

VALIDATION OF RUC

ASSIMILATION OF GPS

MOISTURE INFORMATION

Mohamed Zied SASSI
INM, Tunisia

ARSO, Slovenia

October 2015

OUTLINE

| | |
|--|----|
| Introduction | 3 |
| I. Model ZTD computation | 4 |
| II. White list computation..... | 5 |
| 1) Production of OBSOUL files..... | 5 |
| 2) BATOR and screening run..... | 5 |
| 3) Extraction of first guess departures..... | 6 |
| 4) White list generation..... | 7 |
| 5) Selection criteria..... | 8 |
| 6) Bias corrected obsouls..... | 11 |
| III. Experiments and settings..... | 12 |
| A. GPSREF | 12 |
| B. GPSOL | 12 |
| C. SLOVEX..... | 13 |
| D. VARBEX | 13 |
| IV. Scores and comparison | 15 |
| 1) Surface pressure | 15 |
| 2) Relative humidity..... | 16 |
| 3) Surface temperature..... | 17 |
| 4) Cloud cover..... | 18 |
| 5) Cumulative rain..... | 19 |
| 6) Wind speed..... | 20 |
| 7) Wind direction..... | 21 |
| Conclusions | 22 |
| References | 23 |

Introduction

Regarding the increasing number of available data and the results shown beneficial during the last years, more attention has been accorded to the GPS Zenith Tropospheric Delay (ZTD) data assimilation. GPS ZTD data improves the water vapour observation over the GNSS (Global Navigation Satellite System) sites. In fact, The total ZTD is composed of a dry delay which is proportional to the local pressure and a wet delay which is proportional to the water vapour.

Regarding the better performance of NWP systems in pressure modeling compared to water vapour, the use of GPS ZTD data can improve the observation of the water vapour which is a very important parameter in precipitation forecast and atmospheric dynamics.

The idea was to investigate during this stay the impact of GPS data on Slovenian operational forecast system. Therefore, a set of available EGVAP (The EUMETNET GPS Water Vapour Programme) and Slovenian GPS data has been used in order to select trusted sites that will be considered during this stay.

Two methods have been considered namely static method and variational bias correction method. During the static method, non-variable bias correction parameters will be calculated and integrated to the observational files before assimilation, however, during the variational bias correction experiment, bias correction parameters will be initialized to zero but we will introduce some modifications in order to calculate and update them at every cycle of assimilation.

The last part of this report includes scores calculation and comparison between the different experiments and methods. Since GPS data has more impact on medium and heavy rain cases (EUREF Symposium, 200506), one heavy rain situation is also described.

I. Model ZTD computation

The ZTD parameter is generally presented as being the integral of refractivity over a vertical column of neutral atmosphere, commonly written as:

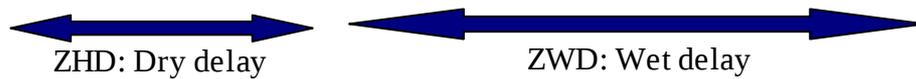
$$ZTD = 10^{-6} \int_{site}^{top} [k_1 R_d \rho_d + (k_2 + k_3/T) R_\omega \rho_\omega] \delta_z$$

where:

| | |
|-----------------------|---------------------------------|
| ρ : Density; | R : Gas constant; |
| z : Geometric high; | k : Refractivity of the lower |
| d : Dry; | neutral atmosphere for radio |
| ω : Wet; | waves in the GPS range; |

For NWP models, parameters as pressure p , specific humidity q and temperature T are useful to compute ZTD, the equation becomes as follows:

$$ZTD = 10^{-6} \int_0^{P_{site}} k_1 \frac{R_d}{g} \delta_p + 10^{-6} \int_0^{P_{site}} \frac{R_d}{g\epsilon} q ((k_2 - k_1\epsilon) + k_3/T) \delta_p$$



where:

| |
|--|
| ϵ : Ratio of the molecular weight of water vapour to that of dry air; |
| g : Local vertical acceleration; |

The ZHD component is proportional to the local pressure, however, the ZWD component is proportional to the amount of water vapour above the GPS site.

On the other hand, the lower part of the integral was set to zero since its contribution is negligible above 80km (Vedel et al., 2001).

II. White list computation

Unlike satellite and ground data, the increasing number of sites conducted to a different approach which assumes that all GPS ZTD data are passive except data provided from trusted list of sites commonly named a white list. During a static bias correction method, the white list includes a set of trusted stations with corresponding calculated values for bias correction, however, during a variational bias correction method, these values are set to zero and are calculated further during the assimilation process. During this stay, both methods were applied. The following part summarizes the white list calculation:

1) **Production of OBSOUL files:**

The calculation of white list was based on EGVAP and Slovenian merged GPS obsoul data. As a first step, the EGVAP obsoul files were produced from “COST” files within the “for_gergo_creat_egvap_obsoul.sh” script available in the package “for_gergo.tar” under “/scratch/ms/hu/hu8”. Before processing, the extraction script should be updated in order to take into account the name of data centers as follows:

In line 23:

```
center=`echo $file | cut -c 46-49 | tr 'a-z' 'A-Z'`
```

Then, “obsoul_merge_new.pl” was applied in order to merge the produced osboul files to the Slovenian GPS data obsouls. The final obsoul files were considered in the following steps.

2) **BATOR and screening run:**

This step consists on running a miniature experiment which includes BATOR and screening steps without minimization and is based only on GPS ZTD data and no other observations were considered.

This run was performed over a period of 30days corresponding to the month of July

2015. The modification in the default odb.ecf file was as follows:

In line 63:

```
cp ~/july_merge/OBSOUL_EGVAP_and_SLOVAVG_YYYYMM$DD$HH.dat
OBSOUL_YYYYMM$DD$RR.conv
cp ~/exp/rundir/as$YYYYMM$DD$RR.gpsol_assim/gpsol/assim/analysis/3dvar/obsprep/
obstypes_YYYYMM$DD$RR obstypes_YYYYMM$DD$RR
tar czf observations.tar.gz OBSOUL_YYYYMM$DD$RR.conv
obstypes_YYYYMM$DD$RR
```

3) Extraction of first guess departures:

The output “odb_CCMA” files are then processed in order to extract the necessary data and especially the first guess departures thanks to the script “get_list_ecma_batch_2011_MAY.sh11-merge” under “~/sbc_ztd”. The package “sbc_ztd.tgz” is available under “/scratch/ms/hu/hu8”.

Some updates had to be introduced in the “get_list_ecma_batch_2011_MAY.sh11-merge” script in order to run cy38 experiments with the “odbviewer” command. In fact, the variable “modoro” has to be replaced by “orography”, and add “modsurf” to the modules list, “status.active@body” should be replaced by “report_status.active@hdr” and “status.blacklisted@body” should be replaced by “report_status.blacklisted@hdr”.

The execution command line can be considered as follows:

```
“odbviewer -q 'select statid,date,time,an_sens_obs,lat,lon,stalt,orography,obsvalue,
fg_depar,report_status.active@hdr,report_status.blacklisted@hdr,an_depar from
hdr, body,errstat,modsurf,index' < "."”
```

As output, the script “get_list_ecma_batch_2011_MAY.sh11-merge” produces listing files that contain the extracted fg_departures which are necessary for the white list calculation.

4) White List generation:

The script “select_gpssol.x” which calculates the white list reads the set of files that are listed in the “list_files_slov” file. This list contains the output file from the previous step C. that contains fg departures information. This script uses also some parameters that should be defined in the “gpssol_mod.F90” module in order to personalize the white list selection criteria.

The outputs of this script are divided into the 7 following files:

- 1) stats.txt_slov: This file includes a list of all the input GPS stations;
- 2) stats_active.txt_slov: This file includes only input active stations;
- 3) stats_1.txt_slov, stats_2.txt_slov, stats_3.txt_slov, stats_4.txt_slov: These files include different levels of selection files according to the selection criteria;
- 4) list_gpssol_slov: This file file includes the White List and the bias correction parameter calculated for every selected station.

Example of the white list:

| | | | | | | | |
|----------|---------|---------|-------|-----|--------------|-------|--------|
| LAMAMETO | 53.9197 | 20.6804 | 158. | 15. | 0.005366078 | 11.51 | 0.0115 |
| DOULSGN1 | 50.1875 | 2.3533 | 67. | 15. | -0.001042003 | 13.16 | 0.0132 |
| ENTZSGN_ | 48.5739 | 7.6438 | 156. | 15. | -0.001120341 | 11.90 | 0.0119 |
| GLRASGN_ | 44.8620 | 4.5264 | 763. | 15. | 0.000621061 | 12.28 | 0.0123 |
| SCDAMETO | 44.8172 | 3.2691 | 1062. | 15. | 0.001091727 | 11.71 | 0.0117 |
| TERSMETO | 53.3898 | 5.2220 | 15. | 15. | -0.001659964 | 15.19 | 0.0152 |
| CT50SGN1 | 49.1656 | 2.7927 | 163. | 15. | -0.002197834 | 12.33 | 0.0123 |
| OSANSGN1 | 42.3456 | 8.6368 | 205. | 15. | 0.006686612 | 11.65 | 0.0116 |
| GIE1SGN_ | 47.6371 | 2.7830 | 166. | 15. | 0.001980725 | 14.74 | 0.0147 |
| HOBUMETO | 53.0775 | 10.4816 | 111. | 15. | 0.006354197 | 12.54 | 0.0125 |
| MARGMETO | 46.1078 | 6.5140 | 474. | 15. | -0.004082245 | 13.88 | 0.0139 |

The white list variables are:

- 1- Name of the data center;
- 2- Station latitude (in degrees);
- 3- Station longitude(in degrees);
- 4- Station altitude;
- 5- Data averaging time;
- 6- Bias correction;
- 7- 10 days stdev (junk)
- 8- Observation error stdev;

5) **Selection criteria:**

The white list calculation is based on some selection criteria that are defined by the user. These criteria are:

- 1- Bias max;
- 2- STDEV max;
- 3- Altitude not constant;
- 4- Stations out of domain;
- 5- Stations altitude above max value;
- 6- Time coverage;
- 7- Analysis center;
- 8- Time averaging;
- 9- Gauss function;

During the current stay, we have set selection criteria parameters as follows:

Table 1: Selection criteria for GPS stations white list computation

| Selection criteria | value |
|--------------------------------------|--------|
| biasmax | 40.E-3 |
| stdevmax | 40.E-3 |
| ndata_min_in_class_gauss (threshold) | 4 |
| nclass_gauss | 70 |
| nconf_gauss | 3 |
| Lonmin | -1 |
| lonmax | 28 |
| latmin | 36 |
| latmax | 55 |
| diff_alt_max | 2000 |
| alt_max | 2500 |
| min_time_coverage | 1 |
| Altitude variation max | 0.1 |
| Longitude variation max | 0.1 |
| Latitude variation max | 0.1 |
| Orography variation max | 0.1 |
| Max Active Stations | 450 |

The Gaussian function is defined as follows:

$$G = \frac{e^{-\frac{1}{2}\left(\frac{Xb-mean}{stdev}\right)^2}}{dstep * ndata * stdev * \sqrt{2\pi}}$$

where: $dstep = (Xmax - Xmin) / nclass_gauss$; $Xmax$ and $Xmin$ being respectively the maximum and the minimum fg departures values.

The stations which do not respect this function within the defined parameters were rejected from the white list.

For the current stay, 49 stations have been rejected as follows given that a station can be rejected for more than one criteria:

| | | |
|------------------------------|-----------------|--------------------------------|
| Rejected counts | : out of | 575 |
| Time_coverage7 | : | 3 |
| Gauss | : | 39 |
| Time averaging | : | 0 |
| Altitude | : | 0 |
| Domain | : | 0 |
| Altitude not constant | : | 2 |
| Analysis center excl | : | 0 |
| Bias above max | : | 11 |
| Stdev above max | : | 11 |
| ----- | | |
| Total | : | 526 (Selected stations) |

Then, more thinning has been applied on remaining stations in order to detect too nearby stations and remove problematic ones based on stdev method.

Finally, **294** stations have been selected making thus the final white list.

Concerning the Slovenian GPS network stations, 29 stations were introduced but only 16 were white listed however 11 were rejected according to the Gaussian function criteria and 2 were rejected because of station high variability. In fact, different station high values are defined in EGVAP and Slovenian networks which leads to station rejection.

For example, Ljubljana GPS station “GSR1METO” is defined to be at 304,587 m high for EGVAP data but defined at 351,662 m high for Slovenian network data. The second rejected station “KARLMETO” is defined to be 134,788 m in EGVAP data however it's at 174,136 m in Slovenian GPS network data.

The following figures illustrate the white list stations geographic positions:



Figure 1: white list stations geographic positions

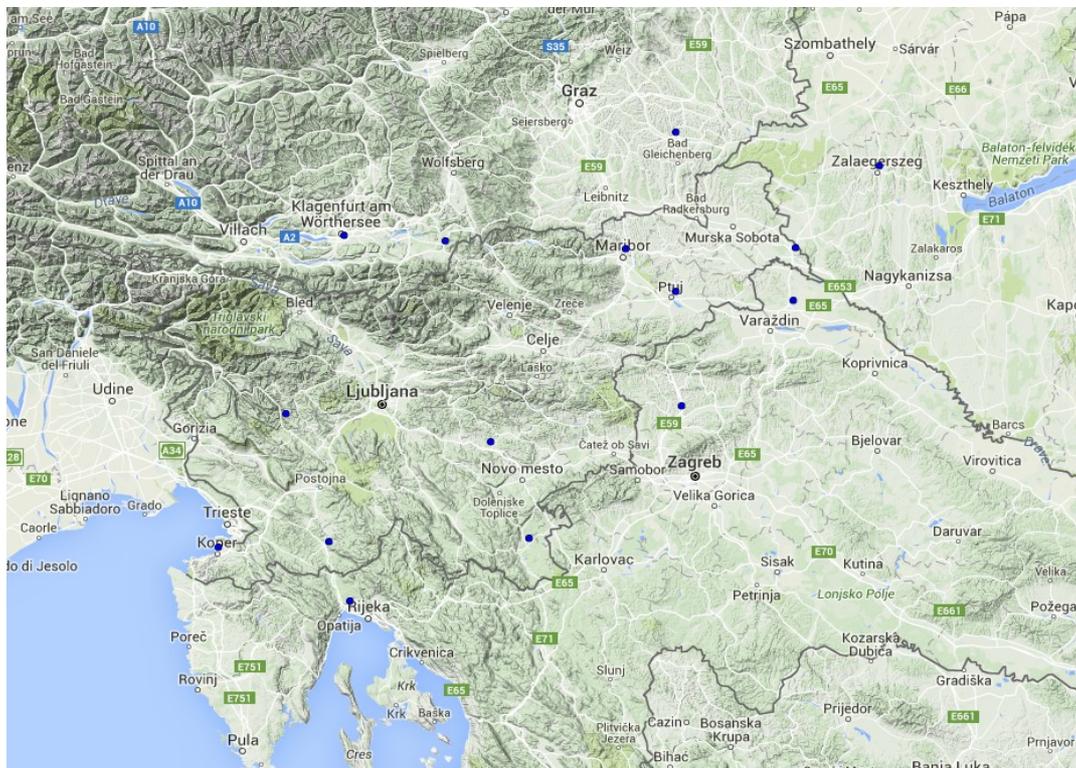


Figure 2: white list stations geographic positions – Slovenian network

6) Bias corrected obsouls:

The final step before running an assimilation experiment was to produce bias corrected obsoul files, therefore, the new white list was needed as input for the script “pregpssol.x” under “~/sbc_ztd/pregpssol”. This script transforms an input original obsoul file into a bias corrected one, which can be used later in assimilation before BATOR.

The following graph summarizes the different steps that have been precessed to produce bias corrected obsoul files starting from raw data files:

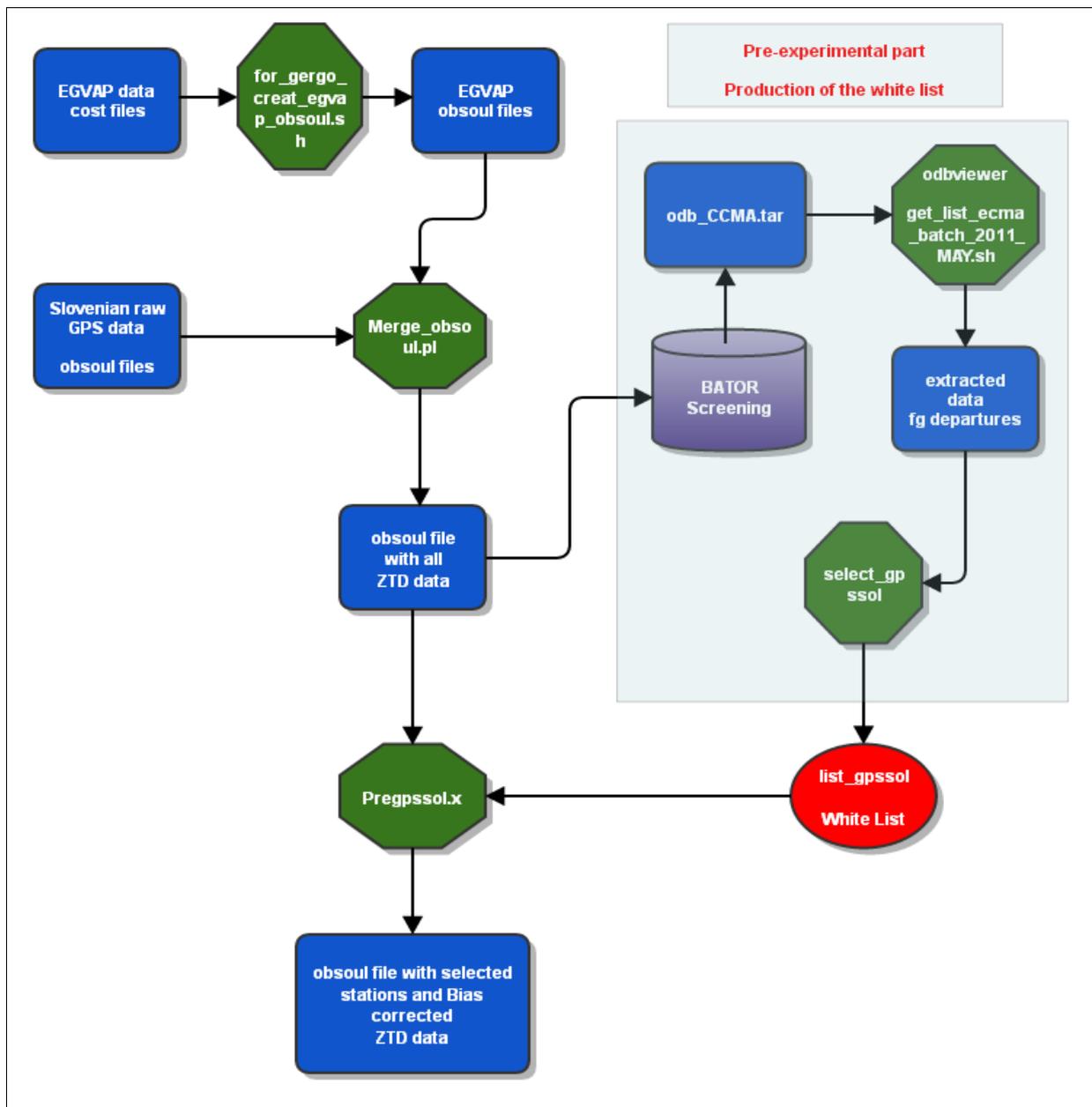


Figure 3: Processed steps to produce bias corrected obsouls starting from raw GPS data files

III. Experiments and settings

Four different experiments have been run namely reference experiment, EGVAP GPS data assimilation experiment, EGVAP and Slovenian GPS data assimilation experiment and variational bias correction experiment:

A. GPSREF: Reference experiment

The reference experiment was run over a twenty days period from September 1st to 20th 2015. This experiment includes the operational suite settings based on a standard 3D-Var assimilation system with a 3 hours cycling and two forecast runs per day for 24 hours range. A different binary was applied and it doesn't include the last year improvements in the operational suite.

No GPS ZTD data was considered during this experiment.

B. GPSOL: Assimilation of EGVAP data

This experiment assimilates EGVAP GPS data within a static bias correction method. It uses the same binary as the reference experiment. Few modifications have been introduced to “obsprep.ecf” in order to assimilate GPS data as follows:

At line 109:

```
cp ~/sbc_ztd/obsoul_bias/OBSOUL_EGVAP_bias_${YYYY}$MM$DD$RR.plusproch
OBSOUL.egvap
# merge all conv data into a single OBSOUL file
LISTF="obsoul_1_XXXXXX_hu_${YYYY}$MM$DD$RR
obsoul_5_XXXXXX_xx_${YYYY}$MM$DD$RR obsoul_1_viswe_si_${YYYY}$MM$DD${RR}00
obsoul_1_axxxxy_si_${YYYY}$MM$DD${RR}00 OBSOUL.egvap"
```

The white list was introduced as “list_gpssol” file in namelists folder.

C. SOLVEX: Assimilation of EGVAP and Slovenian data

This experiment assimilates EGVAP and Slovenian network GPS data within a static bias correction method using the same reference experiment binary.

The modification in obsprep.ecf script is as follows:

At line 110:

```
cp
~/sbc_ztd/obsoul_bias/OBSOUL_EGVAP_and_SLOAVG_bias_YYYY$MM$DD$RR.pluspr
och OBSOUL.egvap
# merge all conv data into a single OBSOUL file
LISTF="obsoul_1_XXXXXX_hu_YYYY$MM$DD$RR
obsoul_5_XXXXXX_xx_YYYY$MM$DD$RR obsoul_1_viswe_si_YYYY$MM$DD${RR}00
obsoul_1_axxxxxy_si_YYYY$MM$DD${RR}00 OBSOUL.egvap"
```

This experiment is also based on a static bias correction approach and uses the same white list as GPSOL was used during this experiment.

D. VARBEX: Variational bias correction experiment

The fourth and last experiment was based on a different approach namely the variational bias correction method. This experiment includes EGVAP and Slovenian network GPS data.

In order to run VARBEX, few supplementary settings were needed. In fact, we had to modify aldnml_minim and aldnml_screen namelists as follows:

```
&NAMVARBC_SFCOBS
LBC_PS=.TRUE.,
LBC_APD=.TRUE.,
LBC_SFCOBS110=.TRUE.,
```

Then, bias correction values had to be set to zero in the white list (list_gpssol_slov file). Finally, the “pregpssol.x” script had to be applied before

BATOR in order to select observations without modifying bias correction values.

Moreover, since variational bias correction for GPS ZTD data is supported since cy40, a back phasing has been made by Hungarian colleagues in order to enable this method for cy38. The “cy38t1_gpsvarbcmfcy40.01.INTEL111059.x.pack ” package includes the modified routines:

arpifs/obs_preproc/upecma.F90
arpifs/obs_preproc/defrun.F90
arpifs/op_obs/hopad.F90
arpifs/op_obs/gpszen_delay.F90
arpifs/op_obs/gpszen_delayad.F90
arpifs/op_obs/gpszen_delaytl.F90
arpifs/op_obs/hoptl.F90
arpifs/op_obs/hop.F90
arpifs/module/varbc_pred.F90
arpifs/module/varbc_table.F90
arpifs/module/varbc_setup.F90
arpifs/module/yomobs.F90

A new binary including these modified routines has been produced and considered during the current experiment.

IV. Scores and comparison

We compared the different experiments performance according to the reference run. Standard deviation and bias were computed for Surface pressure, Temperature, wind, relative humidity, cloud cover and precipitations parameters:

1) Surface Pressure:

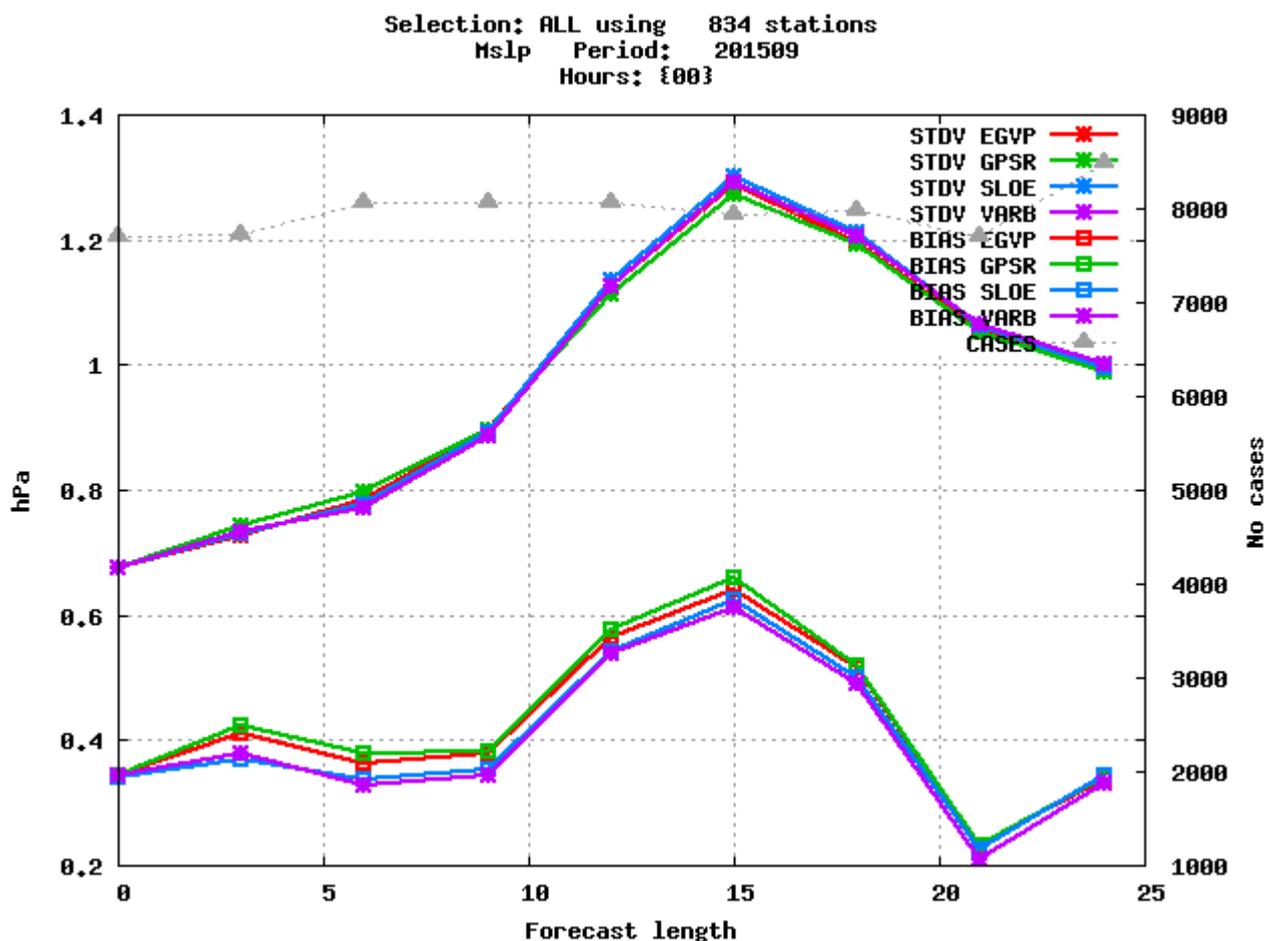


Figure 4: Verification against reference (green) for surface pressure over the operational domain for all network times

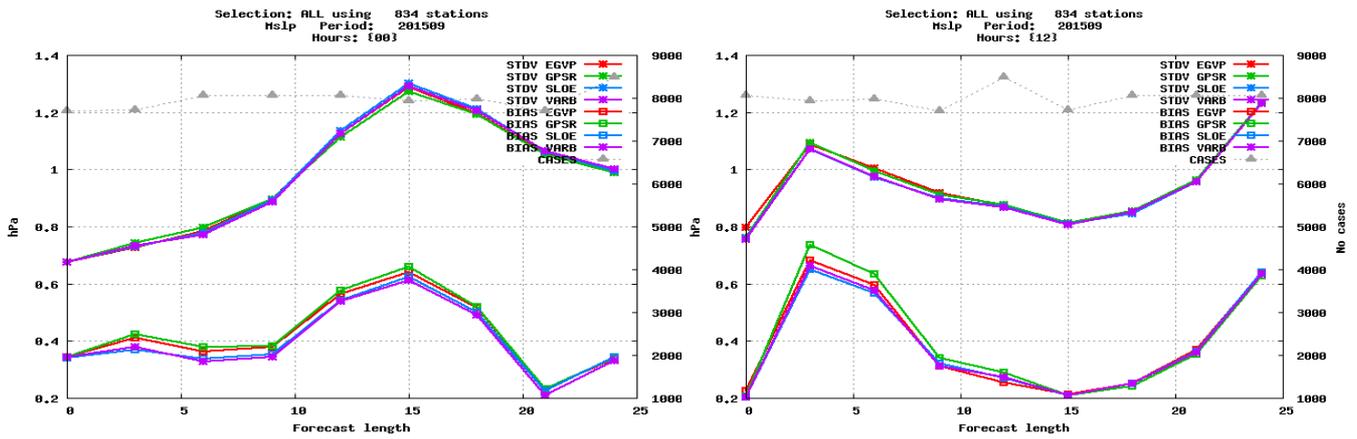


Figure 5: Verification against reference (green) for surface pressure over the operational domain for 00H (left) and 12H (right) forecasts

2) Relative Humidity:

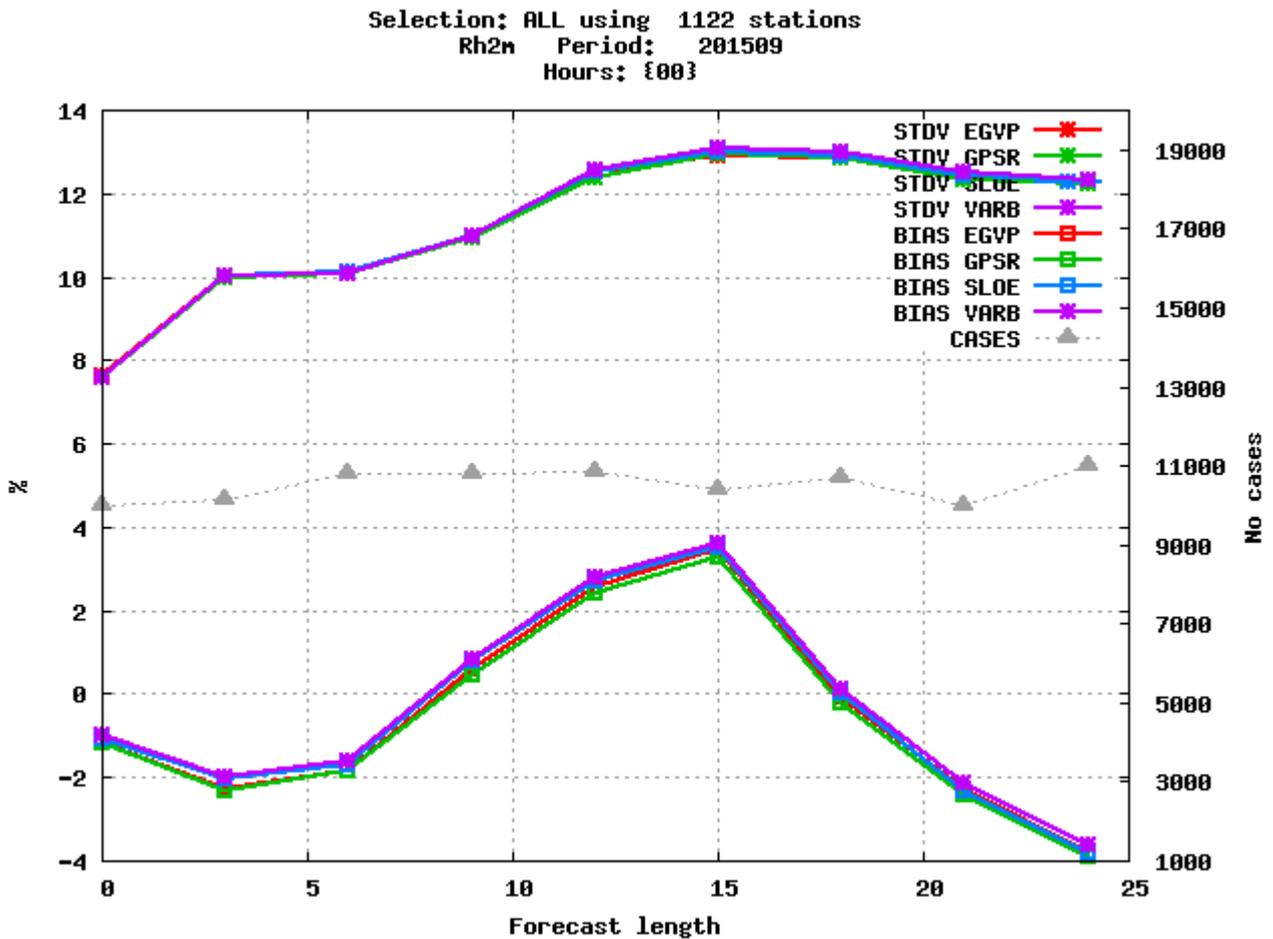


Figure 6: Verification against reference (green) for Relative Humidity over the operational domain for all network times

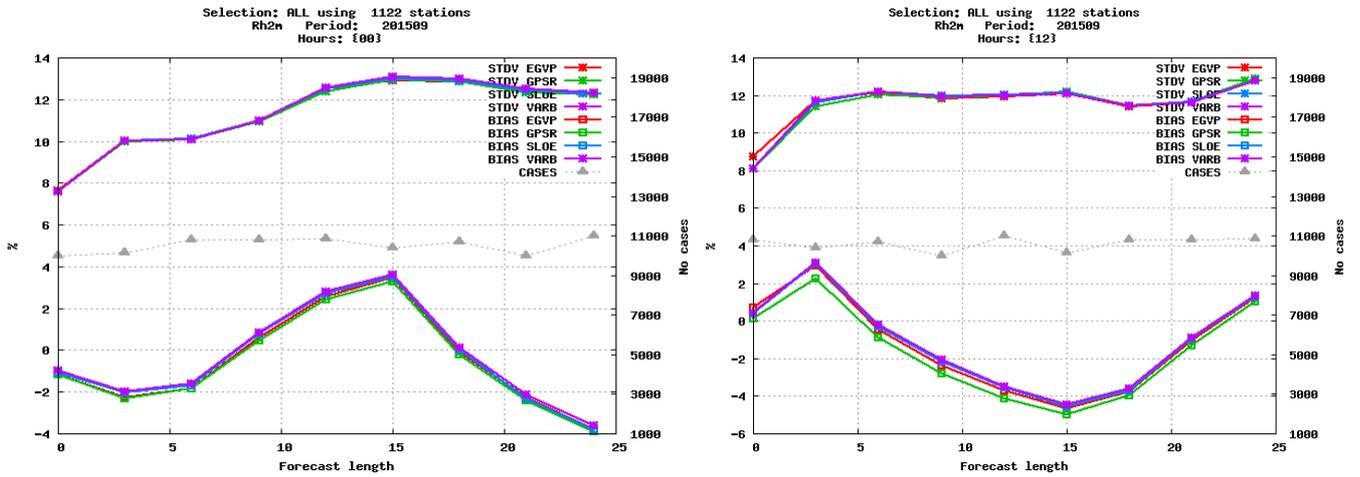


Figure 7: Verification against reference (green) for Relative Humidity over the operational domain for 00H (left) and 12H (right) forecasts

The verification against reference for surface pressure and relative humidity parameters shows a marked decrease of bias.

The assimilation of EGVAP data improved the scores but larger improvements are noticed while adding Slovenian data and especially while using the variational bias correction method.

3) Surface Temperature:

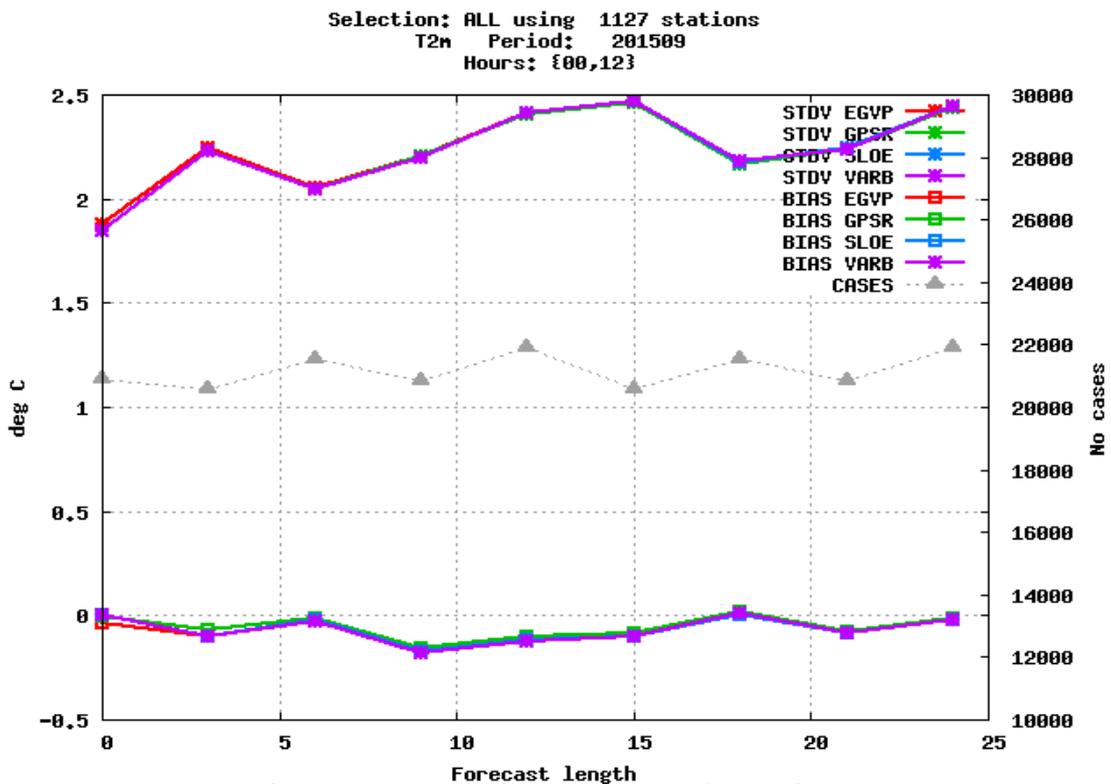


Figure 8: Verification against reference (green) for Surface Temperature

4) Cloud Cover:

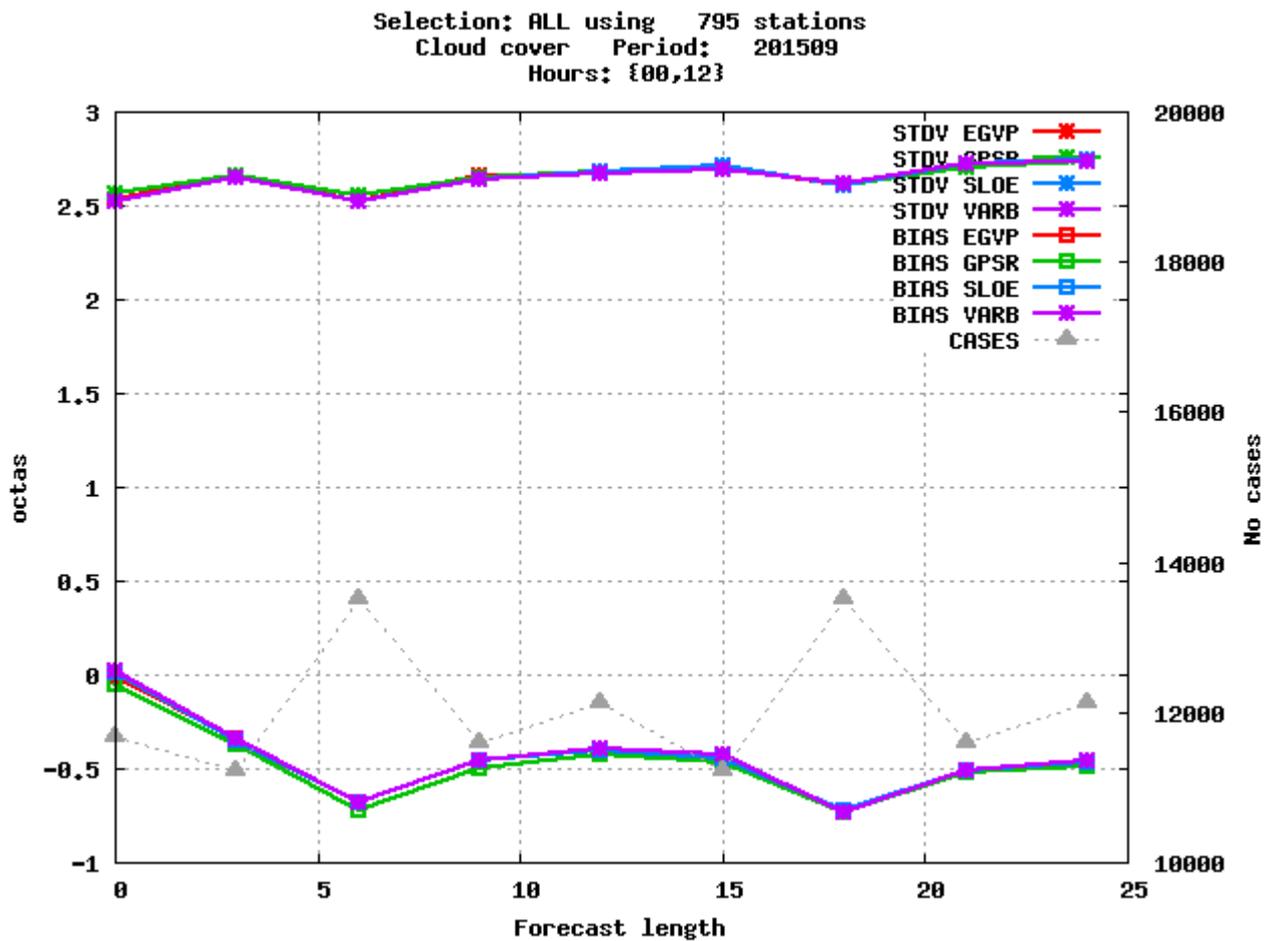


Figure 9: Verification against reference (green) for Cloud Cover over the operational domain for all network times

A slight decrease in bias is noticed for the Cloud Cover parameter over the operational domain for the ensemble of networks. The different experiments are presenting a very near performance.

5) Cumulative rain:

We considered the case of September 13th when important rain quantities have been observed over northern part of Slovenia as GPS data has more impact during medium and high raining situations. We noticed a slight impact on rain localization but no significant result can be concluded as illustrated in the following figures:

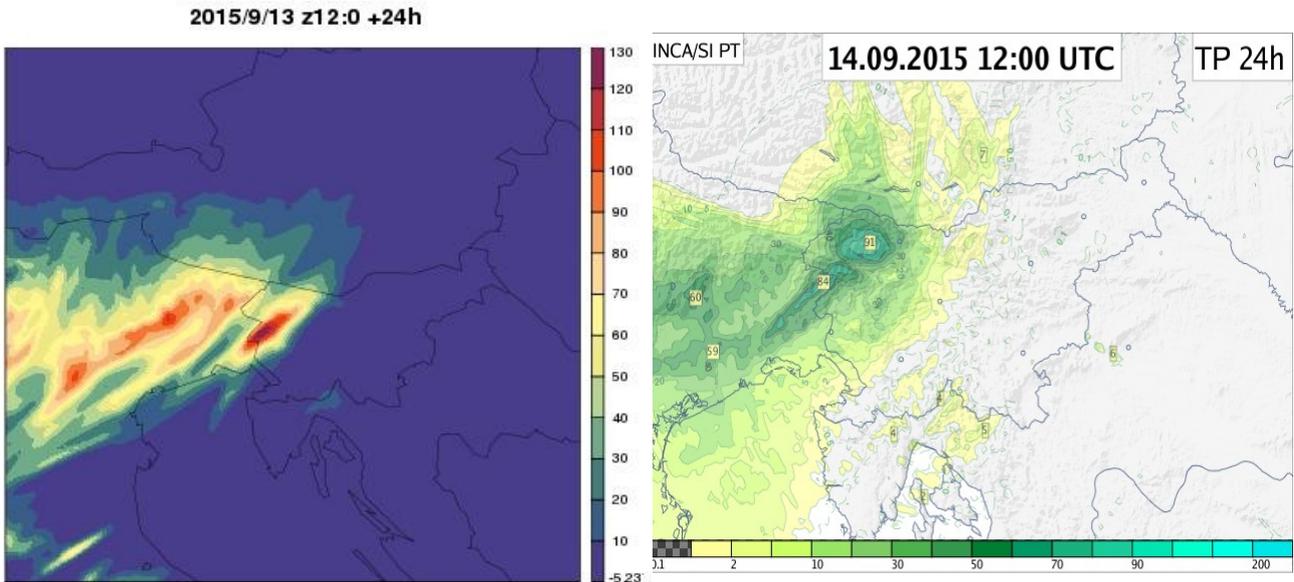


Figure 10: 24H cumulative rain forecast for September 13th 2015 (12H network)
Reference with no GPS observations (left) - Observation (right)

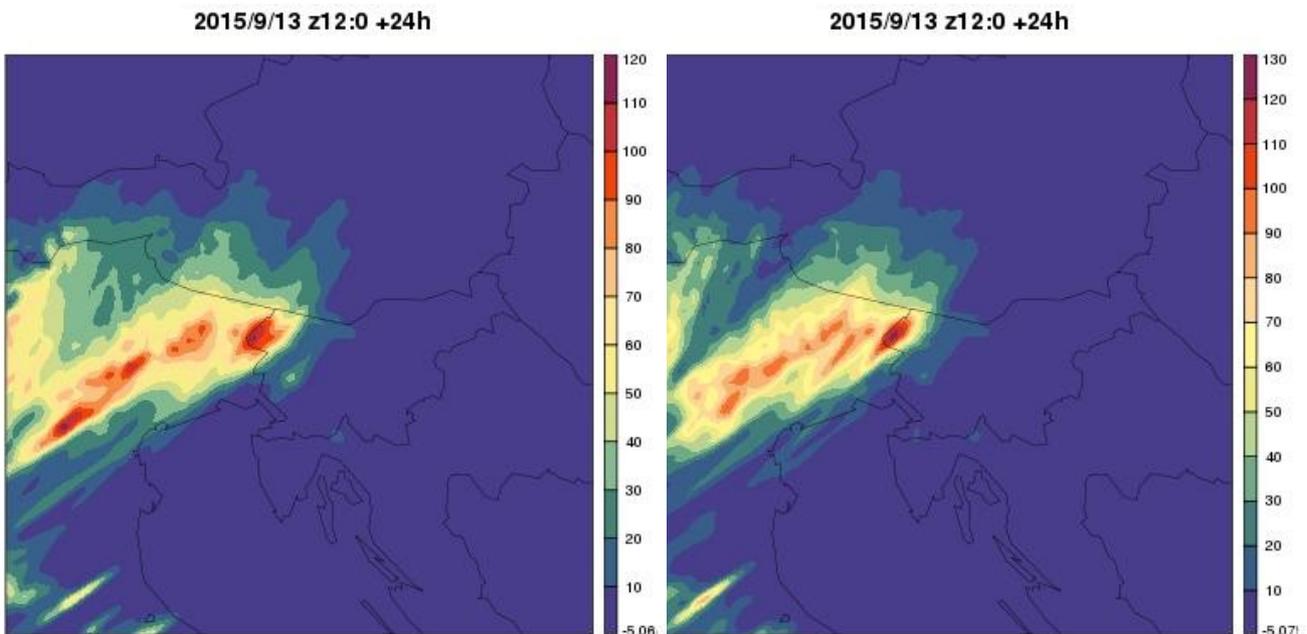


Figure 11: 24H cumulative rain forecast for September 13th 2015 (12H network) including GPS data on top of all observations used in reference experiment
Static bias correction method (left) - Variational bias correction method (right)

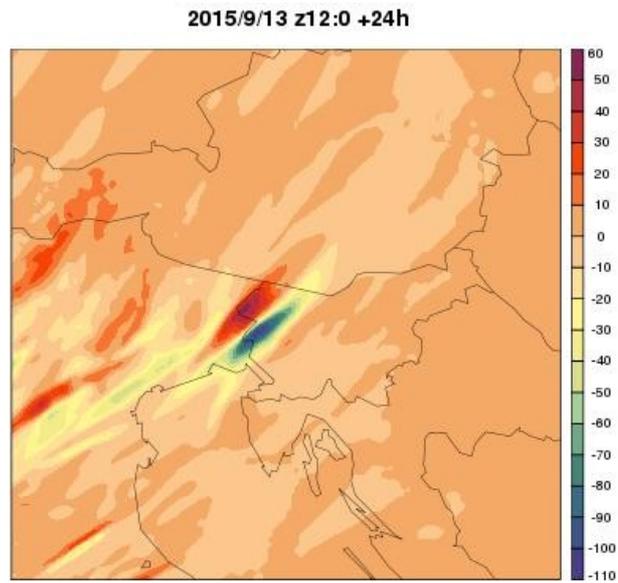


Figure 12: Difference in 24H cumulative rain forecast for September 13th 2015 (12H network)
Variational bias correction GPS assimilation - Reference

6) Wind Speed:

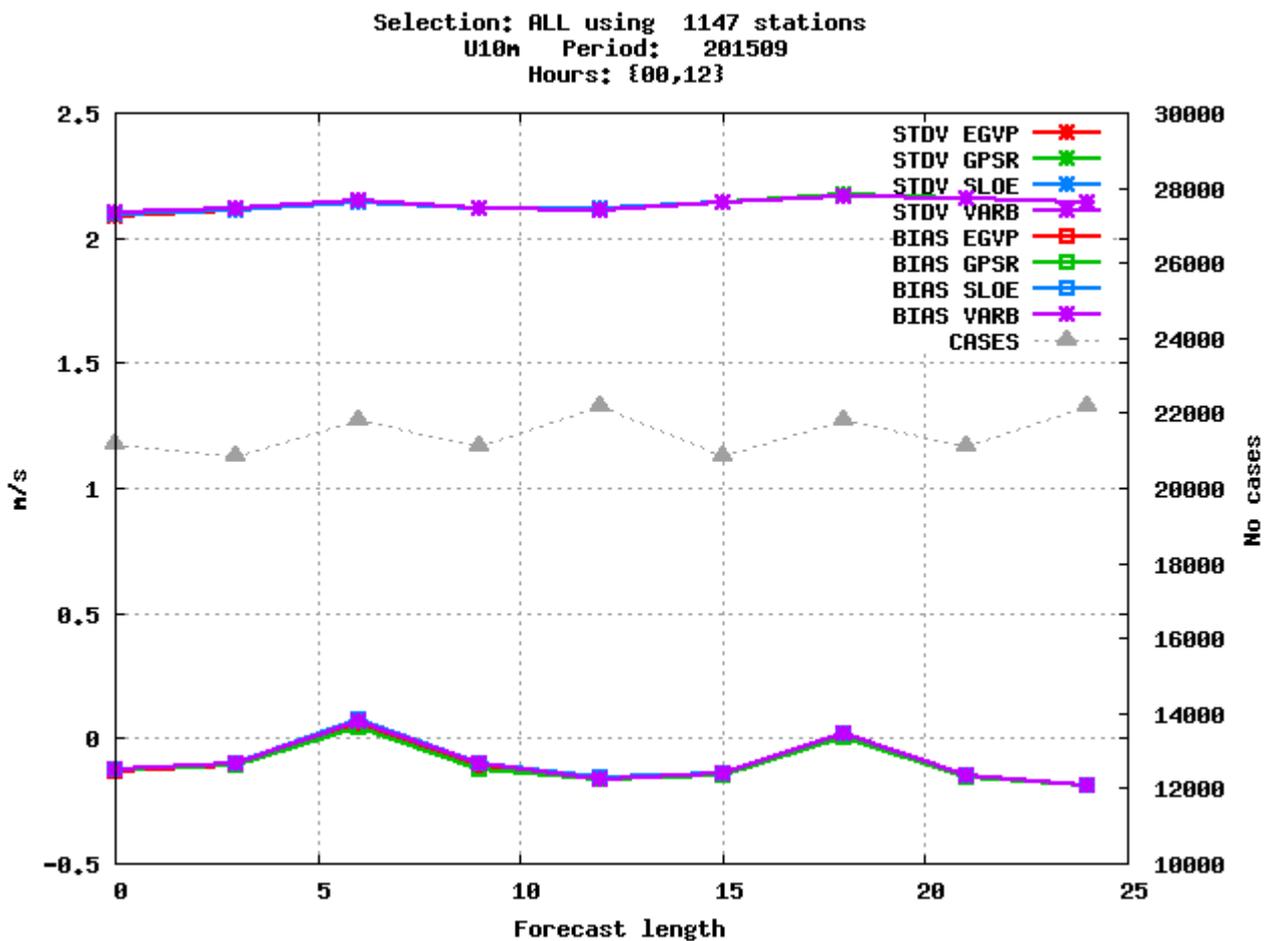


Figure 13: Verification against reference (green) for Wind Speed over the operational domain for all network times

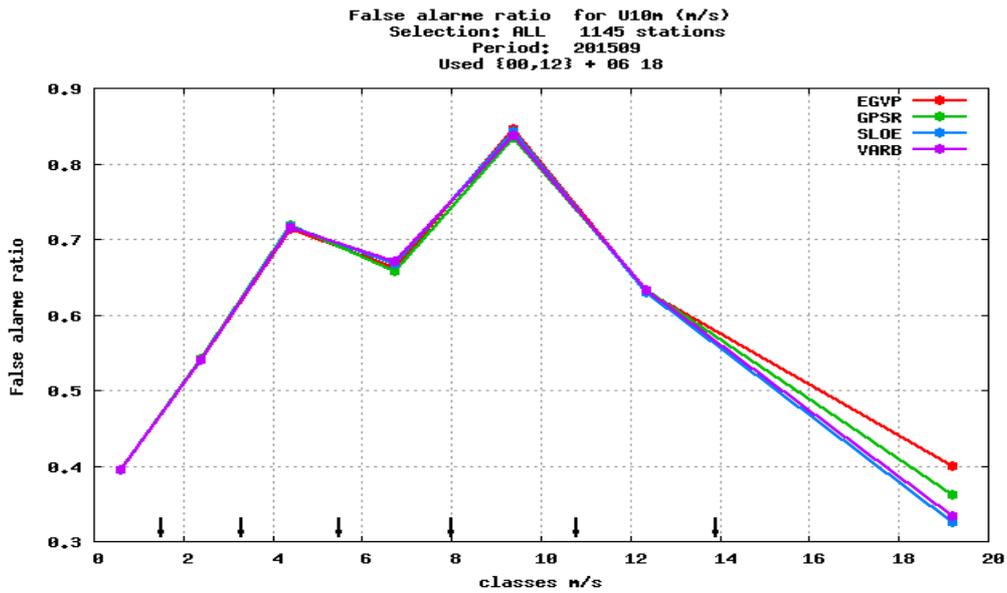


Figure 14: False Alarm Ratio for Wind Speed over the operational domain

7) Wind Direction:

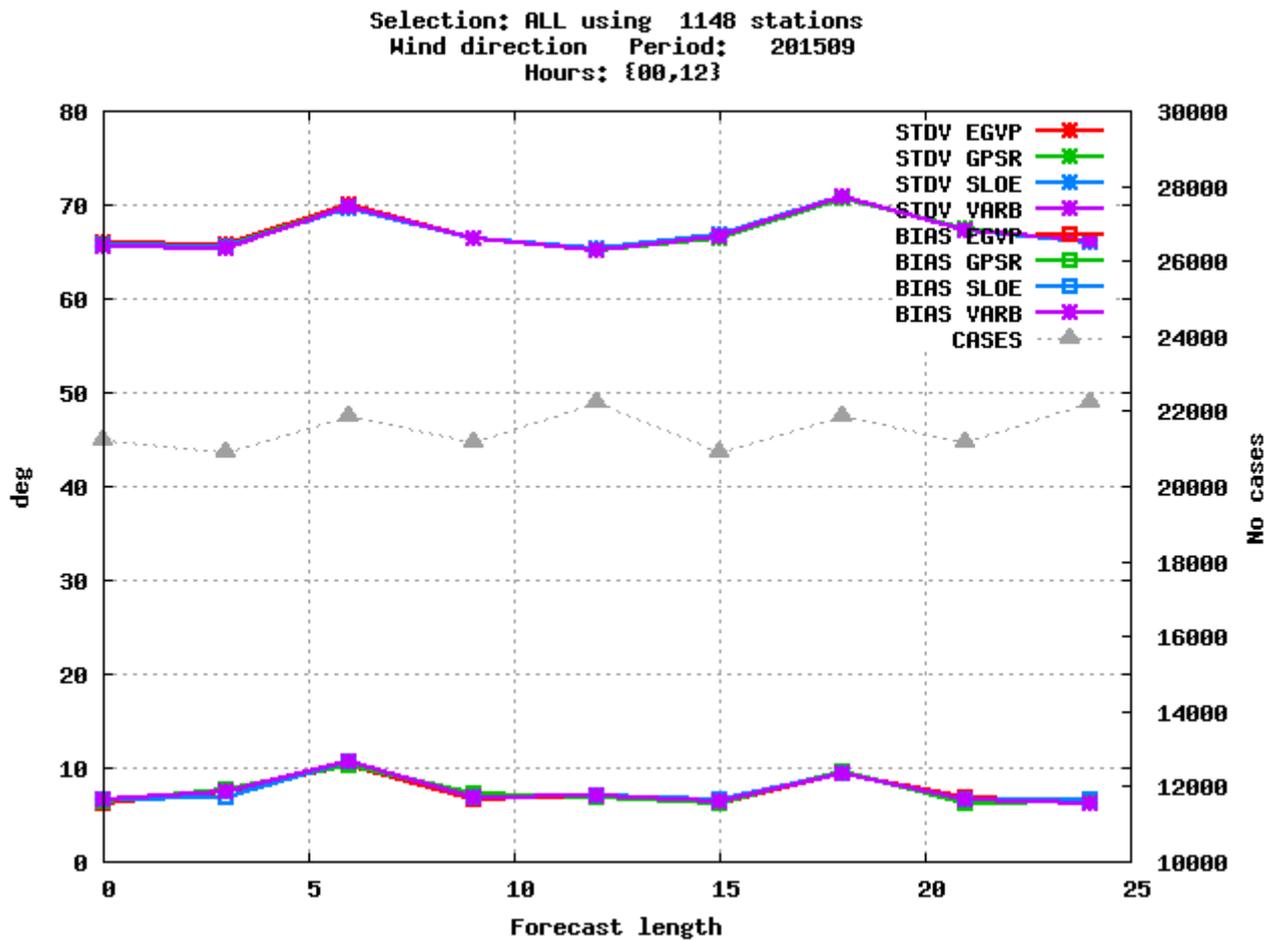


Figure 15: Verification against reference (green) for Wind Direction over the operational domain for all network times

Conclusions

It was a good opportunity to work during this stay on a very interesting topic such as GPS ZTD data assimilation regarding the increasing number of available data and its benefits in water vapour observation.

In order to use trustful data, we began by creating a white list of GPS stations which respect a set of selection criteria. We applied the selection process on the ensemble of available data namely EGVAP and Slovenian network stations. The selection process was based on July 2015 observations, however, in order to get more objective results, we decided to apply experiments on a different period, September 2015.

Four experiments were run during this stay; the reference experiment included operational Slovenian suite parameters which does not include GPS data assimilation. A second experiment has been set in order to assimilate only EGVAP data. Then, So that we study the Slovenian network data impact, a third experiment has been set including all the available observations. Finally, counter to previous experiments which used a static bias correction approach, a fourth experiment based on variational bias correction method has been run including all the available data.

As a first finding, the GPS data assimilation had beneficial impact especially on surface pressure, relative humidity and also cumulative rain and cloud cover, a noticeable bias reduction has been observed for these parameters. The Slovenian data has been shown also beneficial. The variational bias correction method had positive impact as well compared to reference experiment and static bias correction method. On the other hand, no significant impact has been noticed for surface temperature and wind parameters, but further investigation is required for utter conclusions.

References

- H. Vedel, K. S. Mogensen, and X.-Y. Huang 2001 Calculation of zenith delays from meteorological data, comparison of NWP model, radiosonde and GPS delays.
- EUREF Symposium 2005 EGVAP Meteorology and geodesy synergy
- CNRM/GMAP/OBS 2007 Note sur l'assimilation des données GPS sol
- Patrick Moll , Paul Poli, Véronique Ducrocq 2008 Use of ground based GNSS data in NWP at Météo-France
- Stefan Schneider, Florian Meier, Xin Yan 2010 Data assimilation activities at ZAMG (Austria)