

DestinE

Climate Digital Twin

Observation Application

(OBSALL Apps)

University of Helsinki team:

Alexander Mahura

Heikki Järvinen

Jouni Räisänen

Lauri Tuppi

Madeleine Ekblom

April 2024

DestinE

Climate Digital Twin

Observation Application

(OBSALL Apps)

OBSALL Apps includes

3 components for observations:

- Synoptic surface ([SYNOP](#)),
- Upper air sounding ([TEMP](#)),
- Satellite-based ([AMSU-A](#)).

The OBSALL Apps is developed as a set of scripts (*written in bash, fortran, python*):

- 1. Pre-processing modeled data**
- 2. Extracting observations (from Observation DataBase, ODB) - from synoptic and radiosounding stations, and satellites**
- 3. Extracting modeled data at observation times and locations, applying appropriate observation operator, and ODB added with model simulation values**
- 4. Calculating/ producing relevant statistics for monitoring**

Note, the end result is the augmented ODB, which can be used:

- on-line monitoring of the simulation in near-real time**
- posterior analysis of the ClimateDT model quality**

The OBSALL Apps is developed as a set of scripts (*written in bash, fortran, python*):

1. **Pre-processing** (with operators of cdo, Climate Data Operators, software) gsv_extracted **modeled data** (over latitude-longitude domain) for selected meteorological variables into hourly time-slots/slices (00...23 UTCs);
2. **Extracting** (using sql with odp_api software) selected meteorological variables at hourly time-slices for **observations (from ODB) from**: (i) ground-based **synoptic** stations at fixed geographical latitude-longitude locations on surface; (ii) **radiosounding stations and satellites** at fixed latitude-longitude locations on surface and on multiple pressure levels;
3. **Extracting and applying appropriate observation operator** (using cdo operators - for synop; & expecting/using polytope - for radiosounding and satellite) **to modeled data** for the same time-slices for same selected meteorological variables into corresponding locations/points of (synop, radiosounding, satellite) observations, **with end result - augmented ODB**;
4. **Calculating/ producing relevant statistics for monitoring** such as quantile rank histogram statistics and plots.

Observation Data currently available for OBSALL Apps

Observational data **with a free and unrestricted license** have been collected for initial development and testing purposes in ODB format, consisting of the following types:

- **SYNOP** - surface synoptic observations from the open data server of the Finnish Meteorological Institute covering Finland. The dataset covers the standard meteorological quantities (air temperature, wind, precipitation, pressure, etc.) from about 80+ stations reporting every hour during 2010 - 2021 (**12 years**).
- **TEMP** - openly available radio-soundings for air temperature, wind characteristics, relative humidity and geopotential from the RHARM dataset from Jan 1978 to Apr 2020 (**42 years**), typically at 00 and 12 UTCs. Periods available vary with station.
- **AMSU-A** Channel 5 observations retrieved from the ECMWF MARS archive, covering the period 2010 - 2020 (**11 years**) and all AMSU-A channels covering the period from Dec 2016 to Nov 2017 (**1 year**).

OBSALL Apps: Branch

https://earth.bsc.es/gitlab/digital-twins/de_340/obsall/-/tree/newbranch

ObsAll

- Project information
- Repository
- Files
- Commits
- Branches
- Tags
- Contributor statistics
- Graph
- Compare revisions
- Issues (0)
- Merge requests (0)
- CI/CD
- Security and Compliance
- Deployments
- Packages and registries
- Infrastructure
- Monitor
- Analytics
- Wiki
- Snippets

Digital Twins > DE_340 > ObsAll > Repository

Merge branch 'branch-repo-adds' into 'newbranch'
 Alexander Mahura authored just now 2eb14c95

newbranch ▾ obsall / + ▾ History Find file Web IDE Clone ▾

Name	Last commit	Last update
FIGSEXAMPLES	updated destine_obs_satellite_figs.jpg and main_synop.sh	1 day ago
RADSOUND	updated scripts for RADSOUND	3 minutes ago
SATELLITE	updated README-file and set_env.sh	1 day ago
SYNOP	Merge branch 'newbranch' into 'branch-repo-adds'	1 day ago
README.md	Merge branch 'branch-repo-adds' into 'newbranch'	1 hour ago
request_radsound.yml	updated yml-files used fro synop and radsound	36 minutes ago
request_synop.yml	updated yml-files used fro synop and radsound	36 minutes ago
run_obsall.py	updated run_obsall.py with SATELLIET Part	5 days ago
run_obsall.sh	add updated script for run_obsall.sh	6 days ago

README.md

OBSALL Apps

This repository contains the scripts related to the Observation Apps - components for ground-based/surface (SYNOP), radiosounding (TEMP), and satellite (AMSU-A) obs data) of the Climate Adaptation Digital Twin (Climate DT). The development is done in the frameworks of the Destination Earth initiative.

Description

The OBSALL application is currently being developed as a set of scripts (written in `bash`, `fortran`, `python`):

- for pre-processing (with operators of `cdo`, Climate Data Operators, software) gsv_extracted modeled data (over latitude-longitude domain) for selected meteorological variables into hourly time-slices (00...23 UTCs);
- for extracting (using `sql` with `odp_api` software) selected meteorological variables at hourly time-slices for observations from: (i) ground-based synoptical stations at fixed at the surface geographical locations/points (i.e., 2D: latitude, longitude); (ii) vertical radiosounding of the atmosphere at changing locations/points (3D: latitude, longitude, single pressure level), and (iii) satellite at changing locations/ points (3D: latitude, longitude, multiple pressure levels);

OBSALL Apps: README.md

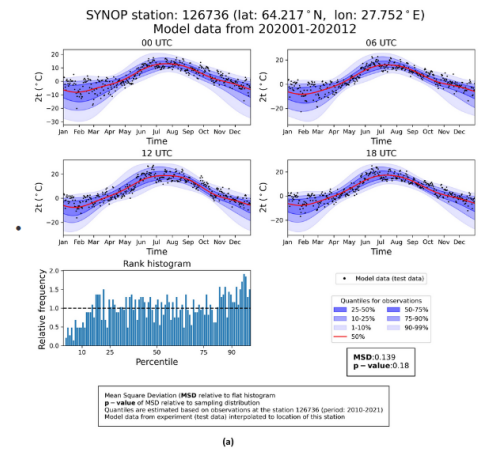
https://earth.bsc.es/gitlab/digital-twins/de_340/obsall/-/blob/newbranch/README.md

Part SYNOP (ground-based/surface observations)

1. `main_synop.sh` - main bash-script to run Apps for observational data for
2. `gsv_mod_data.sh` - bash-script to run pre-processing (using `cdop` operators) of modelled data extracted with `gsv` interface over global domain for air temperature variables (measured at synoptical station) into hourly time-slices and saving
3. `synop_obs.sh` - bash-script to run reading and extraction (using `sql` with `odb_api` software) of above-mentioned air temperatures at 12 hourly time-slices for radiosounding observations at surface and at altitude at geographical (latitude, longitude, pressure level) locations of radiosounding stations over selected geographical domain (or over the globe), calculating difference between two temperatures, and saving (temporary) to separate nc-files at 00 and 12 UTC time-slices.
4. `synop_mod.sh` - bash-script to run reading, extraction and interpolation (using `polytope`) of modelled data for air temperature at surface and at altitude at the same time-slices to geographical (latitude, longitude, pressure level) locations of radiosounding stations, calculating difference between two temperatures, and saving (temporary) to dat-file, and adding (using `import` with `odb_api`) such data to ODB.
5. `graph_mod_obs.py` - python-script to run an internal self-control in calcu
6. `synop_stats.sh` - bash-script to run (additional `bash`, `fortran`, `pyth`

- (i) rank histograms for all synop stations for 00, 06, 12 and 18 UTCs (with `produce_rank_histograms_all_stations.sh` using `rank_histogram_one_station.f95` and `sql/import` with `odb_api`),
- (ii) standard plots for each synop station (with `produce_standard_plots_all_stations.sh` using `plot_quantiles_rankhist.py` and `sql` with `odb_api`; see an example of such plot in **Figure 1a**), and
- (iii) summary rank histograms for all stations (`summary_rank_histograms_all_stations.sh` using `rank_histogram_summary_statistics.f95`, `plot_p_values_map.py`, `plot_rank_hist_sum_all_stations.py` and `sql` with `odb_api`; see examples of such plots in **Figure 2b** and **Figure 2c**).

- Note, needed input includes: list of synop stations with geographical coord and rank histogram bootstrap mean square deviation (MSD) values.

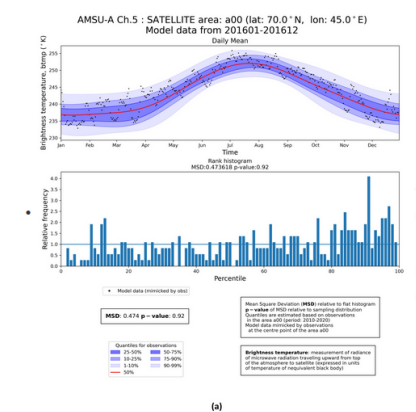


Part AMSU-A (satellite observations)

1. `main_amsua.sh` - main bash-script to run Apps for observational functions, and required modules to be loaded and used in the work
2. `gsv_mod_data.sh` - bash-script to run pre-processing (using `cdop` operators) of modelled data extracted with `gsv` interface over global domain. The "model"-levels (note, workflow will initially be demonstrated on a expanded using `polytope`) and preprocessed data saved in (ter `c_lwc` - specific liquid water content (246); and 2D variables are: pressure (134), 10u and 10v - U and V components of wind speed (Z - geopotential (for orography) (129).
3. `amsua_obs.sh` - bash-script to prepare a metadata file for comp of climate simulation: (i) reality following data assimilation-like, a metadata from AMSU-A observations in the vicinity of the selecte compared to observations only in climatological terms, so metad
4. `amsua_mod.sh` - bash-script to run the Radiance Simulator format), which (using `convert_netcdf2txt.py`) is converted to
5. `amsua_stats.sh` - bash-script to run (additional `bash`, `fortran`

- (i) rank histograms for all locations-areas from satellite observatio `produce_rank_histograms_all_locations.sh` using `rank_histogram_one_station.f95` and `sql/import` with `odb_api`; see an example of such plot in **Figure 2a**), and
- (ii) standard plots for each location-area from satellite observatio and `sql` with `odb_api`; see an example of such plot in **Figure 2a**), and
- (iii) summary rank histograms for locations-areas from satellite of `rank_histogram_sumstats_locations.f95`, `plot_p_value` such plots in **Figure 3b** and **Figure 3c**).

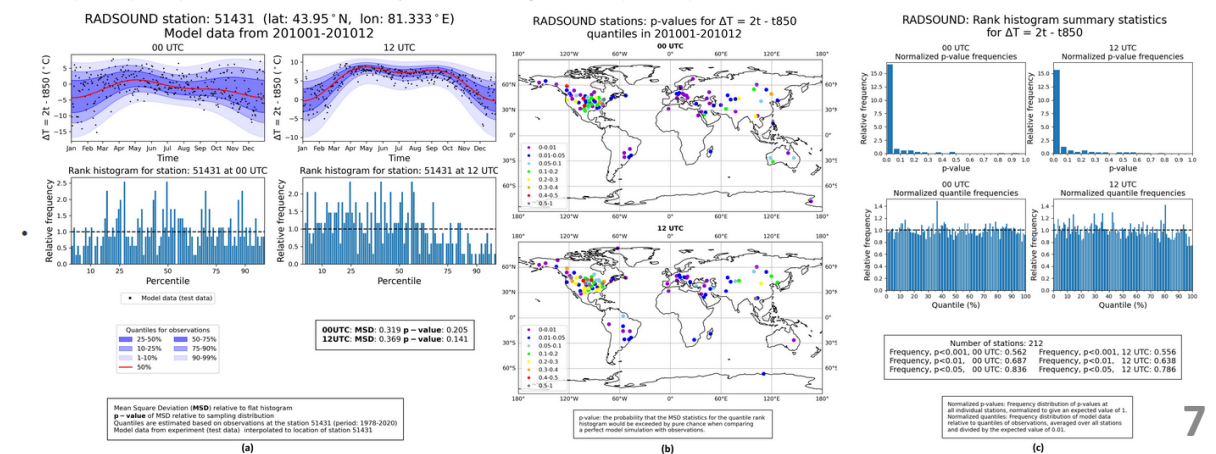
- Note, needed input includes: list of locations-areas from satellite at sfc-, pl-, and o2d-levels, pre-computed quantiles as a function



Part TEMP (radiosounding observations)

1. `main_radsound.sh` - main bash-script to run Apps for observational data for radiosounding.
2. `gsv_radsound_mod_data.sh` - bash-script to run pre-processing (using `cdop` operators) of modelled data extracted with `gsv` interface over global domain for air temperature at 2 metre (T2) and at 850 hPa pressure level (T850) into time-slices at 00 and 12 UTCs, calculating difference (T2-T850) between these two temperatures, and saving (temporary) to separate nc-files at 00 and 12 UTC time-slices.
3. `radsound_obs.sh` - bash-script to run reading and extraction (using `sql` with `odb_api` software) of above-mentioned air temperatures at 12 hourly time-slices for radiosounding observations at surface and at altitude at geographical (latitude, longitude, pressure level) locations of radiosounding stations over selected geographical domain (or over the globe), calculatig difference between two temperatures, and saving (temporary) to dat-file.
4. `radsound_mod.sh` - bash-script to run reading, extraction and interpolation (using `polytope`) of modelled data for air temperature at surface and at altitude at the same time-slices to geographical (latitude, longitude, pressure level) locations of radiosounding stations, calculatig difference between two temperatures, and saving (temporary) to dat-file, and adding (using `import` with `odb_api`) such data to ODB.
5. `radsound_stats.sh` - bash-script to run (additional `bash`, `fortran`, `python` scripts) calculating/producing:

- (i) rank histograms for all radiosounding stations for 00 and 12 UTCs (with `produce_rank_histograms_all_stations.sh` using `rank_histogram_one_station.f95` and `sql/import` with `odb_api`),
 - (ii) standard plots for each radiosounding station (with `produce_standard_plots_all_stations.sh` using `plot_quantiles_rankhist.py` and `sql` with `odb_api`; see an example of such plot in **Figure 2a**), and
 - (iii) summary rank histograms for all radiosounding stations (`summary_rank_histograms_all_stations.sh` using `rank_histogram_summary_statistics.f95`, `plot_p_values_map.py`, `plot_rank_hist_sum_all_stations.py` and `sql` with `odb_api`; see examples of such plots in **Figure 2b** and **Figure 2c**).
- Note, needed input includes: list of radiosounding stations with geographical coordinates, modelled data at stations coordinates at 2 metre height and at 850 hPa pressure level, pre-computed quantiles as a function of time of year and rank histogram bootstrap mean square deviation (MSD) values.



Installation

In order to install, copy "obsall" repository to your local directory on Lumi HPC using:

- ✓ login to your DestinE account on Lumi HPC
- ✓ mkdir your_local_directory
- ✓ cd your_local_directory

Option 1: *cloning obsall repository to your_local_directory*

- ✓ git clone https://earth.bsc.es/gitlab/digital-twins/de_340/obsall.git
- ✓ cd obsall
- ✓ git checkout newbranch

OR

Option 2: *copying tar-file containing obsall repository to your_local_directory*

- ✓ cp /projappl/project_465000454/ama/ARC_OBSALL/arc_obsall_apps.tar .
- ✓ tar -xvf arc_obsall_apps.tar
- ✓ cd obsall

Required Modules

to set environment in order to run OBSALL Apps on Lumi, execute bash-script containing list of required modules: source load_modules.sh

- ✓ # Load modules for OBSALL Apps
- ✓ module use /project/project_465000454/devaraju/modules/LUMI/23.03/C
- ✓ module load LUMI/23.03
- ✓ module load partition/C
- ✓ module load PrgEnv-gnu
- ✓ module load ecCodes/2.32.0-cpeCray-23.03.lua
- ✓ module load odb_api/0.18.1-cpeCray-23.03.lua
- ✓ module load python-climagedt/3.11.3-cpeCray-23.03.lua
- ✓ module load pyfdb/0.0.2-cpeCray-23.03.lua
- ✓ module load cray-hdf5/1.12.2.3
- ✓ module load cray-netcdf/4.9.0.3
- ✓ module load rtov/13.2
- ✓ module load radsim/3.2
- ✓ module load fdb/5.11.94-cpeCray-23.03.lua
- ✓ module load eckit/1.25.0-cpeCray-23.03.lua
- ✓ module load metkit/1.11.0-cpeCray-23.03.lua

+ *later upgrade OBSALL Apps for Polytope tool (required for SATELLITE observations) in Phase-2*

Data Requests (yaml-file)

file **request*.yaml** (when with all variables included in same request, as **mother_request.yaml**) is needed for gsv extracting selected modeled data

Data request (sfc): SYNOP Part (*synop)

At sfc-level: 2t & other available meteorological variables

Data request (sfc+pl): TEMP Part (*radsound)

At sfc-level: 2t

At pl-level: t (at 850 hPa)

Data request (sfc+pl+o2d): AMSU-A Part (*satellite)

At sfc-level: 2t, 2d, sp, skt, 10u, 10v & z, lsm (n/a – IFS; tmp. in static grib file)

At pl-levels: t, q, clwc, z (at 1000, 925, 850, 700, 600, 500, 400, 300, 250, 200, 150, 100, 70, 50, 30, 20, 10, 5, 1 hPa) – 19 pl-levels

At o2d-level: avg_siconc – from NEMO model (or ci – from IFS, n/a now)

How to Run Apps

How to run: runs script `run_obsall.py` is used to execute OBSALL Apps: `python run_obsall.py`

```
#!/scratch/project_465000454/devaraju/SW/LUMI-23.03/C/python-climatedt/bin/python
```

```
# OBSALL Apps (3 parts: SYNOP, TEMP, AMSU-A observations)
```

```
# Import required libraries
```

```
import sys
```

```
import subprocess
```

```
## --- Processing ground-based observations (SYNOP)
```

```
print('*****')
```

```
print('DestinE Climate Digital Twin - OBSALL Apps')
```

```
print('--- Processing ground-based observations (SYNOP)')
```

```
print('*****')
```

```
command_synop_run = "cd SYNOP; pwd; ./main_synop.sh; exit 0"
```

```
subprocess.run(command_synop_run, shell=True, check=True, executable="/bin/bash")
```

```
## --- Processing radiosounding observations (TEMP)
```

```
print('*****')
```

```
print('DestinE Climate Digital Twin - OBSALL Apps')
```

```
print('--- Processing radiosounding-based observations (TEMP)')
```

```
print('*****')
```

```
command_radsound_run = "cd RADSOUND; pwd; ./main_radsound.sh; exit 0"
```

```
subprocess.run(command_radsound_run, shell=True, check=True, executable="/bin/bash")
```

```
## Processing satellite observations (AMSU-A)
```

```
print('*****')
```

```
print('DestinE Climate Digital Twin - OBSALL Apps')
```

```
print('--- Processing satellite-based observations (AMSU-A)')
```

```
print('*****')
```

```
command_satellite_run = "cd SATELLITE; pwd; ./main_amsua.sh; exit 0"
```

```
subprocess.run(command_satellite_run, shell=True, check=True, executable="/bin/bash")
```

```
sys.exit(0)
```

Demonstration (with illustrations) of OBSALL Apps running for SYNOP, TEMP, and AMSU-A Parts as well as so-called Monitoring for these 3 Parts

*on Lumi's DestinE project account
for a limited size datasets of synop, radiosounding, and satellite observations*

LUMI

About LUMI

Get Started

News and Articles

Events and Training

Newsletters

User Support

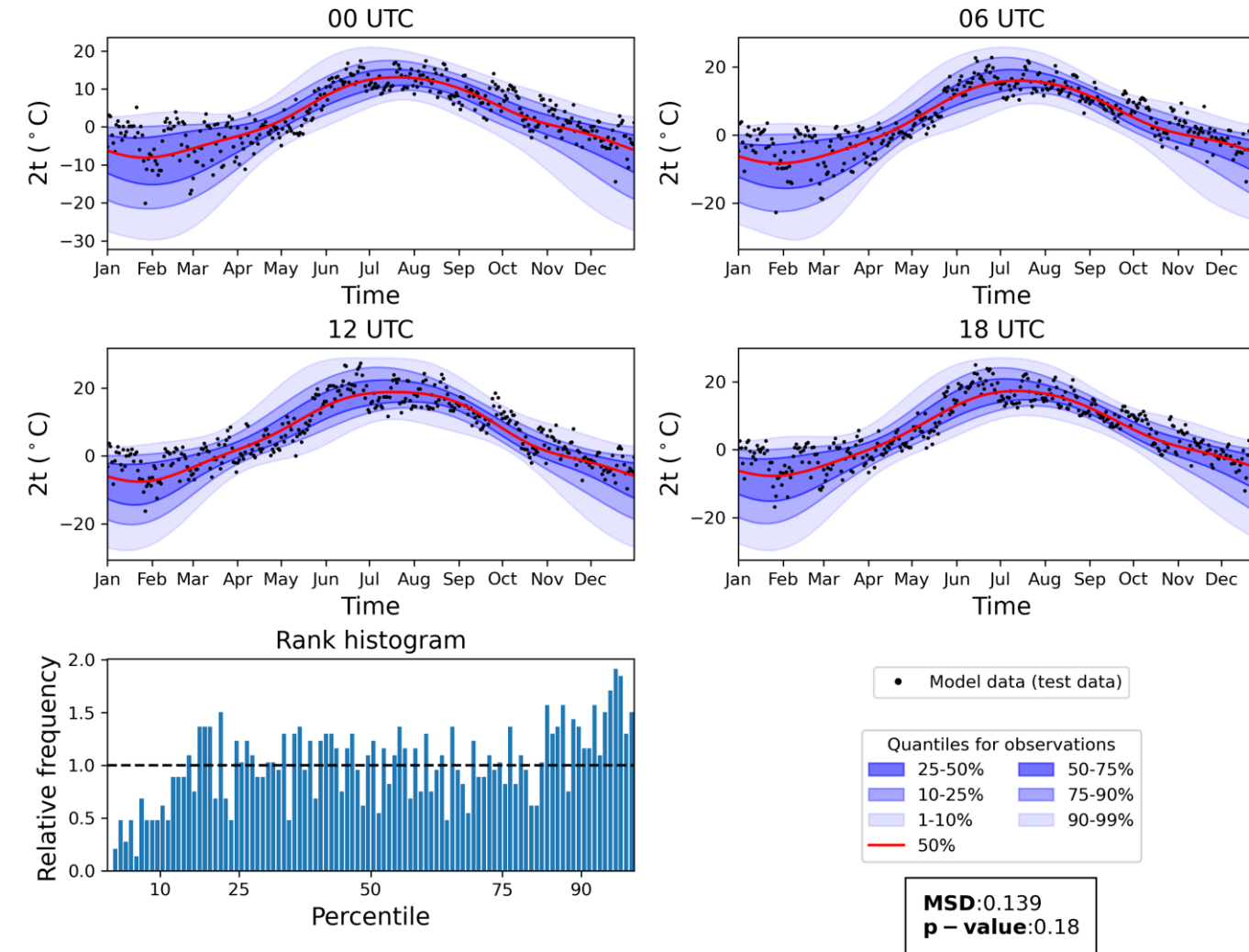


source - <https://www.lumi-supercomputer.eu>

Monitoring for SYNOP (1)

SYNOP station: 126736 (lat: 64.217° N, lon: 27.752° E)

Model data from 202001-202012



Graphical/ figures output files:
obsall/SYNOP/STATS/figures/

**Plots for 2 metre air temperature (2t)
for period of 1 Jan - 31 Dec 2020:**

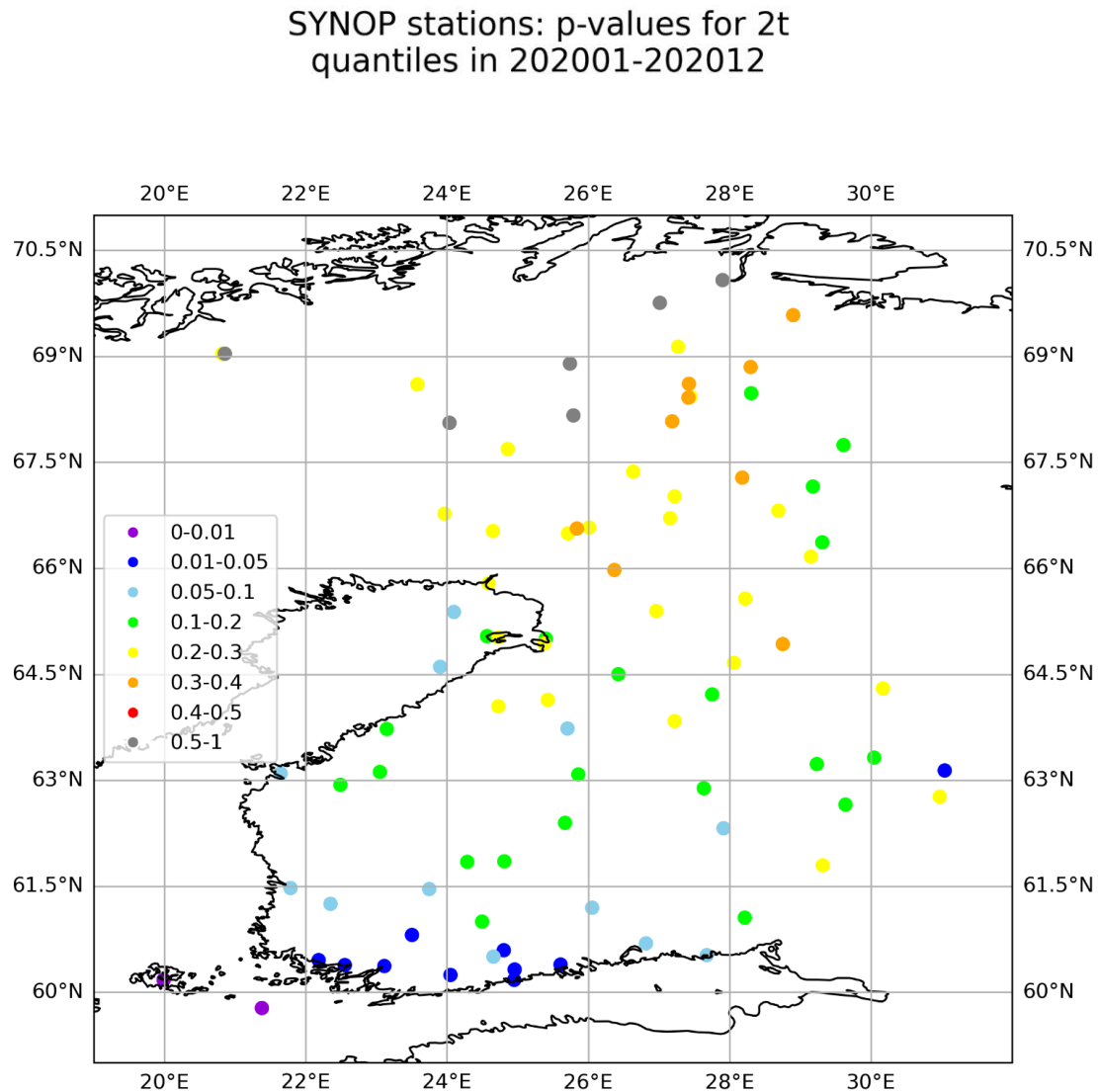
Figure 1a - Quantiles/ time series data (00, 06, 12, and 18 UTCs) and rank histogram (combination of 00, 06, 12, and 18 UTCs) for synop station 126736.

Monitoring for SYNOP (2)

Graphical/ figures output files:
obsall/SYNOP/STATS/figures/

**Plots for 2 metre air temperature (2t)
for period of 1 Jan - 31 Dec 2020:**

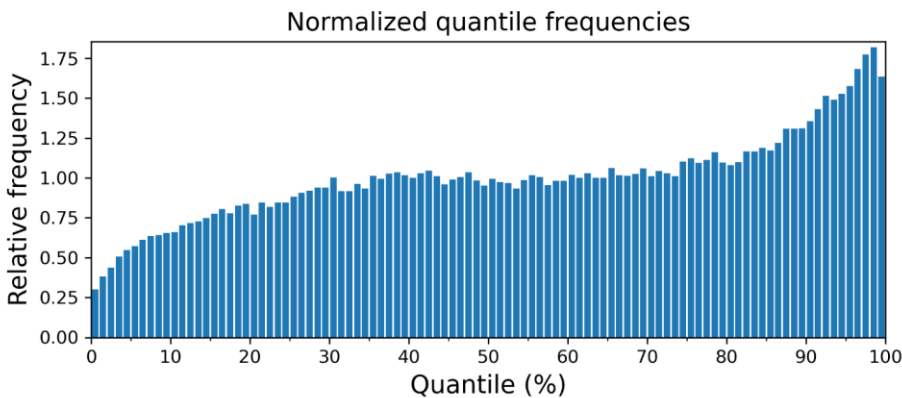
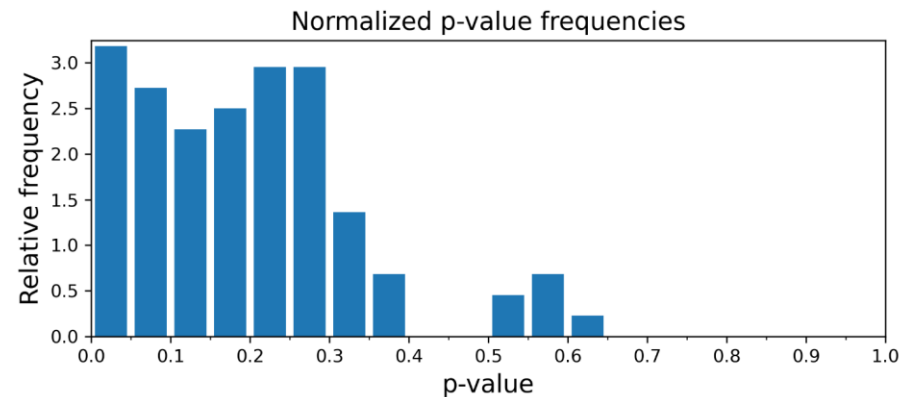
Figure 1b - p-values for 2t quantiles.



p-value: the probability that the MSD statistics for the quantile rank histogram would be exceeded by pure chance when comparing a perfect model simulation with observations.

Monitoring for SYNOP (3)

Graphical/ figures output files:
obsall/SYNOP/STATS/figures/



Number of stations: 88
Frequency, $p < 0.001$: 0.023
Frequency, $p < 0.01$: 0.045
Frequency, $p < 0.05$: 0.159

Normalized p-values: Frequency distribution of p-values at all individual stations, normalized to give an expected value of 1.
Normalized quantiles: Frequency distribution of model data relative to quantiles of observations, averaged over all stations and divided by the expected value of 0.01.

**Plots for 2 metre air temperature (2t)
for period of 1 Jan - 31 Dec 2020:**

Figure 1c - Rank histogram summary statistics:
normalized p-value frequencies and normalized
quantile frequencies.

Monitoring for TEMP (1)

Graphical/ figures output files:
obsall/RADSOUND/STATRS/figures/

**Plots for difference in air
temperatures at 2 metre and at
850 hPa pressure level (2t - t850)
for period of 1 Jan - 31 Dec 2010:**

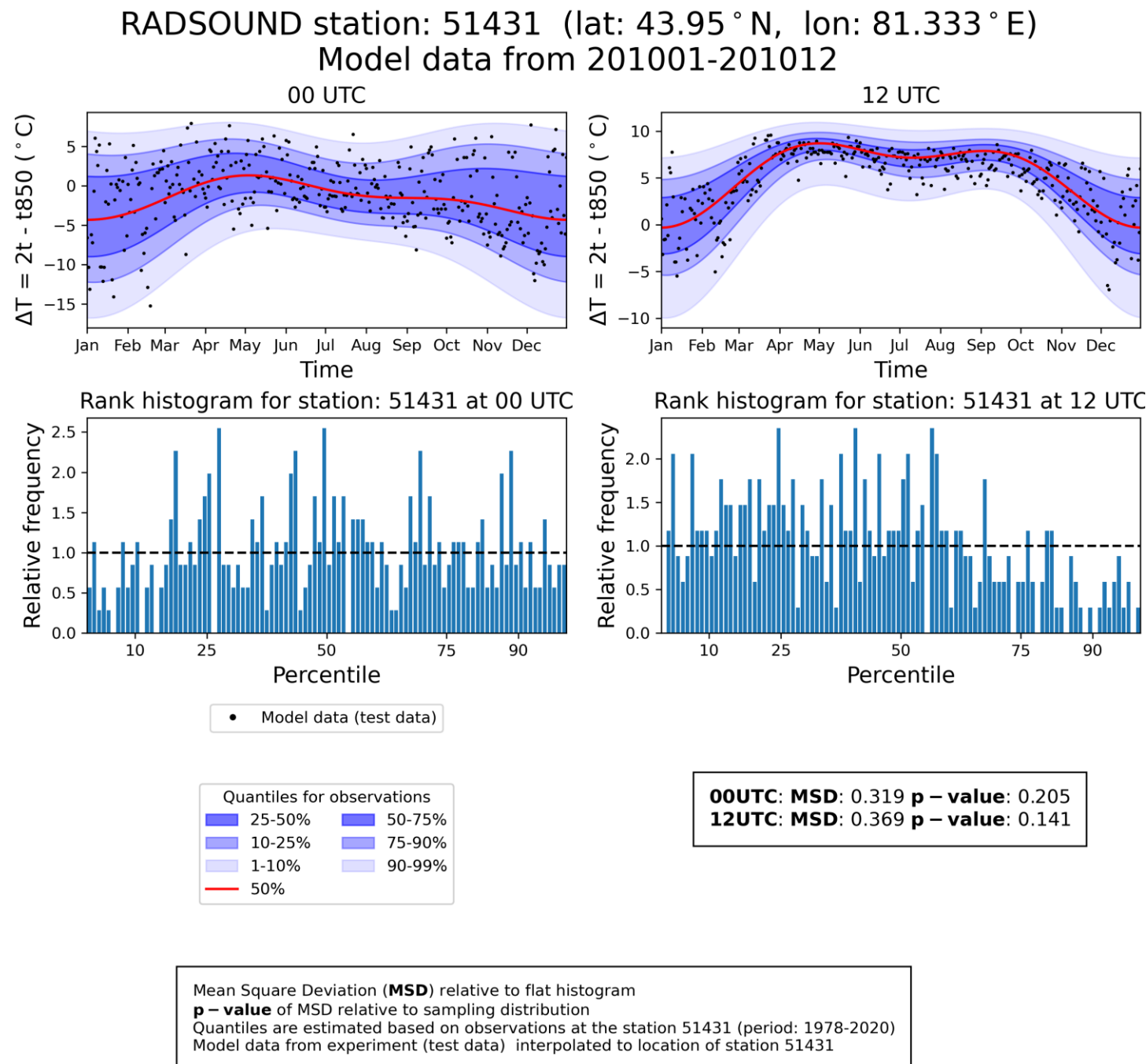
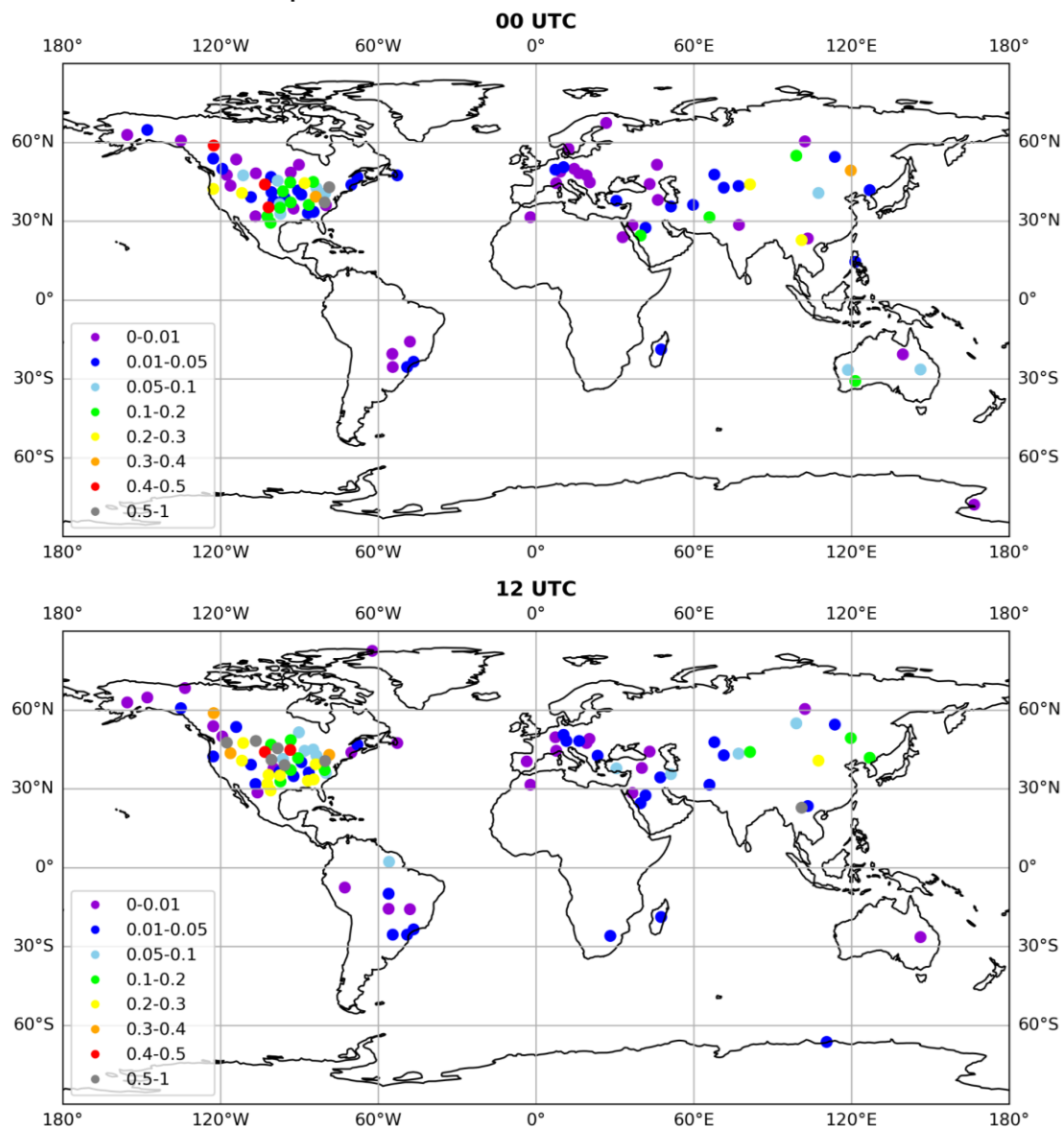


Figure 2a - Quantiles/ time series data and rank histogram (00 and 12 UTCs) for radiosounding station 51431

RADSOUND stations: p-values for $\Delta T = 2t - t850$
quantiles in 201001-201012



p-value: the probability that the MSD statistics for the quantile rank histogram would be exceeded by pure chance when comparing a perfect model simulation with observations.

Monitoring for TEMP (2)

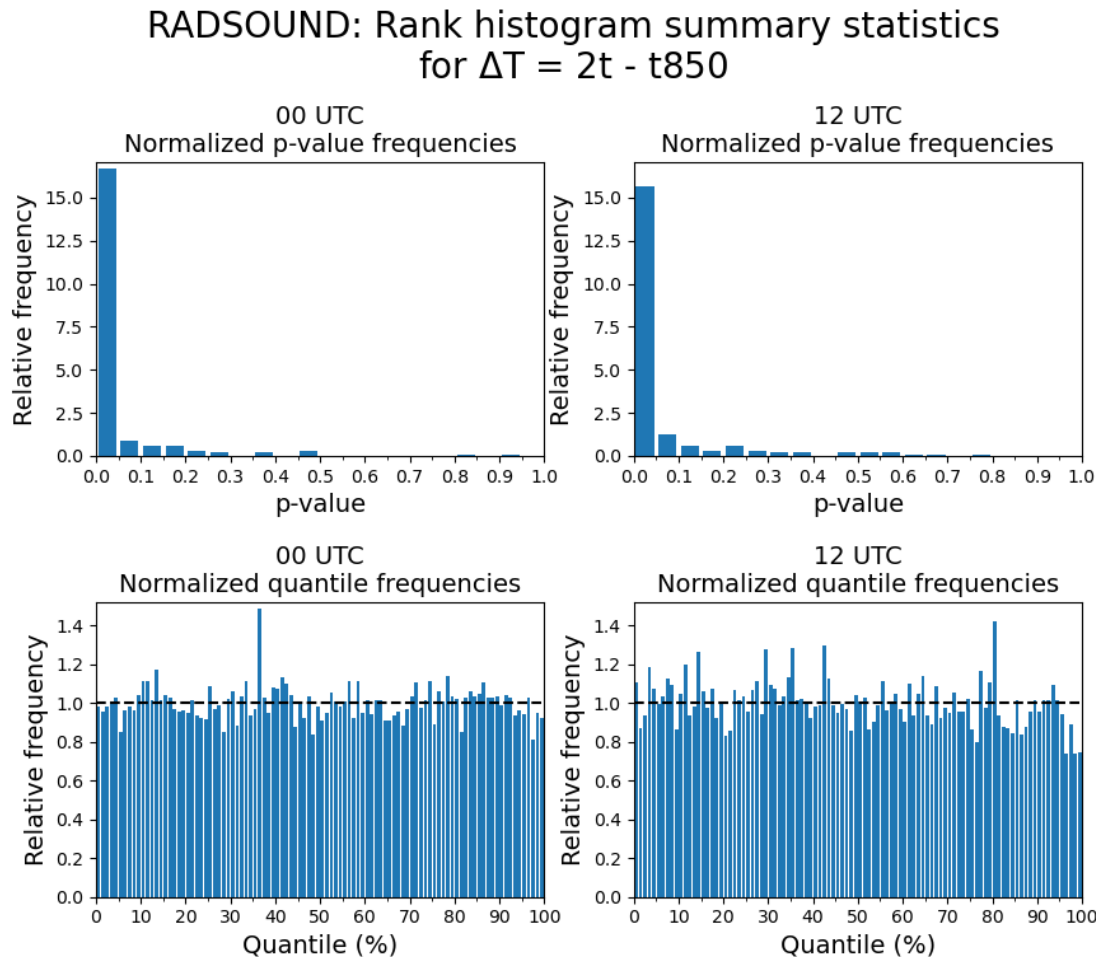
Graphical/ figures output files:
obsall/RADSOUND/STATRS/figures/

**Plots for difference in air
temperatures at 2 metre and at
850 hPa pressure level ($2t - t850$)
for period of 1 Jan - 31 Dec 2010:**

**Figure 2b - p-values for $2t - t850$
quantiles.**

Monitoring for TEMP (3)

Graphical/ figures output files:
obsall/RADSOUND/STATRS/figures/



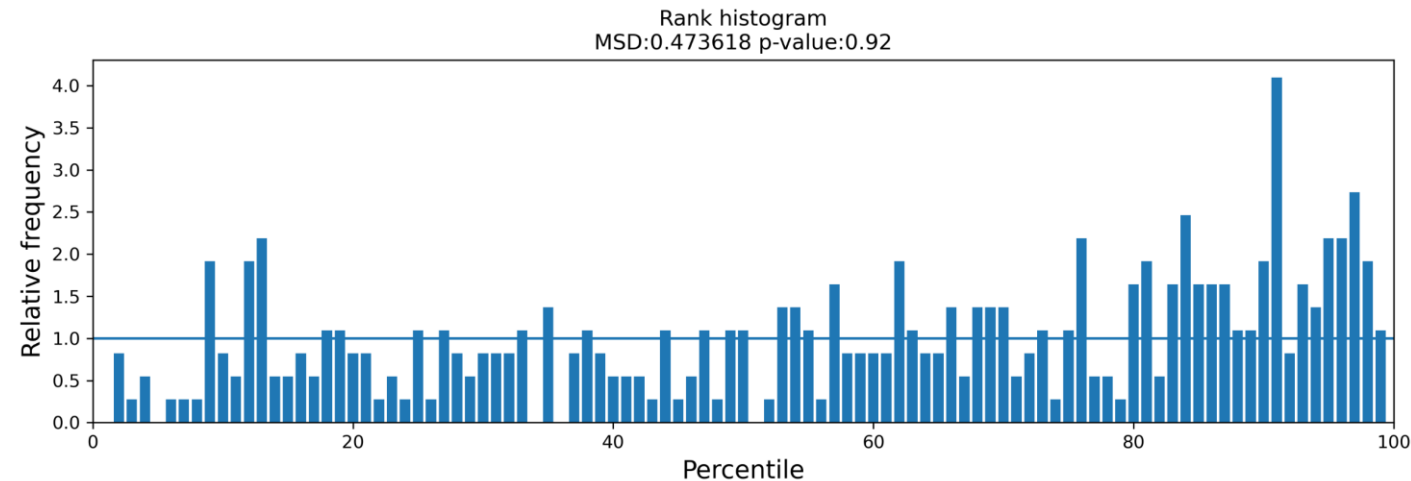
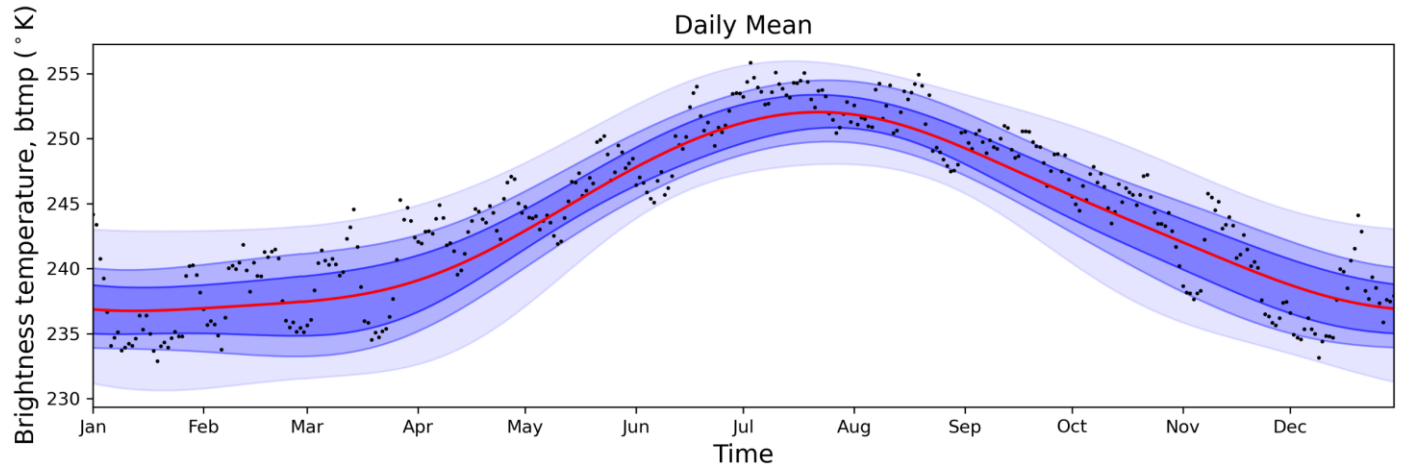
Number of stations: 212	
Frequency, $p < 0.001$, 00 UTC: 0.562	Frequency, $p < 0.001$, 12 UTC: 0.556
Frequency, $p < 0.01$, 00 UTC: 0.687	Frequency, $p < 0.01$, 12 UTC: 0.638
Frequency, $p < 0.05$, 00 UTC: 0.836	Frequency, $p < 0.05$, 12 UTC: 0.786

Normalized p-values: Frequency distribution of p-values at all individual stations, normalized to give an expected value of 1.
Normalized quantiles: Frequency distribution of model data relative to quantiles of observations, averaged over all stations and divided by the expected value of 0.01.

**Plots for difference in air
temperatures at 2 metre and at
850 hPa pressure level ($2t - t850$)
for period of 1 Jan - 31 Dec 2010:**

Figure 2c - Rank histogram summary statistics for 00 and 12 UTCs: normalized p-value frequencies and normalized quantile frequencies.

AMSU-A Ch.5 : SATELLITE area: a00 (lat: 70.0 ° N, lon: 45.0 ° E)
Model data from 201601-201612



• Model data (mimicked by obs)

MSD: 0.474 p – value: 0.92

Quantiles for observations

- 25-50%
- 10-25%
- 1-10%
- 50-75%
- 75-90%
- 90-99%
- 50%

Mean Square Deviation (MSD) relative to flat histogram
p – value of MSD relative to sampling distribution
Quantiles are estimated based on observations
in the area a00 (period: 2010-2020)
Model data mimicked by observations
at the centre point of the area a00

Brightness temperature: measurement of radiance
of microwave radiation traveling upward from top
of the atmosphere to satellite (expressed in units
of temperature of nequivalent black body)

Monitoring for AMSU-A (1)

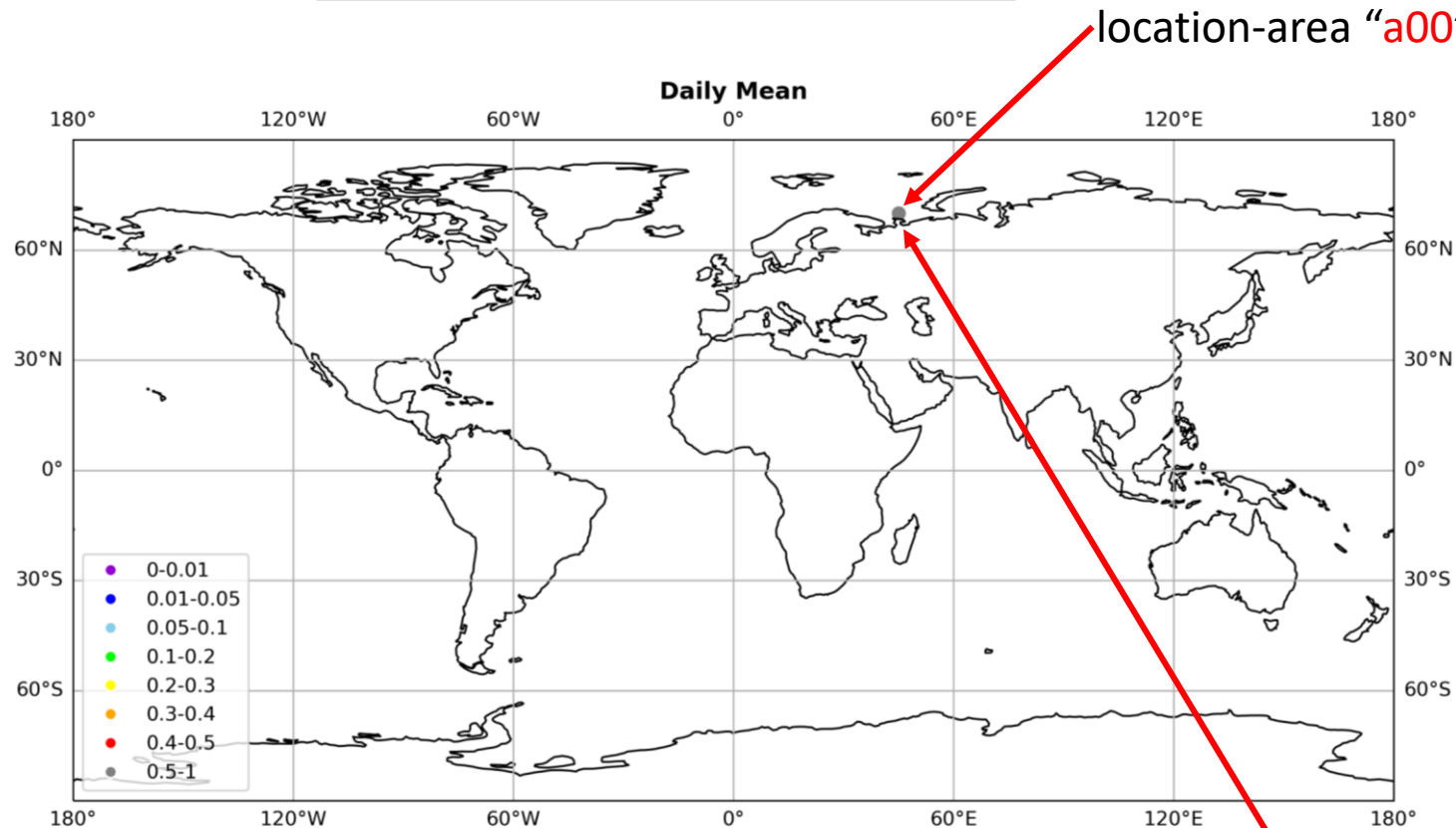
Graphical/ figures output files:
obsall/SATELLITE/STATST/figures/

**Plots for brightness temperature
for period of 1 Jan - 31 Dec 2016:**

Figure 3a – Quantiles/ time series data
and rank histogram for daily mean
brightness temperature for area /a00;
about 100x100 km with a center at 70N,
45E).

AMSU-A Ch.5 : SATELLITE areas:
p-values for brightness temperature (btmp)
quantiles in 201601-201612

p – value: probability that the MSD statistics for the quantile rank histogram would be exceeded by pure chance when comparing a perfect model simulation with observations.



Brightness temperature: Measure of upwelling microwave radiation at top of the atmosphere (expressed in units of temperature of equivalent black body)

Monitoring for AMSU-A (2)

Graphical/ figures output files:
obsall/SATELLITE/STATST/figures/

**Plots for brightness temperature
for period of 1 Jan - 31 Dec 2016:**

Figure 3b – p-values for brightness
temperature quantiles for areas.

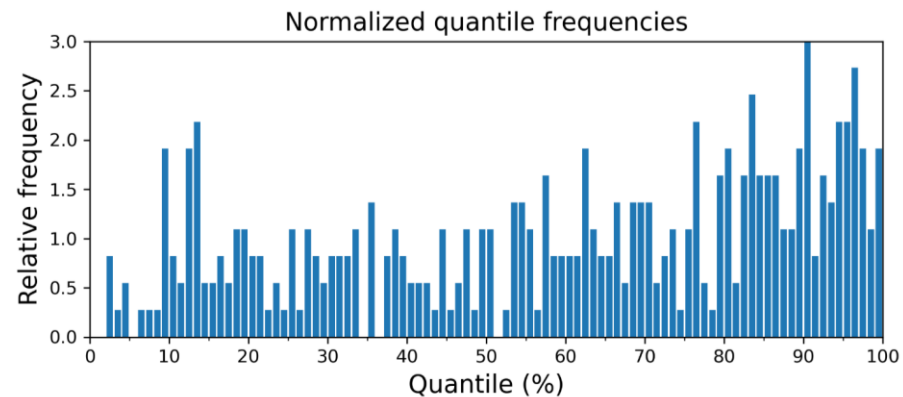
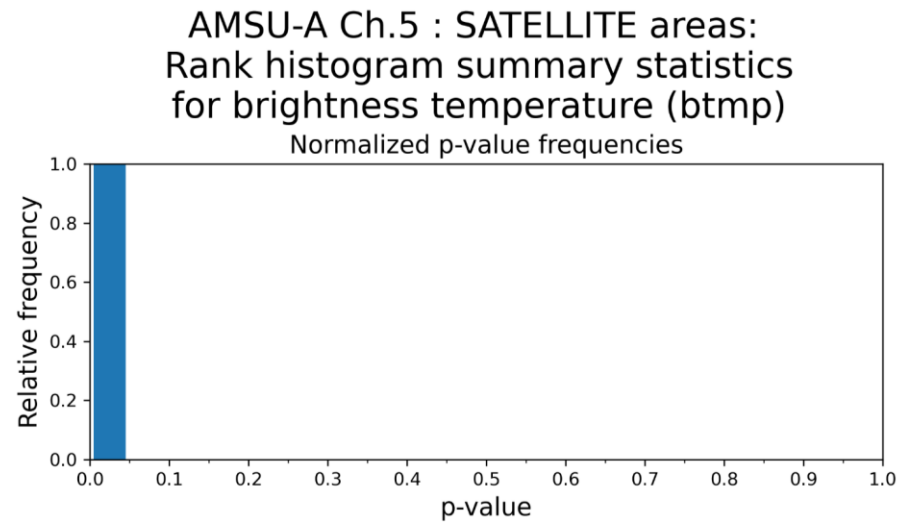
The plotted dot as a colored circle represents the central point (70 N, 45 E) of the area "a00", which is situated within boundaries of [69.5, 70.5] N and [43.5, 46.5] E.

Monitoring for AMSU-A (3)

Graphical/ figures output files:
obsall/SATELLITE/STATST/figures/

**Plots for brightness temperature
for period of 1 Jan - 31 Dec 2016:**

Figure 3c – Rank histogram summary statistics for brightness temperature for areas: normalized p-value frequencies and normalized quantile frequencies



Number of areas: 1
Frequency, $p < 0.001$: 1.0
Frequency, $p < 0.01$: 1.0
Frequency, $p < 0.05$: 1.0

Normalized p – values: Frequency distribution of p-values at all individual stations, normalized to give an expected value of 1.
Normalized quantiles: Frequency distribution of model data relative to quantiles of observations, averaged over all stations and divided by the expected value of 0.01.

Brightness temperature: Measure of upwelling microwave radiation at top of the atmosphere (expressed in units of temperature of equivalent black body)

Authors are Thankful

- *for advise on issues with autosubmit, gsv extraction and other tools/apps, integration in workflow, etc. to BSC colleagues:* **Miguel Castrillo, Francesc Roura, Iker Gonzalez, Marvin Axness, Aina Gaya i Àvila**
- *for constructive comments, remarks, suggestions to reviewers:* **Steger Christian (DWD), Bruno Kinoshita (BSC), Miguel Castrillo (BSC) & Aina Gaya i Àvila (BSC), Pablo Ortega (BSC), Paolo Davini (CNR), Jost Hardenberg (POLITO), Natalia Nazarova (POLITO)**
- *for LUMI related advises/help to:* **Juha Lento (CSC), Devaraju Narayanappa (CSC), Henrik Nortamo (CSC), Sami Salonen (CSC), Roman Nuterman (UCPH), Thomas Geenen (ECMWF)**
- *for polytope tool info/examples sharing -* **Matthew Griffith, Nils Wedi (ECMWF)**