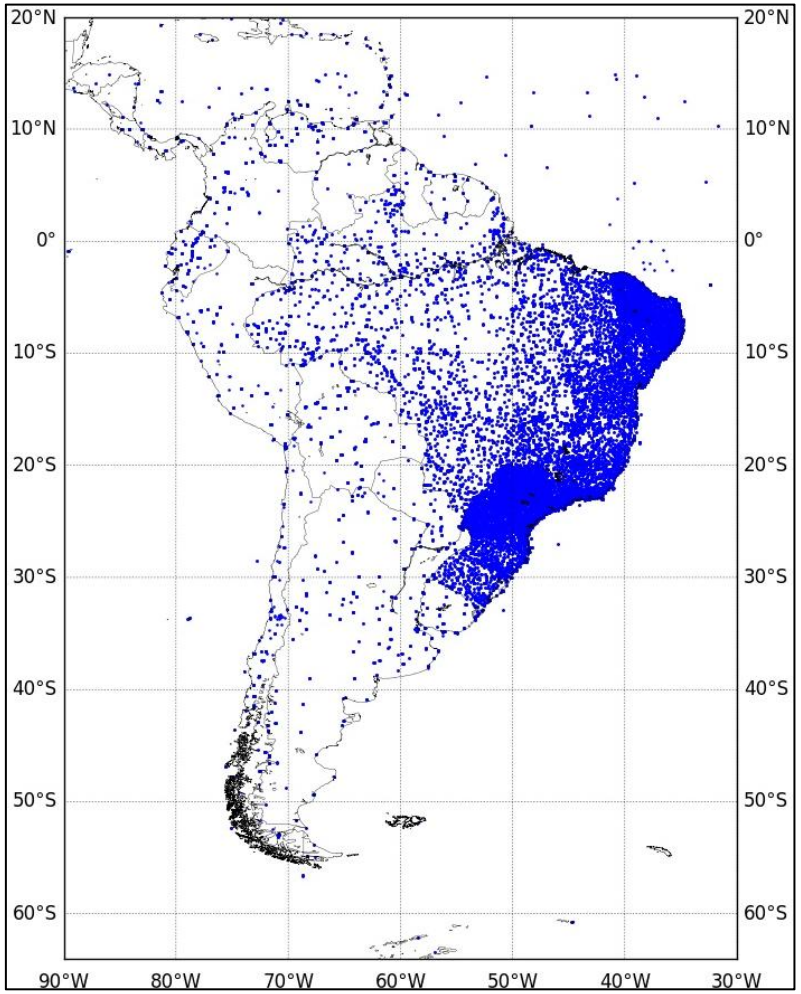


Fig. 12. Data coverage on the SA.

Database

Quality Control

- Flags
  - A: accept
  - S: suspect
  - X: reject
- Levels 1 and 2 (automatic checks)
  - Range
  - Physical
  - Step
  - Neighborhood
- Level 3 (on demand checks)
  - Suspect analisys

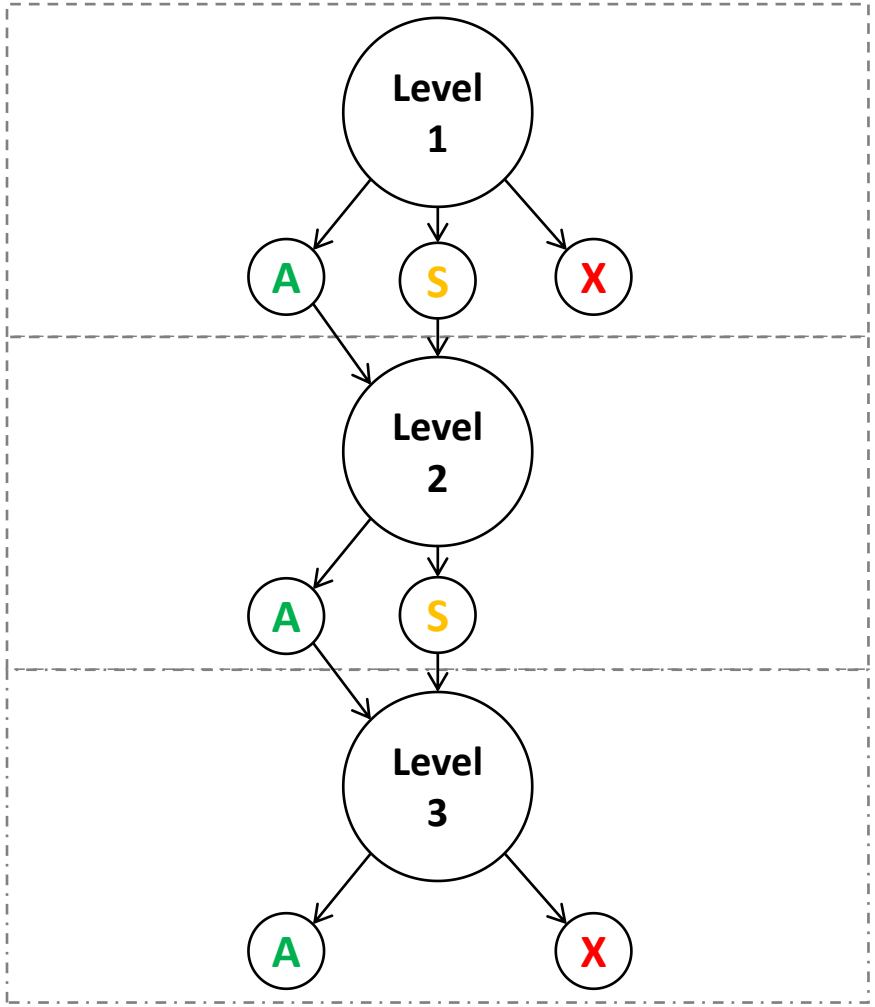


Fig. 13. Database quality control schema.

Data Acquisition and Processing

Questions?