RIHMI INPUT FOR WP3 WITHIN ERA CLIM2 PROJECT

BY ALEXANDER STERIN (RIHMI-WDC, OBNINSK, RUSSIA)

E-MAIL: STERIN@METEO.RU

HTTP://WWW.METEO.RU

Three kinds of in-situ RIHMI observational data:

Upper-air data

Inputs from:

computer media (not all stations, poor vertical resolution - not all levels, few standard pressure levels only)

handwritten tables

computerized views of old punch-card codes formats

There were references to early UA data on computer media (say, in IGRA), but they were episodic and incomplete!

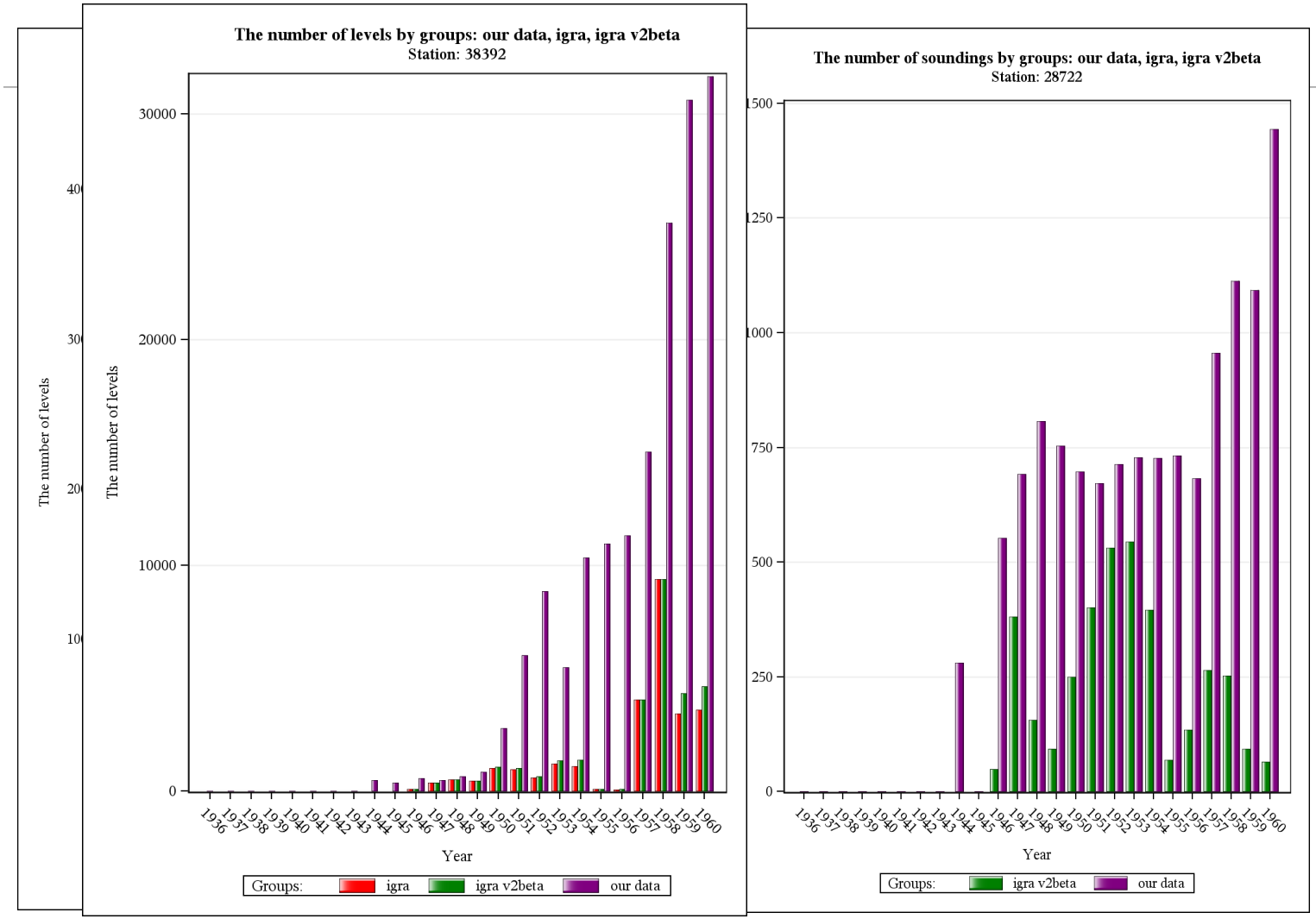
The main effort was to digitize 41 station data (41 stations is about 20% of fUSSR total)

Surface meteorological data

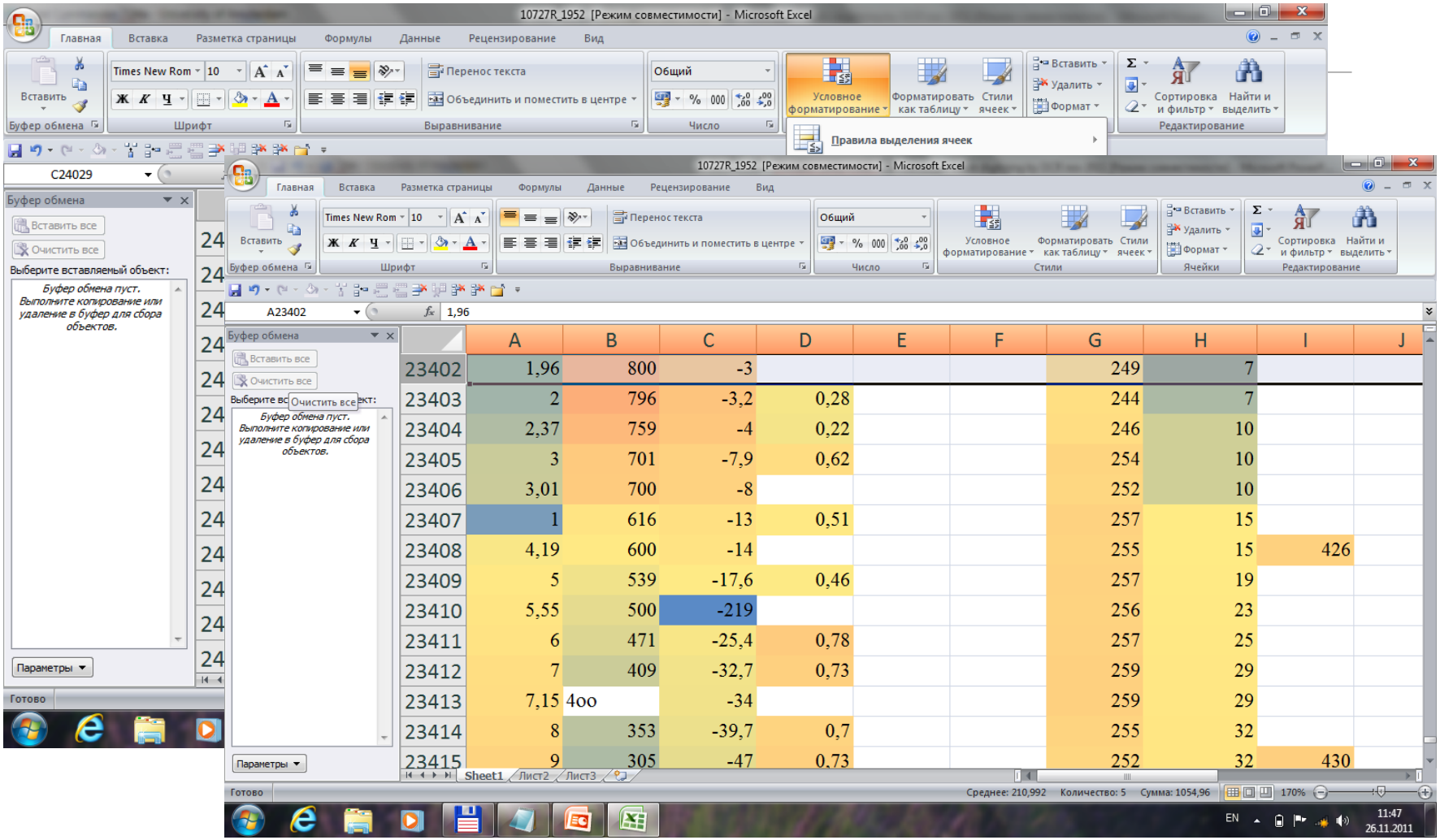
Snow cover data

Upper-air data (incomplete in previous computer sources).

station UFA (28722) – Missing in IGRA, partly present in IGRA2beta, 38392 TASHAUZ – present in IGRA



Observational Upper-Air data: Efforts on digitizing Manual digitizing and OCR digitizing



Observational Upper-Air data: Efforts on QC Statistical Tables and Statistical Graphs

STATISTICS FOR P, T AND U ON SELECTED HEIGHT LEVELS FOR R CAO

STATISTICS FOR P, T and U YEAR=1948

Процедура MEANS

			Analysis Variable : P					
H	mmon	N Obs	Mean	Std Dev	Minimum	Maximum	N	1
1000	1	57	888.21	12.27	861.00	908.00	57	
	2	51	893.38	9.84	864.00	915.00	50	
	3	55	889.15	10.49	868.00	908.00	55	
	4	62	897.19	7.05	882.00	910.00	62	
	5	61	902.00	7.52	883.00	917.00	61	
	6	56	896.39	4.65	887.00	906.00	56	
	7	63	894.52	6.15	878.00	907.00	63	
	8	55	895.35	4.70	885.00	906.00	55	
	9	53	897.02	10.48	871.00	910.00	53	
	10	58	892.60	9.85	871.00	909.00	58	
	11	69	893.57	8.50	874.00	911.00	69	
	12	51	898.55	10.40	881.00	920.00	51	
2000	1	57	780.26	11.61	754.00	798.00	57	
	2	51	782.18	8.57	759.00	799.00	50	

STATISTICS FOR F,T AND U ON STANDARD PRESSURE LEVELS FOR R CAO

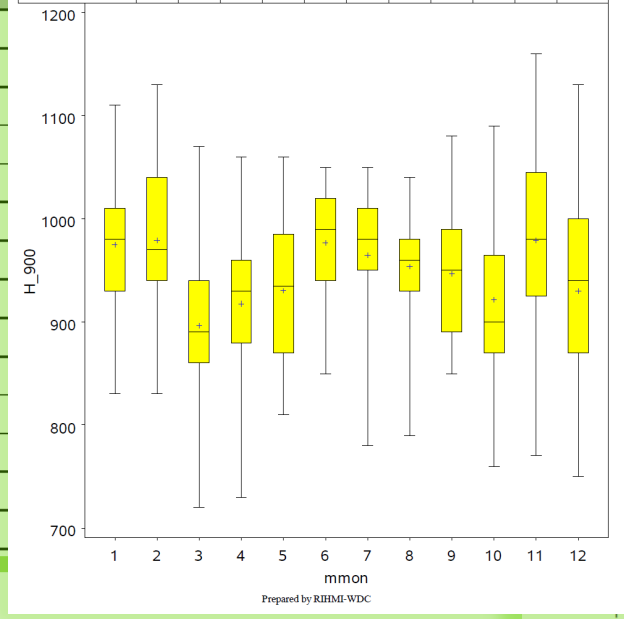
STATISTICS FOR F,T AND U ON STANDARD PRESSURE LEVELS FOR BIG TABLES
STATISTICS FOR F YEAR=1945

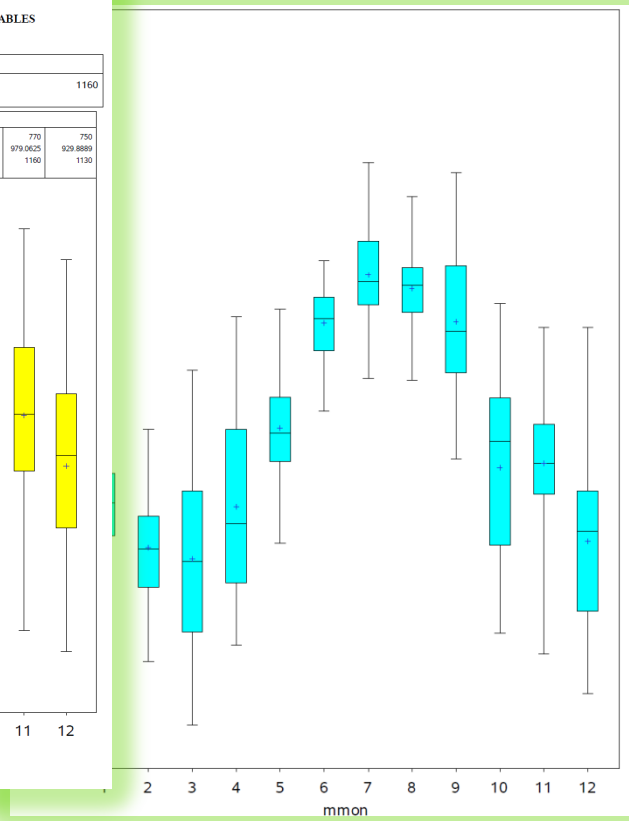
Overall Statistics

Min	720	Mean	947.6622	Max	1160
Pooled Std Dev	71.58791				

Extremes and means by month

Min	830	830	720	730	810	850	780	790	850	760	770	750
Mean	974.898	979.0741	896.2903	917.6667	930.3125	976.6234	964.9254	953.8462	946.7241	921.6667	979.0625	929.8889
Max	1110	1130	1070	1060	1060	1050	1050	1040	1080	1090	1160	1130





Observational Upper-Air data:

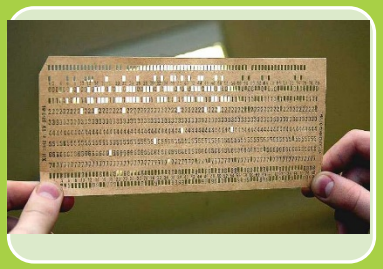
For 41 stations-

Computerized “views” of old punch-card codes formats – what is it?

The handwritten tables were manually keyed by punching machines to paper punch cards in the 1970s and 1980s.

The punch cards were, in fact, the sets of 80 coded sequences of 12 bit each.

Billions of punch cards were keyed in different centers at special divisions that were equipped by punching machines. Similar division was at RIHMI



The 12 bit sequences were corresponding to digits in some cases, but some of them could not be transformed directly into digits, characters or into known additional symbols in currently existing coding systems.

- For UA Data, the punch cards for different kinds of levels were in separated boxes

Observational Upper-Air data:

For 41 stations-

Computerized “views” of old punch-card codes formats – what is it?

In the late 1980s the paper punch cards were copied to 9-track magnetic tapes “as-is”). No transformations and no processing!

- The punch cards were disseminated by kinds of levels, so that standard levels, tropopause levels, significant point data were separated into different chunks and not ready for reconstructing whole radiosonde profiles

These “views” of punch cards, though being put later from 9-track tapes into modern computer media, were not digital data!

- As intermediate media, OVERLAND DATA TAPE UNIT was used
- In the early 2000’s, the 9 track tape content migrated to OVERLAND DATA, and 9 track tape collection was annihilated

In 2008-2009, the “computerized Views” of punch cards were finally put to modern robotic tape library. But still without any processing!

- The processing was within ERA CLIM2 Project

Observational Upper-Air data:

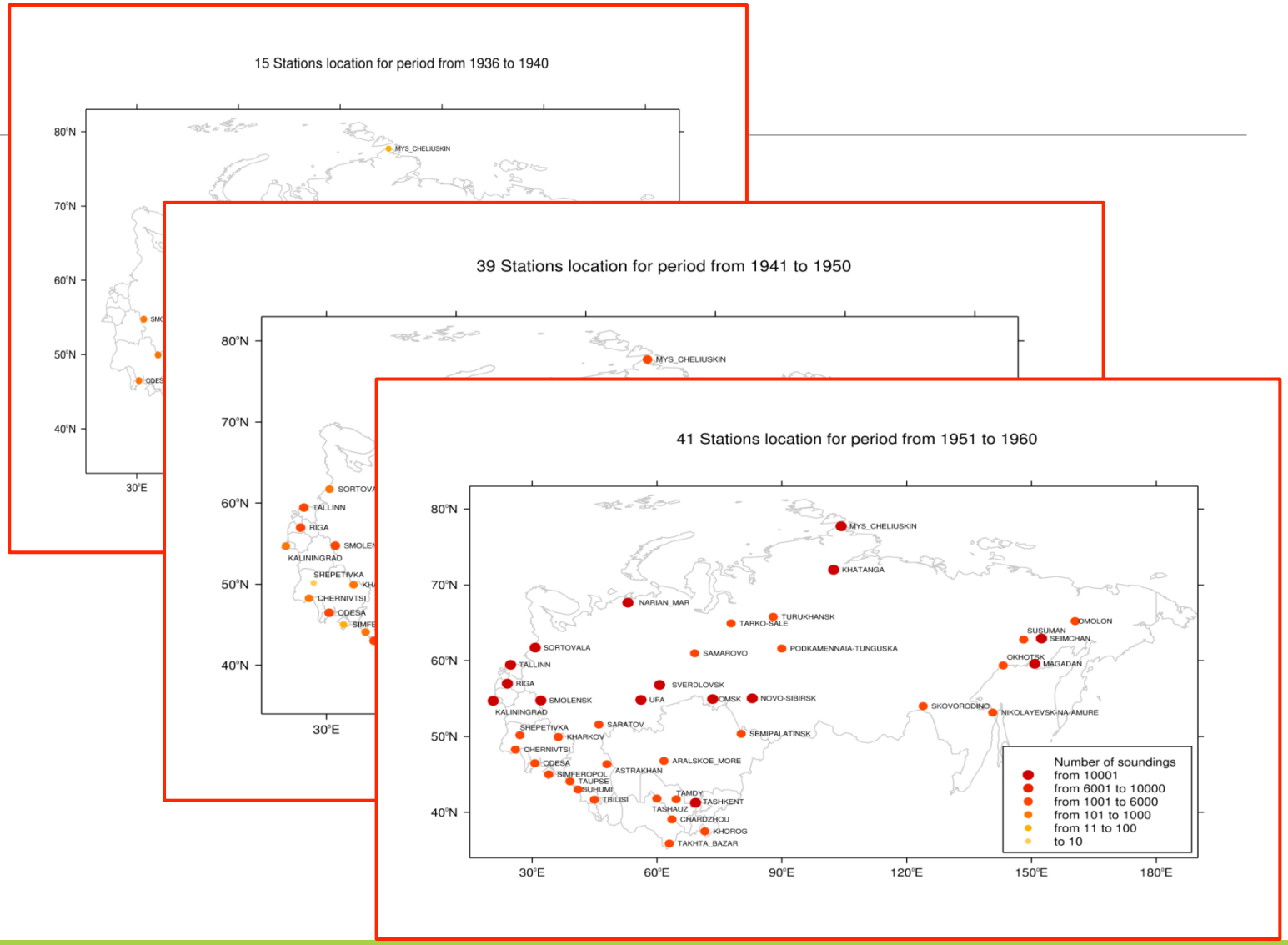
Computerized “views” of old punch-card codes formats – what is it?

An essential set of transforming operations within ERA CLIM2, including the use of special “dictionaries” for transforming 12 bit sequences into codes of values was applied.

Finally, the output symbols that became acceptable for creation of data sets, were checked within this transformation to avoid output decoded symbols that did not correspond to digits, or symbols that compose unrealistic numeric values.

Within this process various kinds of levels were merged, many other operations were applied, so the data sets for 41 stations were produced as a result, that were acceptable for future calculations and analysis!

Upper-air data (41 stations)



The final U/A dataset was compiled from three sources and put into single format as it was done for the U/A data in previous ERA CLIM Project. **The amount of soundings: 390 873. The amount of levels: 8 993 028**

Upper-air data

The QC for U/A data:

For the U/A data the QC was applied based on the following efforts:

Calculation of statistical tables (frequencies to control key element values)

Calculation of statistical tables (univariate statistics for each parameter to detect suspicious and erroneous values)

Statistical graph assessment (boxplots to detect and flag outliers)

Use of climatological statistics for the U/A parameters, **for the same 41 stations, but calculated for a later period (we used climate monthly statistics for 1985-2014 that were calculated from the RIHMI AEROSTAB global radiosonde data set).** Based on this climatology, the flagging was provided.

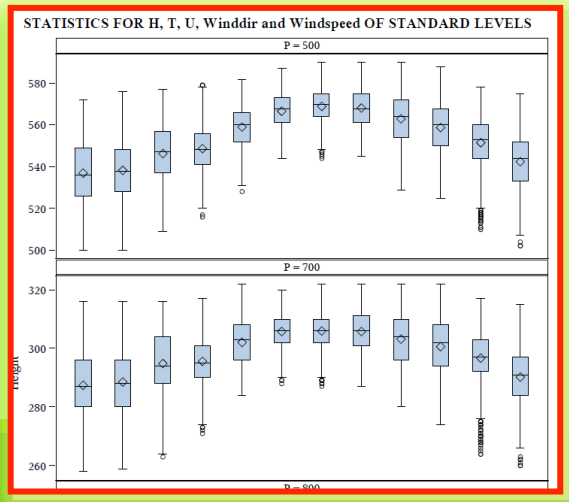
QC for French and Spain data – climate statistics from later periods not available – more rough check was available

STATISTICS FOR P, T AND U ON SELECTED HEIGHT LEVELS FOR R CAO

STATISTICS FOR P, T and U YEAR=1948

Процедура MEANS

Analysis Variable : P										
H	mmon	N Obs	Mean	Std Dev	Minimum	Maximum	N	10th Pctl	Median	90th Pctl
1000	1	57	888.21	12.27	861.00	908.00	57	869.00	891.00	903.00
	2	51	893.38	9.84	864.00	915.00	50	881.00	894.00	903.00
	3	55	889.15	10.49	868.00	908.00	55	874.00	888.00	903.00
	4	62	897.19	7.05	882.00	910.00	62	887.00	899.00	905.00
	5	61	902.00	7.52	883.00	917.00	61	895.00	901.00	911.00
	6	56	896.39	4.65	887.00	906.00	56	890.00	896.50	903.00
	7	63	894.52	6.15	878.00	907.00	63	888.00	894.00	903.00
	8	55	895.35	4.70	885.00	906.00	55	889.00	896.00	901.00
	9	53	897.02	10.48	871.00	910.00	53	879.00	900.00	907.00
	10	58	892.60	9.85	871.00	909.00	58	875.00	895.00	903.00
	11	69	893.57	8.50	874.00	911.00	69	880.00	895.00	903.00
	12	51	898.55	10.40	881.00	920.00	51	884.00	899.00	914.00
2000	1	57	780.26	11.61	754.00	798.00	57	763.00	781.00	795.00
	2	51	782.18	8.57	759.00	799.00	50	769.00	783.00	791.00



The final Russian U/A dataset was compiled from three sources and put into single format as it was done for the U/A data in previous ERA CLIM Project.

The amount of soundings: 390 873.

The amount of levels: 8 993 028

```
23205   67   39   53    1    9 1940   11   19  1    5    2    9   400  1  612  0  -10.8  1  91  1    .  9    .  9
 23205   67   39   53    1    9 1940   11   19  1    5    2    9   500  1  534  0  -17.4  1  91  1    .  9    .  9
 23205   67   39   53    1    9 1940   11   21  1    5    2    9   400  1  581  0  -34.8  1  76  1    .  9    .  9
 23205   67   39   53    1    9 1940   11   21  1    5    2    9   500  1  491  0  -42.0  1   .  9    .  9    .  9
 23205   67   39   53    1    9 1940   11   23  1   11    3    9   400  1  591  0  -21.5  1  75  1    .  9    .  9
 23205   67   39   53    1    9 1940   11   23  1   11    3    9   500  1  514  0  -24.7  1  72  1    .  9    .  9
 23205   67   39   53    1    9 1940   11   23  1   11    3    9   600  1  452  0  -28.8  1  69  1    .  9    .  9
 23205   67   39   53    1    9 1943    1   19  1    5    3    9    20  1  999  0    0.3  1  91  1    .  9    .  9
 23205   67   39   53    1    9 1943    1   19  1    5    3    9    50  1  963  0    0.0  1  91  1    .  9    .  9
 23205   67   39   53    1    9 1943    1   19  1    5    3    9   100  1  902  0   -3.5  1  91  1    .  9    .  9
 23205   67   39   53    1    9 1943   11    9  1   11   10    9    20  1  994  0   -0.6  1  82  1    .  9    .  9
 23205   67   39   53    1    9 1943   11    9  1   11   10    9    50  1  956  0   -1.3  1  79  1    .  9    .  9
 23205   67   39   53    1    9 1943   11    9  1   11   10    9   100  1  895  0   -4.2  1  75  1    .  9    .  9
 23205   67   39   53    1    9 1943   11    9  1   11   10    9   150  1  837  0   -8.7  1  73  1    .  9    .  9
 23205   67   39   53    1    9 1943   11    9  1   11   10    9   200  1  785  0  -11.8  1  73  1    .  9    .  9
 23205   67   39   53    1    9 1943   11    9  1   11   10    9   300  1  688  0  -15.7  1  82  1    .  9    .  9
 23205   67   39   53    1    9 1943   11    9  1   11   10    9   400  1  601  0  -20.0  1   .  9    .  9    .  9
 23205   67   39   53    1    9 1943   11    9  1   11   10    2   451  1  561  0  -21.2  1   .  9    .  9    .  9
 23205   67   39   53    1    9 1943   11    9  1   11   10    9   500  1  525  0  -19.8  1   .  9    .  9    .  9
 23205   67   39   53    1    9 1943   11    9  1   11   10    9   600  1  455  0  -26.2  1   .  9    .  9    .  9
 23205   67   39   53    1    9 1943   11   10  1   10    7    1     1  0 1027  0   -4.4  0  91  1    0  1    0  1
 23205   67   39   53    1    9 1943   11   10  1   10    7    9    20  0 1002  0   -5.8  0  91  1    .  9    .  9
```

SAS Program fragment for output:

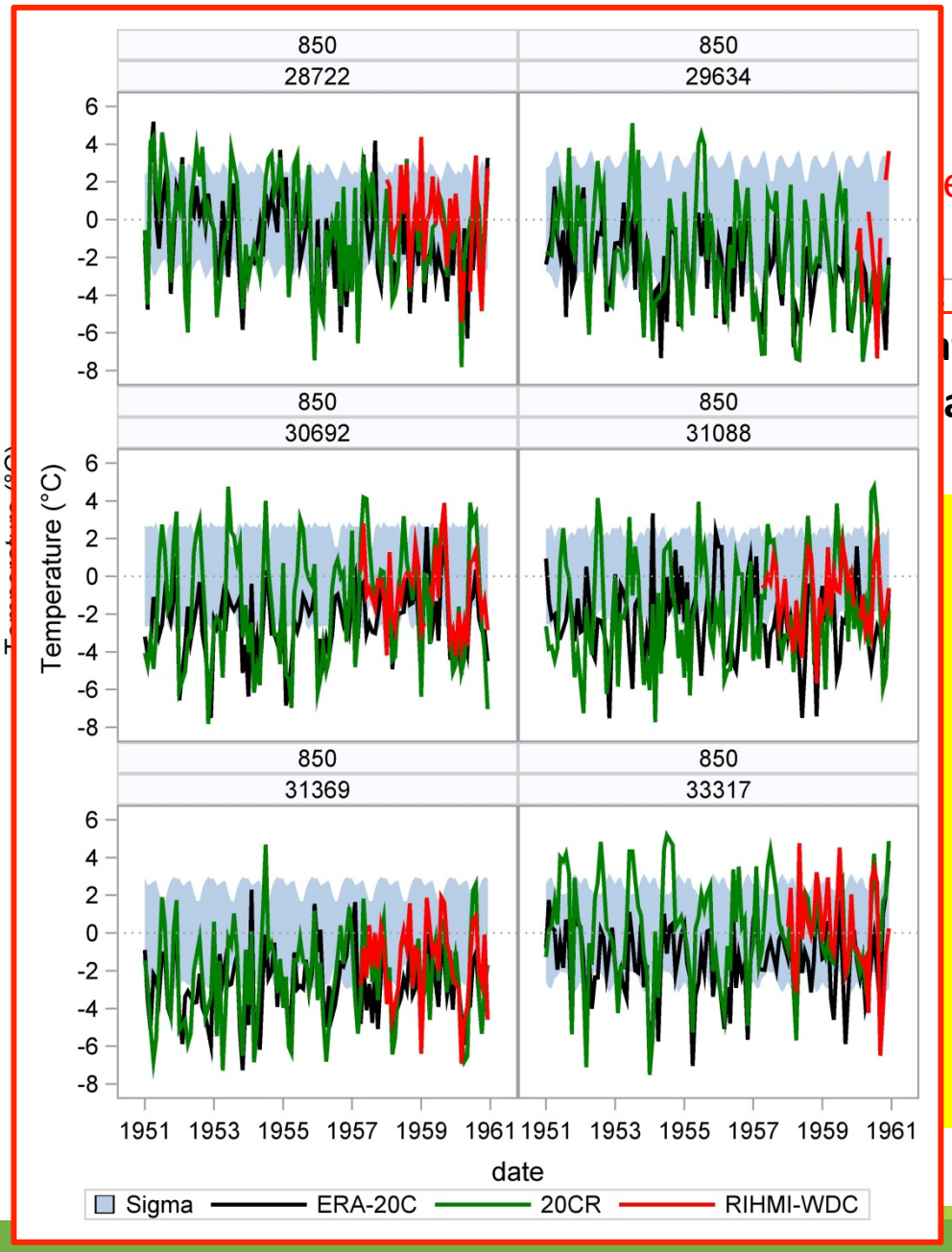
put index 6. (latdeg latmin londeg lonmin hgtstat year month day) (5.) q_day 3. (time nlev code) (5.)

H 6. q_H 3. P 5. q_P 3. T 7.1 q_T 3. U 4. q_U 3. winddir 5. q_winddir 3. windspeed 5. q_windspeed 3.;

eters, **for the same 41**

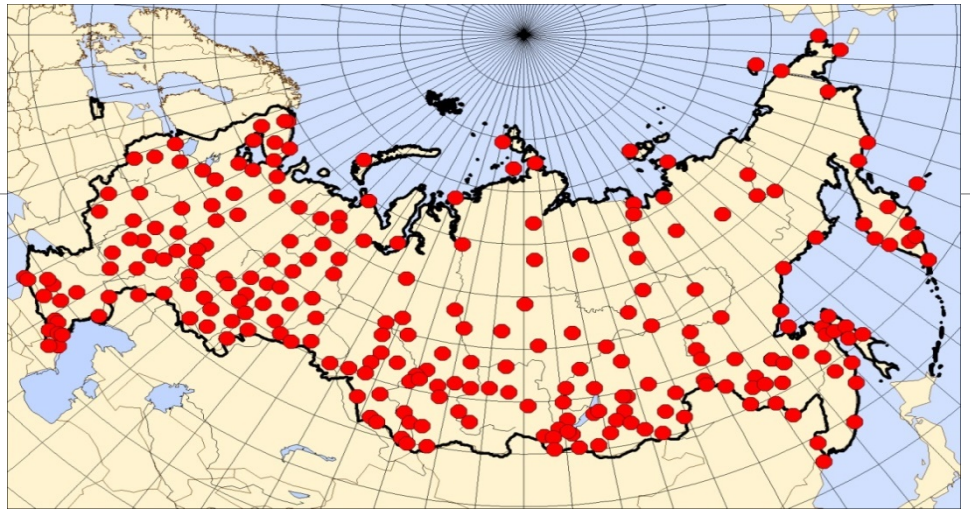
!

ıtion climatologies that were
ails – in WP4)



Panel: monthly anomalies for 100hPa Temperature for group of 6 stations for 1951-1960
Grey shaded – monthly climatology sigmas
ERA-20C – black
20CR v2 – green
RIHMI digitized - red
ERA-20c better corresponds to climatology!

Surface meteorological data



*Map of stations containing sub-daily meteorological observation for 1965 and back till the beginning of observations**
(For most of 246 stations, period begins in middle 1930s)*

The goal within ERA CLIM2 Project was to extend part of stations data for period from the beginning of observation and ending by 1965 (provided for 4 observations per day). The subset of 246 stations of the 518 was eventually selected for this extension and data were prepared

The following data operations were done for 246 stations to obtain the extension for 1965 and back:

- *The assessment of existing sources of data on computer media in odd old formats and in hardcopies, to select subset of stations acceptable for period extension*
- *The reformatting of old odd formatted data, to fit the common format for 1966 and later data set.*
- *Filling gaps in data for period before 1966 by digitizing hardcopy materials, transforming digitized data to common format*
- *Operations with date and time variables, setting Greenwich date-time*
- *QC of data*

Surface meteorological data

DATA QUALITY

The dataset contains the results of main hourly meteorological observations. The data accuracy is appropriate to the measurement accuracy of meteorological parameters indicated in the “Instructions for Meteorological Stations and Observation Rooms” (No.3, part 1, 1985).

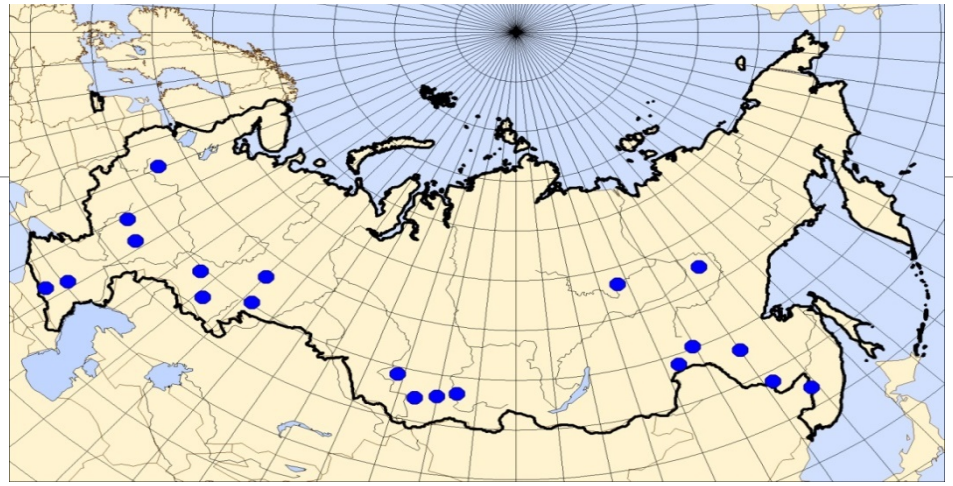
Most of the elements in the dataset have quality flags that can take on the following values:

0 – value is reliable;
1 – value is reliable and recovered manually;
2 - value is reliable and recovered automatically;
3 – value is dubious;
4 – value is rejected by syntactic and semantic check software programs;
5 – value is missing but observations were conducted;
6 – value is rejected at the station;
7 – value is missing since no observations were conducted;

If a value of the element is missing, quality flags take on the values from 4 to 7.

All the elements have been checked for permissible values.

Snow cover data



The goal within ERA CLIM2 Project was to extend the list of stations in the snow parameter data set, by preparing data that were not available on computer media before. This was done for 20 additional stations

Records in data files are arranged by increasing key elements:

- year;
- month;
- day.

```
29956 1936   3 14    14   1 0 0 0
29956 1936   3 15    13   1 0 0 0
29956 1936   3 16    14   1 0 0 0
29956 1936   3 17    16   1 0 0 0
29956 1936   3 18    15   1 0 0 0
29956 1936   3 19    22   1 0 0 0
29956 1936   3 20    20   1 0 0 0
29956 1936   3 21    19   1 0 0 0
29956 1936   3 22    19   1 0 0 0
29956 1936   3 23    17   1 0 0 0
29956 1936   3 24    17   1 0 0 0
29956 1936   3 25    16   1 0 0 0
29956 1936   3 26    15   1 0 0 0
29956 1936   3 27    15   1 0 0 0
29956 1936   3 28    15   1 0 0 0
29956 1936   3 29    14   1 0 0 0
```

Snow cover data

Observation period	Extent of snow cover around the station	Value Q
Before July 1959	50% and less than 50% of the area around the station	0
	More than 50% of the area around the station	1
From August 1959 up to the present day	Extent of snow cover around the station is estimated from eleven-number scale. For example, the in case of lack of snow is 0; 20% of the area around the station covered with snow is 2; 50% of the area around the station covered with snow is 5, etc.	From 0 to 10

Situation	Q1 value
Value of snow depth is correct	0
Lack of snow	1
Snow cover absent at site, however snow is recorded in the vicinity of the station	2
Snow cover is less than 0.5 cm	3
Observations were not made or snow depth value is rejected	9

Snow cover data

Situation	Q2 value
Value is doubtful on conditions that Abs(**D**)>30 & abs(Δ**H1**)>10 & abs(Δ**H2**)>10 **Notes: D**=ΔH_2-ΔH_1, **ΔH** is the difference between the neighboring snow depth values	1
For the rest of the cases	0

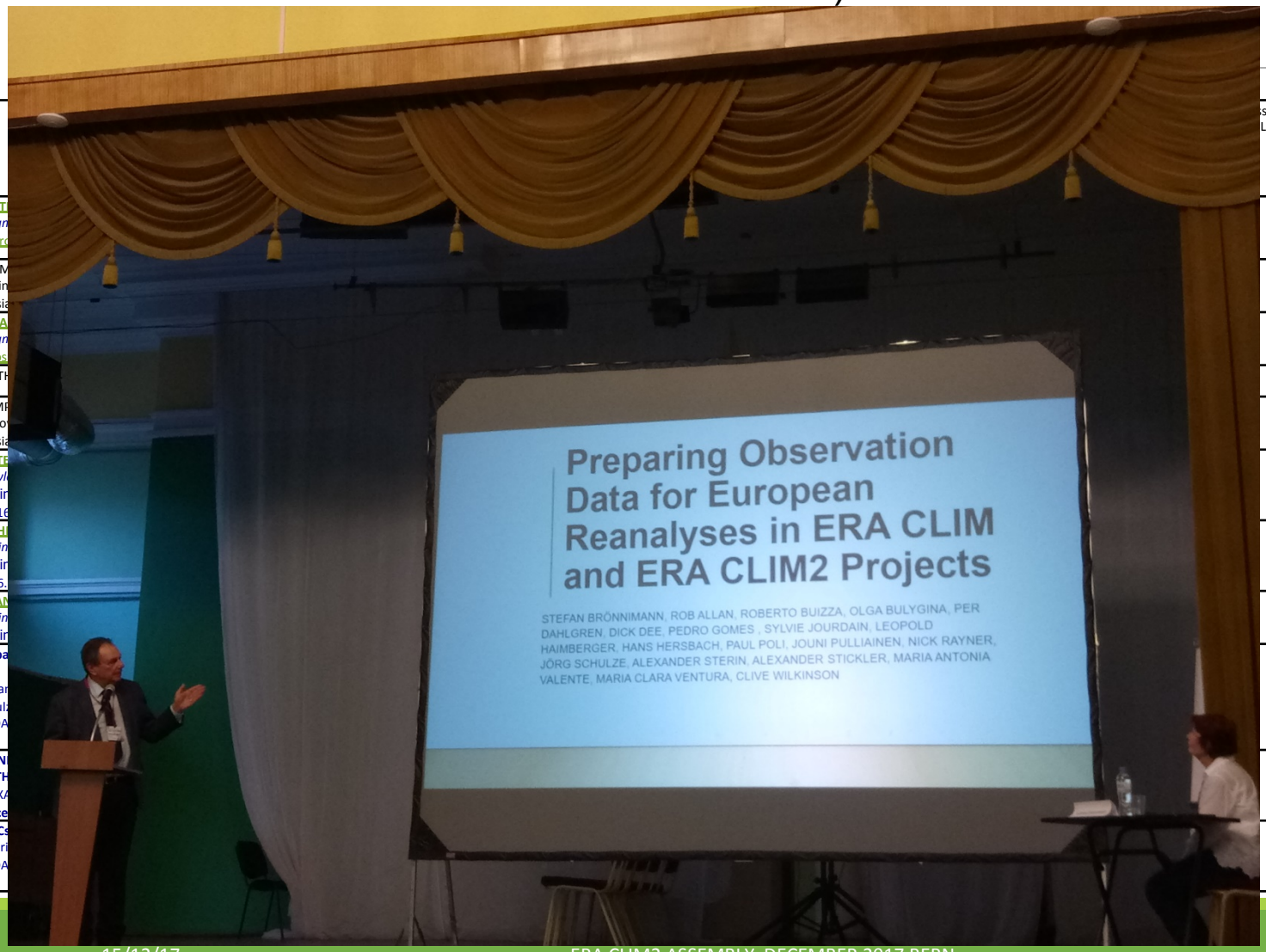
* Successive snow depth values are compared. **Snow depth is a cumulative inertial characteristic.** This QC is oriented to detection of "jumps" in the snow depth data. It is considered that if we have three successive values, and there is a "jump" from first to second value and opposite "jump" from second to third value, so that third value becomes close to first value, - then second value is doubtful.
Value 1 is set to Q2, provided that the difference between successive snow depth values is 30 cm and over, with snow depth from consecutive observations differing by 10 cm and over.

Situation	Q3 value
H^=0 & **Tmean** >5 & **Tmin**>0	1
For the rest of the cases	0

** Snow depth values are analyzed at different mean and minimum daily air temperatures. Value 1 is set to Q3, provided that with positive minimum and mean daily air temperatures higher than 5°C, snow depth is more than 0

Publications by RIHMI ERA CLIM2 TEAM

(basic publications on ERA CLIM2 direct results authored by international group members, such as BAMS Papers, are not included)



		publication that use ERA-CLIM2 data or acknowledges ERA CLIM2 support	publication written by people funded by ERA-CLIM2
ARCT… Wegn… Enviro…			Yes
ESTIM… Sterin… Russia…			Yes
EURA… Wegn… Cryos…			Yes
ON TH…		Yes	Yes
COMP… Lavro… Russia…			Yes
SYSTE… Kopyl… Geoin… . 2016			Yes
TECH… Sterin… Geoin… 2016.		Yes	Yes
QUAN… Sterin… Geoin…		Yes	Yes
Prepa… Stefan… Schul… CODA…		Yes	Yes
MON… METH… ALEXA… **Proce…**			Yes
WDCs Valeri… CODA…			Yes

The ERA-PreSAT experimental re-analysis: The value of using upper-air data in historical reanalyses

Paper from group of authors was sent to QJRMS in 2016 (Hans Herzbach et al.)

The contribution of U/A data was assessed

What about the contribution of surface meteorological data? Of snow cover data?

Continue to present results at conferences: ENVIROMIS 2018 (Tomsk, July 2018

Future ERA reanalysis projects (maybe ERA CLIM3):

RIHMI is interested in participating Future ERA Reanalysis project

What are the mechanisms of participation for non-EU countries (for Russia in particular) in Horizon 2020?

Two options:

1. If Ministry of Science (MOS) of Russia supports “mirror” project for institutions participating in H2020 consortium
   - Not all H2020 have “mirror” projects from MOS
   - “Mirror” project may be for only one organization
2. If EU participant of H2020 consortium involves non-EU participant as sub-contractor

The ERA-CLIM2 project: lessons learned and open questions for the future

OCR plus simple solutions based on MS Office options came to be highly effective for data digitizing!

Early data (though episodic and incomplete), - provide improvement for reanalysis, that was demonstrated for UA data, but should be studied for other data inputs

Good and complete data for late period are needed to provide climatology for early data check (1950s and back)

Need to summarize the references to ERA CLIM data from all sources worldwide (maybe, by WebOfScience or Scopus facilities)

For ERA CLIM3:

What can RIHMI offer?

What mechanism will work for involving non-EU participants?

Thank you for attention!

Snow cover data

№	Index WMO	Name	Latitude	Longitude	Beginning	End
1	24639	NJURBA	63°17’	118°20’	1936	2014
2	24768	CURAPCA	62 ° 02’	132 ° 36’	1936	2014
3	26393	VYSNIJ VOLOCEK	57 ° 32’	34 ° 33’	1893	2014
4	27935	MICURINSK	52 ° 53’	40 ° 29’	1927	2014
5	28240	NIZHNYJ TAGIL	57 ° 53’	60 ° 04’	1936	2014
6	28645	CHELJABINSK	55 ° 09’	61 ° 18’	1900	2014
7	28711	BUGUL’MA	54 ° 38’	52 ° 48’	1932	2014
8	28925	MELEUZ	52 ° 57’	55 ° 58’	1936	2014
9	29736	MASLJANINO	54 ° 20’	84 ° 13’	1937	2014
10	29869	ERMAKOVSKOE	53 ° 18’	92 ° 25’	1936	2014
11	29956	TASTYP	52 ° 48’	89 ° 53’	1936	2014
12	30499	TYNDA	55 ° 11’	124 ° 40’	1936	2014
13	30683	EROFEJ PAVLOVIC	53 ° 58’	121 ° 56’	1936	2014
14	31318	NORA	53 ° 21’	130 ° 00’	1936	2014
15	31702	OBLUC’E	49 ° 00’	131 ° 05’	1936	2014
16	31832	BIKIN	46 ° 48’	123 ° 416’	1922	2014
17	34056	RTISCEVO	52 ° 15’	43 ° 47’	1936	2014
18	34964	ARZGIR	45 ° 24’	44 ° 12’	1936	2014
19	36061	TUROCAK	52 ° 15’	87 ° 08’	1936	2014
20	37212	NAL’CIK	43 ° 32’	43 ° 38’	1936	2014

U

Table of year by month

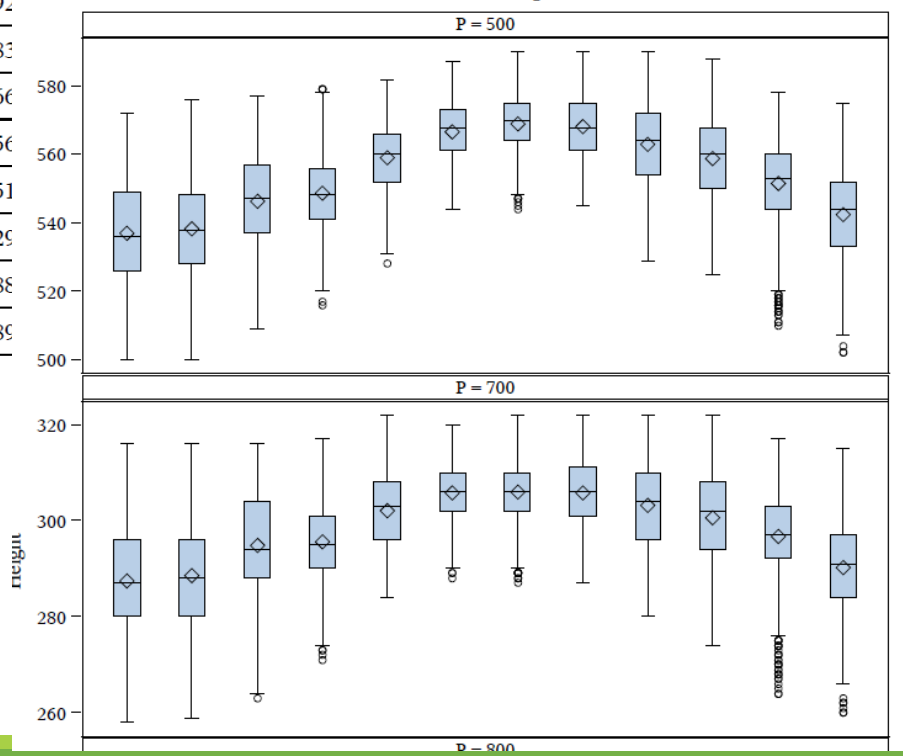
month

year	1	2	3	4	5	6	7	8	9	10	11	12	Total
19													
19													
19													
19													
19													
19													
To													

Analysis Variable : H

P	month	N Obs	Mean	Std Dev	Minimum	Maximum	N	10th Pctl	Median	90th Pctl
300	5	413	907.94	17.78	869.00	947.00	413	886.00	906.00	933.00
	6	405	920.09	13.74	884.00	954.00	405	902.00	921.00	936.00
	7	400	932.85	14.26	892					
	8	375	929.77	13.72	883					
	9	277	913.79	18.03	866					
	10	355	903.45	17.41	856					
	11	382	900.27	17.88	851					
	12	399	881.06	19.31	829					
500	1	423	527.66	15.32	488					
	2	389	527.79	16.22	489					

STATISTICS FOR H, T, U, Winddir and Windspeed OF STANDARD LEVELS



Snow cover data

Field number	Position	Field length	Field name	Notes
1	1-5	5	WMO index of station	Fixed for the file
	6	1	Blank	
2	7-10	4	Year	
	11		Blank	
3	12-13	2	Month	
	14	1	Blank	
4	15-16	2	Day	
	17	1	Blank	
5	18-21	4	H – snow height	In cm
	22	1	Blank	
6	23-24	2	S-extent of snow cover around the station	In numbers on ten-number scale, see Table 3.2
	25	1	Blank	
7	26	1	Q1 - additional information on snow depth	See Table 3.3
	27		Blank	
8	28		Q2 – quality flag from snow depth	See Table 3.4
	29		Blank	
9	30		Q3 – additional information with regard for air temperature	See Table 3.5
	31	1	Line end character	