*Regional Cooperation for Limited Area Modelling in Central Europe*

# RC LACE Stay Report

Topic: Inline data assimilation of ASCAT SSM using SODA/SEKF

OMSZ Budapest 5th June - 30th June 2023

Matjaž Ličar

Supervisor: Helga Tóth

Contents

## 1 Introduction

In the previous work [1] we performed the assimilation of satellite based SSM observations with SURFEX in offline mode, by performing offline data assimilation using SODA with a SEKF algorithm. The satellite product used was available on a daily basis and covered the entire Europe with some surrounding areas. This work proved the technical feasibility of assimilating this type of observations, however future experiments, particularly in an inline setting, were desired. Since then, the Hungarian meteorological service OMSZ has successfully implemented the use of SODA/SEKF algorithm for the assimilation of screen level temperature and relative humidity in their operational AROME model. Given this new expertise and a new pre-operational satellite product, which is available in near real time, it is now possible to perform such an inline experiment.

## 2 Satellite surface soil moisture product

H08–SM-OBS-2 (H08) is a satellite product provided by EUMETSAT, the European Organization for the Exploitation of Meteorological Satellites [2]. It is based on the Advanced Scatterometer (ASCAT) instrument onboard the Metop satellites. The product aims to provide near-real-time (NRT) information on surface soil moisture (SSM) and surface wind vectors over the Earth's land and ocean areas. It utilizes the back-scattered microwave radiation measurements from ASCAT to derive these parameters. The basis for this product is post-processing of SM-OBS-1 product, which is performed at GeoSphere (formerly ZAMG).

The ASCAT sensor has a sampling resolution of 12.5 km. The soil moisture product of the SM-OBS-1 product results from processing these data at a 25 km (research) and 50 km (operational) resolution. Soil moisture data of the SM-OBS-2 product at a resolution of 1km is obtained by disaggregation and downscaling using information from a local scaling layer of the base SM-OBS-1 product. The data is available in the form of netCDF files, containing surface soil moisture [%], surface soil moisture standard error [%] and soil moisture processing and correction flags. It is currently listed as in a pre-operational state.

**2.1 Pre-processing of satellite data.** Due to the near-real-time nature of the product, potentially multiple netCDF files are available for each date with data valid at "random" times during the day with a varying spatial extent. For our purposes we need to select only the data within the Hungarian AROME domain, at suitable times for our assimilation system.

An in-house developed utility (HAWKhoz, written in C++) is used to select data points near and within the AROME domain for a given satellite netCDF file. The output is a netCDF file on a pre-defined regular 1 km grid. In the next step we use the CDO utility to transform these data to the 2.5 km AROME domain. Additional pre-processing and quality control is then performed to obtain the data in ASCII format ready for data assimilation . Figure (1) demonstrates this process.

Data from a single pass of the satellite over central Europe can be split into multiple satellite files with a varying amount of data within the AROME domain. We thus combine several data files into a composite image, with data valid within a window around a target time. A time window length of 3 hours ($\pm$ 90 minutes around a target time) was chosen as a reasonable value for our experiment. Figure (2) demonstrates combining multiple files into a composite.

**2.2 Calibration of satelite data using CDF matching and quality control.** The satellite data is given as a percentage accounting for the soil field capacity, whereas the model soil moisture variables are in absolute units of $[m^3/m^3]$. It is thus necessary to transform the observations to model quantities. The calibration process using CDF matching is analogous to the previous work from the stay at OMSZ in 2022 [1]. A 18 month data set from 2022-01-01 to 2023-06-01 of satelite data and 3h analyses from the operational AROME model were used for this purpose. First, the point wise minimum and maximum values from the model analyses are used to linearly transform the observations to the model equivalent units. Assuming both observation and model data sets follow a normal distribution, equating the CDF functions of both data sets yields [1],[3]:

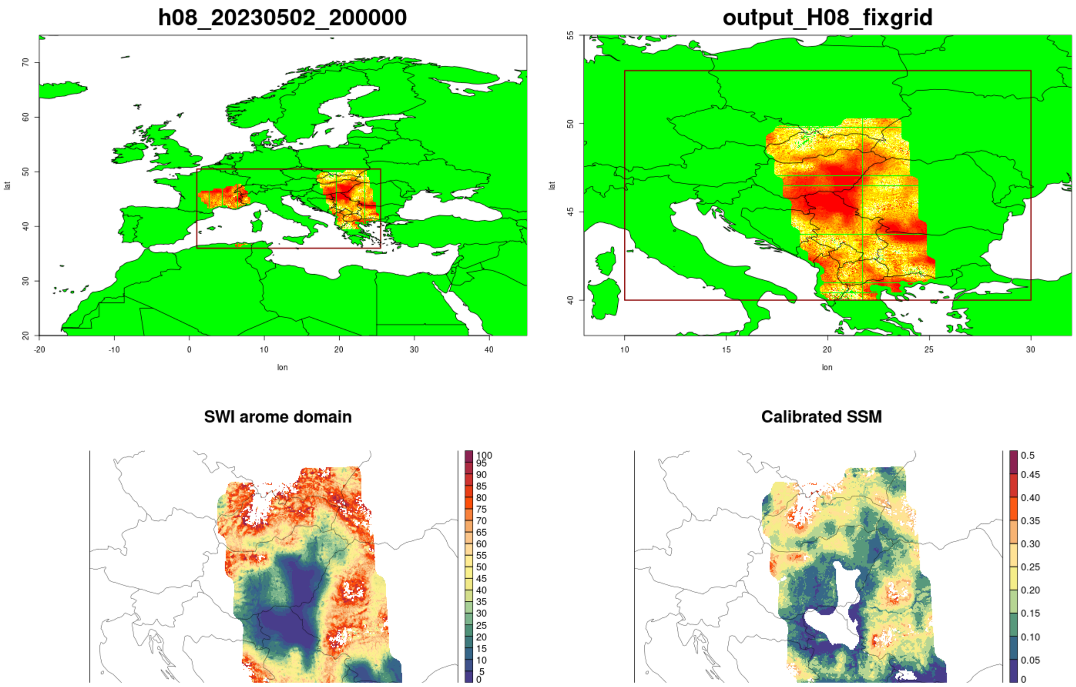



FIGURE 1. Demonstration of pre-processing of satellite data. Top left: Satellite product containing valid data within the AROME domain. Top right: Output from the HAWKhoz utility which selects only data within the red box. Bottom left: Transformation using CDO to the AROME 2.5km grid. Bottom right: Pre-processed satellite data.

(1)
$$
\begin{aligned}
x_m &= p_2 x_s + p_1 \\
p_2 &= \frac{\sigma_m}{\sigma_s} \\
p_1 &= \mu_m - p_2 \mu_s
\end{aligned}
$$

where $x_m$ and $x_s$ denote the modeled and observed SSM, $\mu_m$, $\mu_s$, $\sigma_m$ and $\sigma_s$ are the mean values and standard deviations of the modeled and observation data sets, respectively.

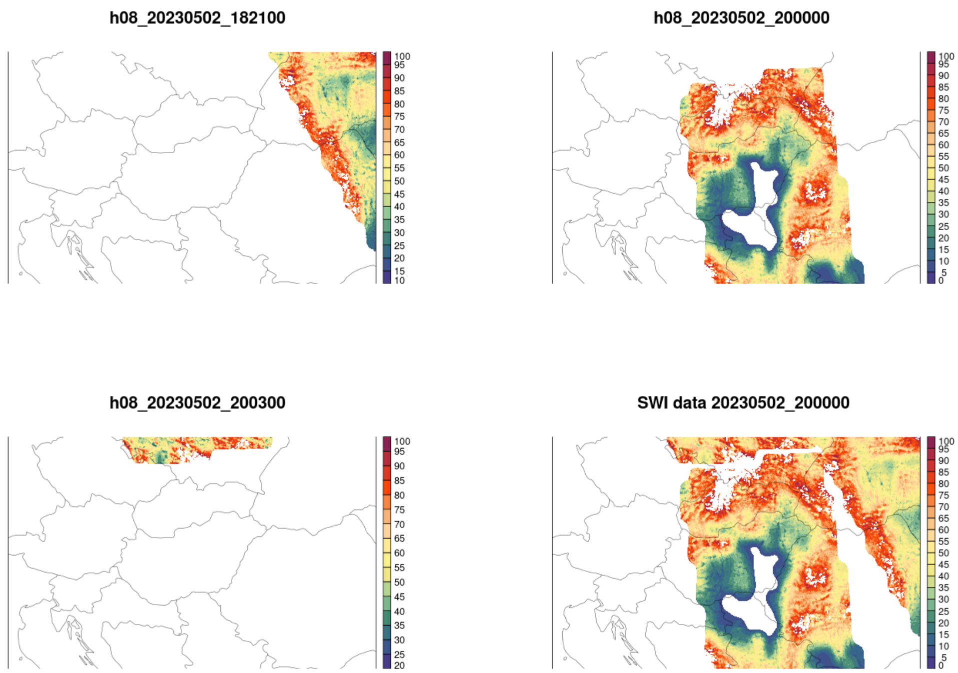



FIGURE 2. Example of combining multiple satellite images into a composite within the time window of 3 hours around 20:00. Top and bottom left: 3 satellite images within the time window. Bottom right: Composite of the images.

By inspecting certain satellite images, we see that some degree of data filtering is desirable, as some images may contain features that do not appear physical. Apart from the standard error, soil moisture processing and correction flags are also available with the product. Currently, in the pre-operational state of the product, only cases with measurements below 0% or above 100% are implemented in these flags. Other possible values are listed in the product manual, presumably to be implemented at a later stage. We therefore apply only rudimentary filtering on the basis of relative error, where a very liberal cutoff value of 1 was used. Some erroneous features mentioned above can display relative errors exceeding 10. The reasoning behind choosing such a high value is, that the standard error has little dependence on the magnitude of the measurement. For very dry points this leads to very high relative errors, to the point where choosing a lower value would lead to filtering dryer areas completely, which could in turn lead to unwanted biases in our assimilation system. Filtering of a very dry area can be seen on figure (1).

**2.3 Availability of satellite data.** In this section we do a short analysis of the temporal and spatial availability of suitable observation data. Figure (3) presents a histogram of satellite images corresponding to respective 3 hour windows during the day, where the data represents an archive of satellite data spanning dates between 2022-01-01 and 2023-06-01. We see, that the satellite passes the AROME domain two times per day, with the most natural choice to perform satellite data assimilation for 09h and 21h.

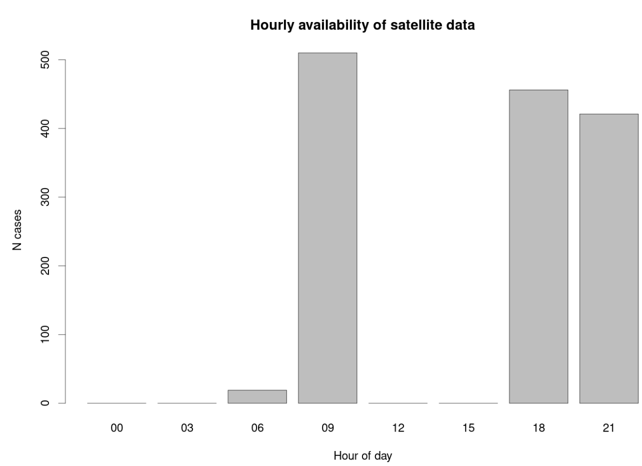



FIGURE 3. Histogram of times of validity for satellite images for an 18 month period.

Figure (4) presents the spatial availability of satellite observations. We see, that on average a satellite image covers about 30% to 40% of the domain, and that domain coverage is relatively even across Hungary, with some mountain regions with fewer data available.

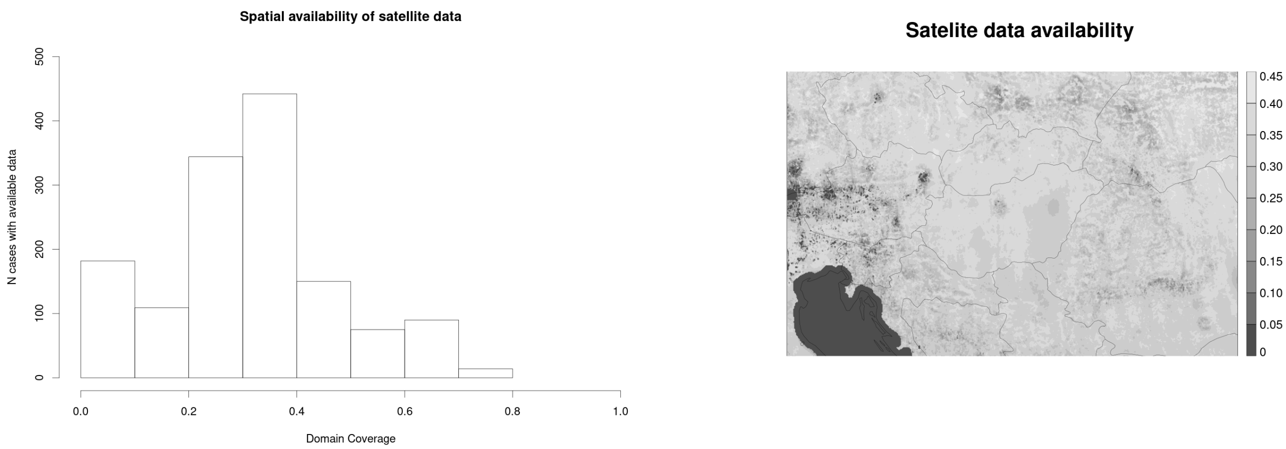



FIGURE 4. Domain coverage of satellite observations. Left: Histogram of domain coverage for images with valid data inside AROME domain. Right: Point wise availability of satellite data, i.e. number of observations at point over total number of images.

**2.4 In-situ observations of soil moisture.** The Operational Drought and Water Scarcity Management System service from Hungary [4] provides in-situ observations of soil moisture. These data are available for 6 layers spanning depths between 10 cm and 75 cm at 116 locations spread uniformly across Hungary at an hourly resolution. Use of these data had been considered, since in-situ observations would provide valuable means of validating both the satellite product, as well as our assimilation system. Unfortunately, none of the measured depths correspond neither to the top most layer of the soil observed by the satellite, nor to any layers represented in the ISBA scheme. The locations of stations themselves however provide us with a convenient set of locations, for cases where we want to examine the model or satellite data in a temporal sense. The locations of these stations are shown on figure (5).

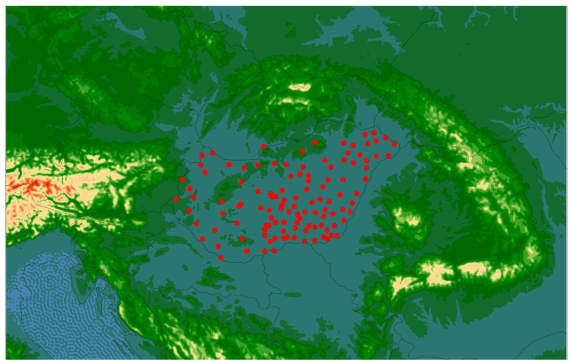



FIGURE 5. Locations of the 116 stations of the Hungarian Operational Drought and Water Scarcity Management System.

**2.5 Verification of satellite data calibration.** In this section, we compare calibrated and uncalibrated satellite data with operational AROME model WG1 3 hourly analyses spanning a period between 2022-01-01 and 2023-06-01. The locations outlined in section (2.4) were used for this purpose. We compute the bias and correlation coefficient between observations and model analyses for each of the locations. Figure (6) shows, that the bias of calibrated data compared to uncalibrated data is considerably reduced, whereas the correlation coefficient remains largely the same. This is consistent with expectations, given the linear nature of the calibration. Figure (7) shows the comparison between model analyses and satellite observations for the best and worst performing stations after calibration with respect to bias.

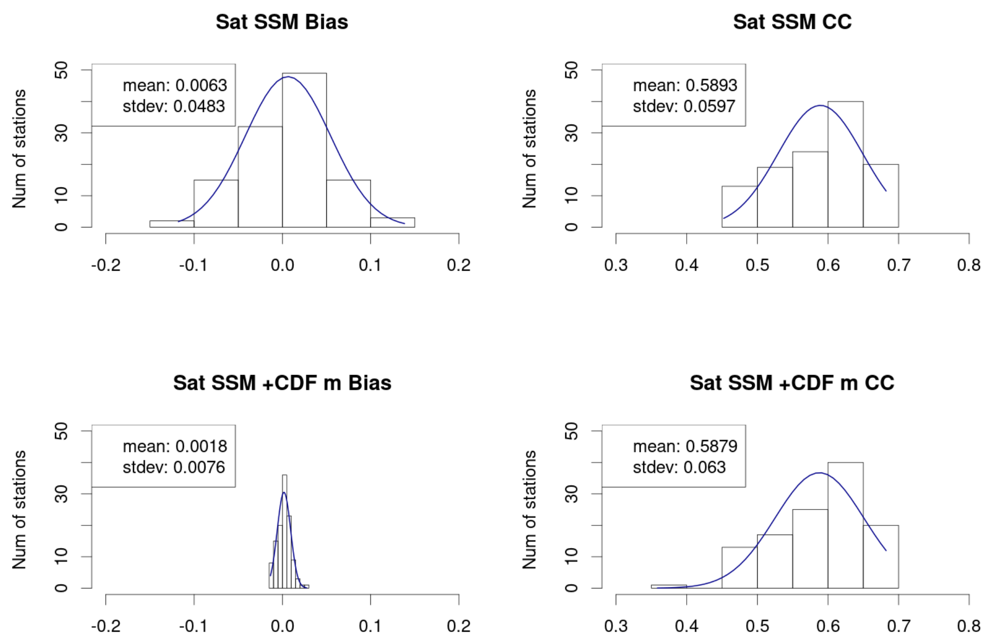



FIGURE 6. Bias (left) and correlation coefficient (right) distribution for 116 locations between observations and non-calibrated (top) and calibrated (bottom) data.

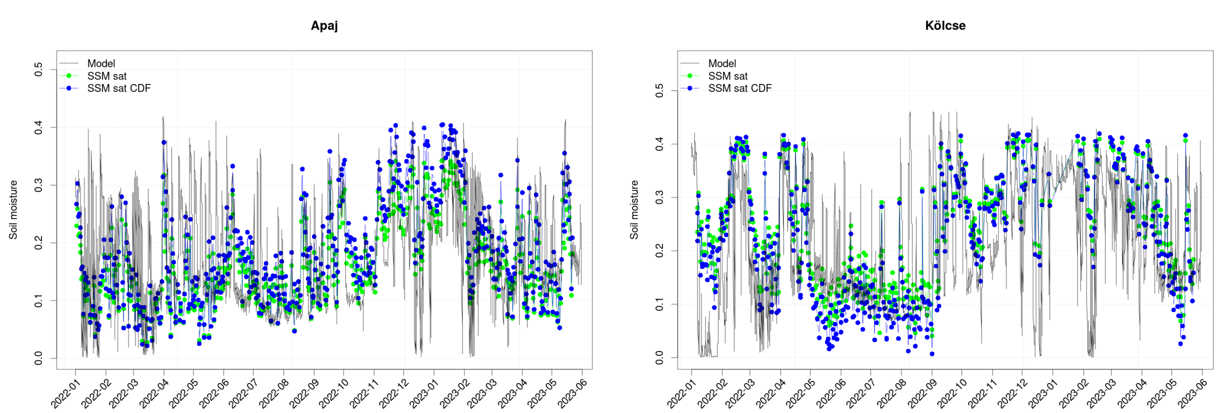



FIGURE 7. Comparison between model analyses, calibrated and non-calibrated satellite observations for two locations.

## 3 INLINE DATA ASSIMILATION EXPERIMENTS

Our experiments were run using the EcFlow system. The suite definitions and other settings were analogous to the operational AROME model, with a 3 hour assimilation cycle and two 24 hour forecast runs per day. The only difference between the experiments was the method of surface data assimilation. The assimilation cycle consists of the following tasks:

- Preparation of boundary conditions from the global ECMWF model
- Preparation of observations
- Surface data assimilation (different between experiments)
- 3Dvar minimisation for atmospheric data assimilation. Analysis used as IC for forecast run.
- 3h hour integration. The result of integration is used as first guess for analysis in the next cycle.
- 3h Offline SURFEX run: preparation of forcing files based on the forecast run, reference run and 8 perturbed runs. The perturbed runs are used to estimate the Jacobians in the SEKF algorithm, where we have 2 perturbed runs (with a positive and negative perturbation) for each of the control variables.


FIGURE 8. EcFlow suite for the experiments.

The experiments were run for a period of 31 days from 2023-05-01 to 2023-05-31, with a 7 day spinup period. Details regarding data assimilation in SURFEX can be found in [1],[5] and [6].

- Satelite SSM observation experiment
  Surface data assimilation is performed using SODA with SEKF algorithm. The assimilation is only performed at 09 and 21 hours. In other cases, the first guess is used as the analysis. Observations are provided in the form of ASCII files. Default values were used in the SURFEX namelist, with four control variables WG1, WG2, TG1 and TG2.
- Operational reference run This experiment is analogous to the operational AROME model used at OMSZ. Surface data assimilation is performed using SODA/SEKF at every 3 hours, where the analysis of T2m and RH2m produced by CANARI is used to update the SRUFEX fields. The same four control variables as in the SSM observation experiment are used.
- Reference run without surface data assimilation No surface assimilation was performed in this experiment. CANARI analysis of T2m and RH2m was performed, however SURFEX fields were not updated.

## 4 Results

**4.1 Validation with OVISYS system.** The Hungarian OVISYS model validation system provides convenient means to produce forecast scores with respect to synop observations. A post processing script produces forecast data in suitable format for the OVISYS system. We can then use a web interface to plot scores on the basis of a selected subset of synop stations, compare the results with existing runs in the OVISYS system, and similar. We choose the entire set of 557 synop stations in Hungary and surrounding areas and compare the scores for our 3 experiments. Bias and RMSE for the 2m temperature, relative humidity and 3h precipitation as a function of lead time are shown on figures (9), (10) and (11). Similarly, figures (12) and (13) show T2m and RH2m analysis bias at 00 h and 12 h. A quick summary of these figures is, that satellite data assimilation has a positive effect on the scores opposed to not performing surface assimilation at all, however the effect is small and the satellite assimilation experiment does not outperform the operational system.

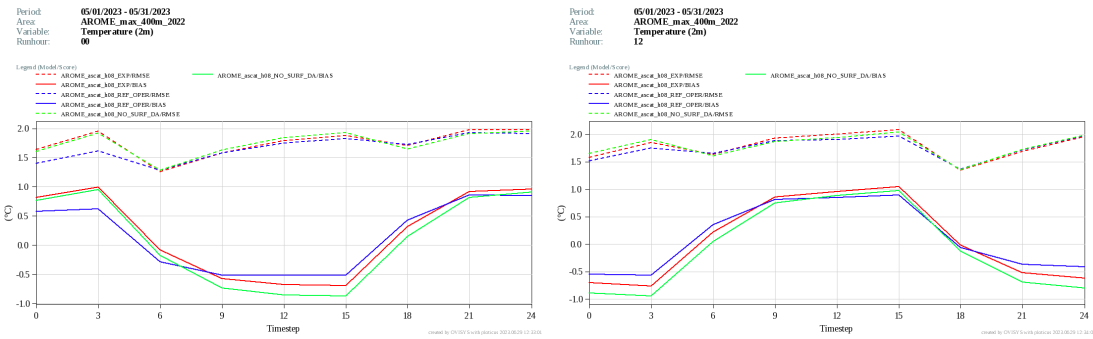



FIGURE 9. Bias and RMSE of T2m with respect to synop observations for 00 forecast runs (left) and 12 forecast runs (right) for each experiment.

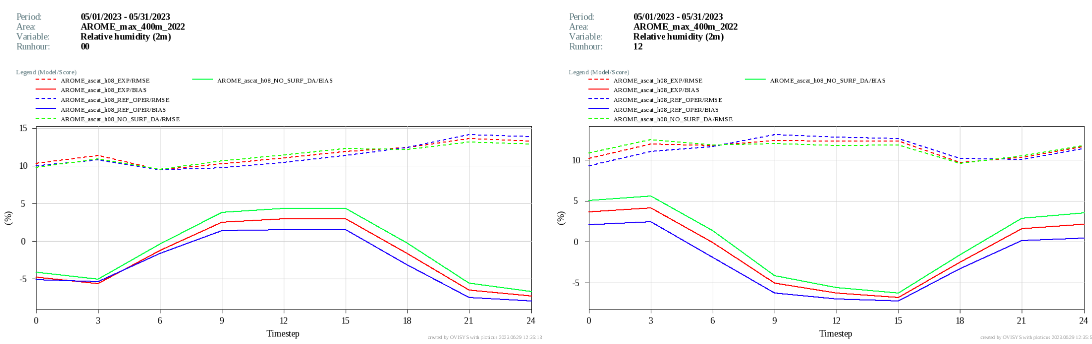



FIGURE 10. Bias and RMSE of RHm with respect to synop observations for 00 forecast runs (left) and 12 forecast runs (right) for each experiment.

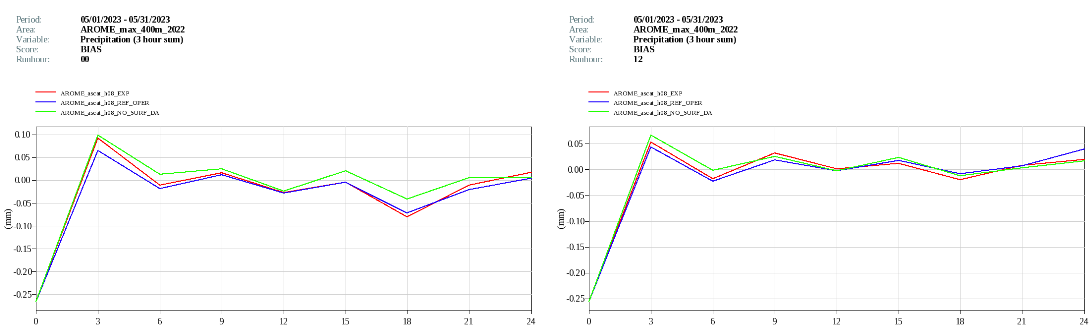



FIGURE 11. 3h precipitation Bias with respect to synop observations for 00 forecast runs (left) and 12 forecast runs (right) for each experiment.

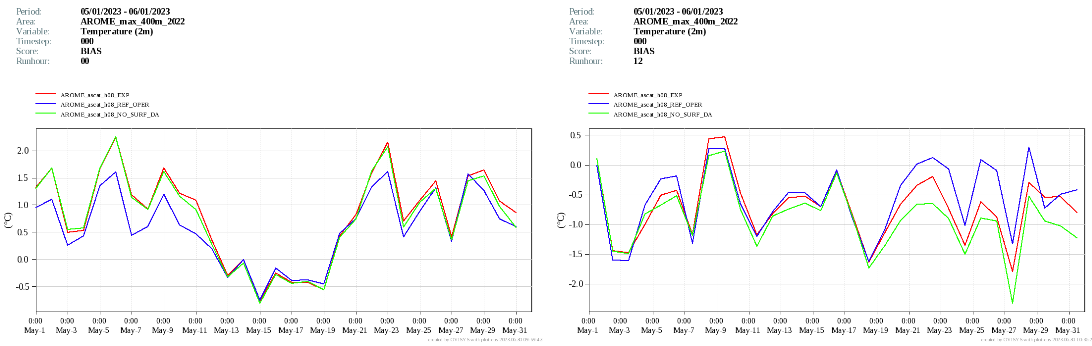



FIGURE 12. Bias of the T2m analysis at 00 h (left) and 12 h (right) for each experiment.

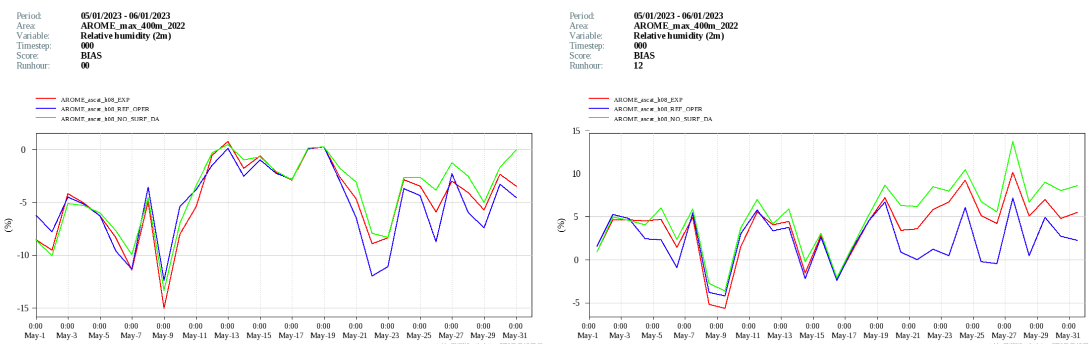



FIGURE 13. Bias of the RH2m analysis at 00 h (left) and 12 h (right) for each experiment.

**4.2 Examination of SURFEX field analyses.** In this section we compare the SURFEX field analyses for each experiment. For this purpose we use the locations outlined in section 2.4. For each of the experiments and each station we compute the bias, RMSE and correlation coefficient between WG1 and satellite observations. These are shown on figures (14), (15) and (16). We see, that by this metrics the WG1 more closely resembles observations in the satellite data assimilation experiment, however the effect is very small. Figure (17) shows observations against modeled WG1 for each experiment, however any systematic biases or major differences between the experiments are not apparent from such plots.

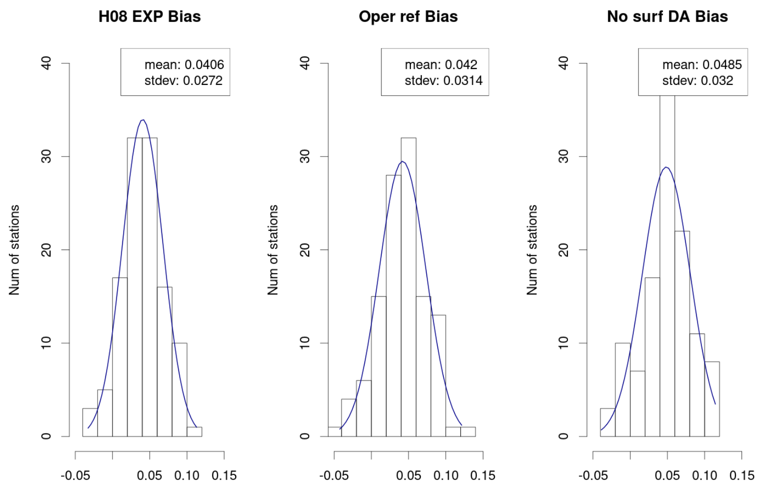



FIGURE 14. Bias of WG1 analyses with respect to satellite observations for 116 locations for (left) satellite observation DA experiment, (center) operational reference experiment, (right) no surface DA experiment.

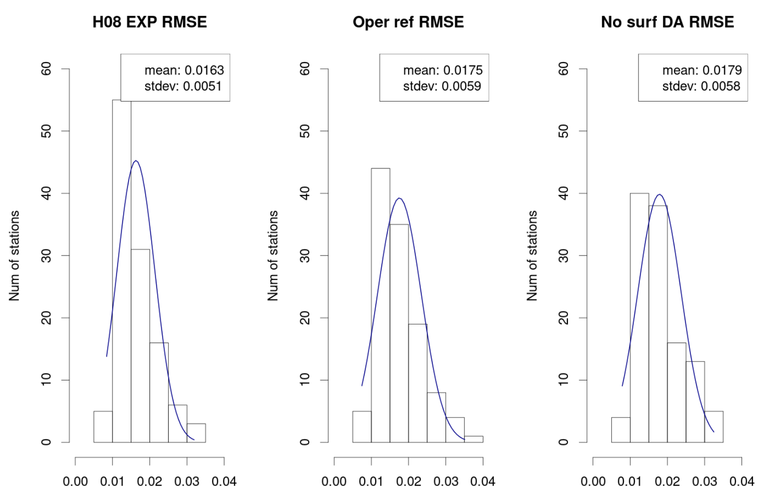



FIGURE 15. RMSE of WG1 analyses with respect to satellite observations for 116 locations for (left) satellite observation DA experiment, (center) operational reference experiment, (right) no surface DA experiment.

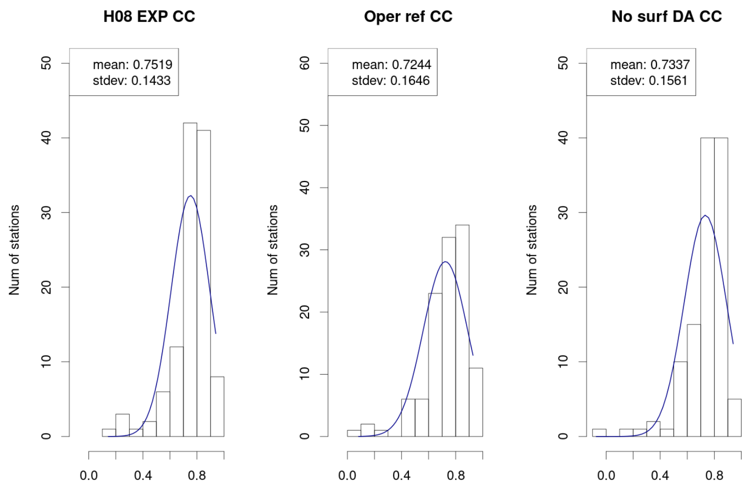



FIGURE 16. Correlation coefficient of WG1 analyses with respect to satellite observations for 116 locations for (left) satellite observation DA experiment, (center) operational reference experiment, (right) no surface DA experiment.

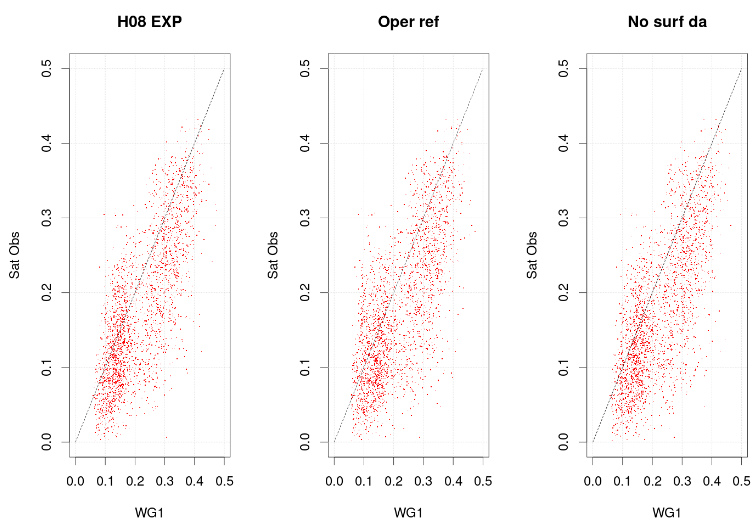



FIGURE 17. WG1 analyses against satellite observations for 116 locations for (left) satellite observation DA experiment, (center) operational reference experiment, (right) no surface DA experiment.

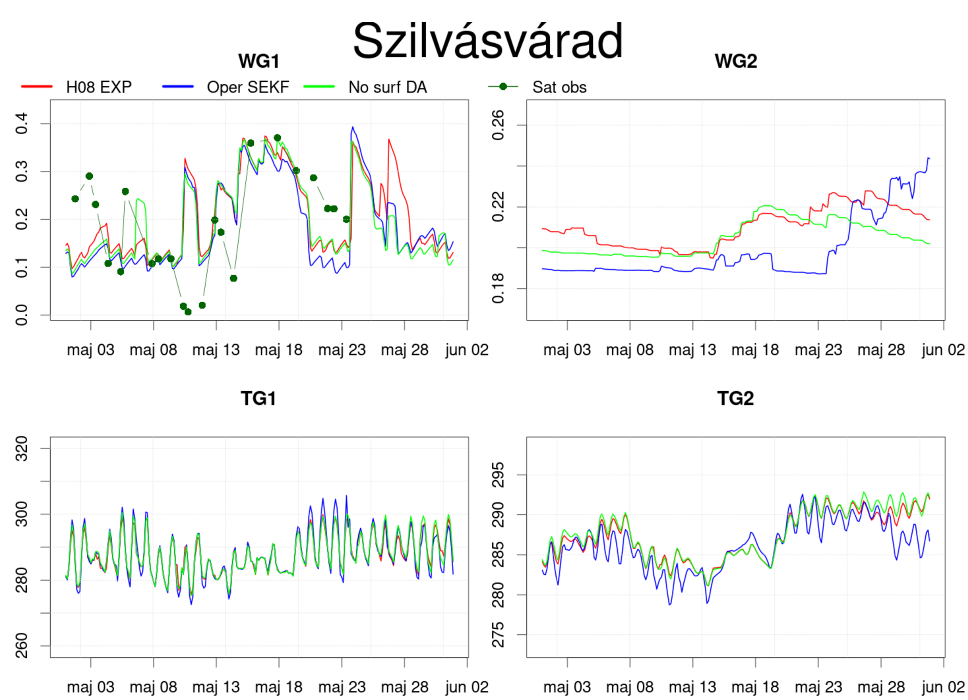
Appendix A shows the time evolution of SURFEX variables along with observation for selected points. Depending on the point, we see a wide variety in terms of agreement between the experiments and observations. In general, WG1 does not follow observations too well and it is difficult to find cases, where the satellite observation assimilation experiment would indicate a clear improvement over the other experiments. Soil temperatures TG1 and TG2 are very similar for all experiments. The difference between the no surface assimilation and satellite observation assimilation experiments is also very small, which leads us to believe that WG1 observations have only a very small impact on soil temperature. WG2 exhibits a far smaller temporal variability compared to WG1. At certain times, one or more of the experiments can feature a somewhat step-wise change in WG2, which either appears or does not appear in the other experiments.

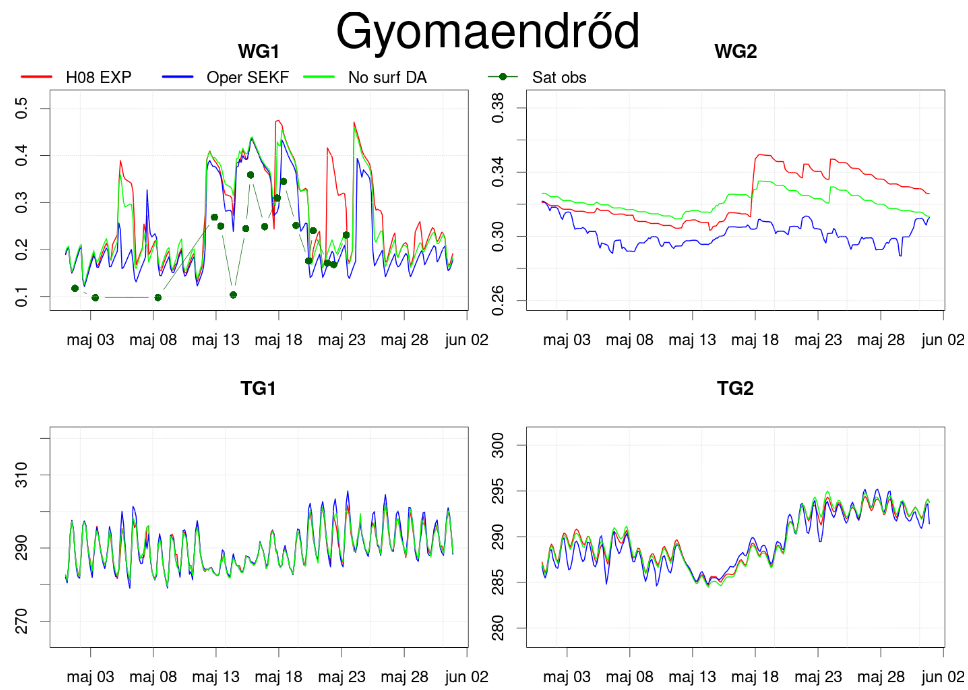
## 5 Conclusion

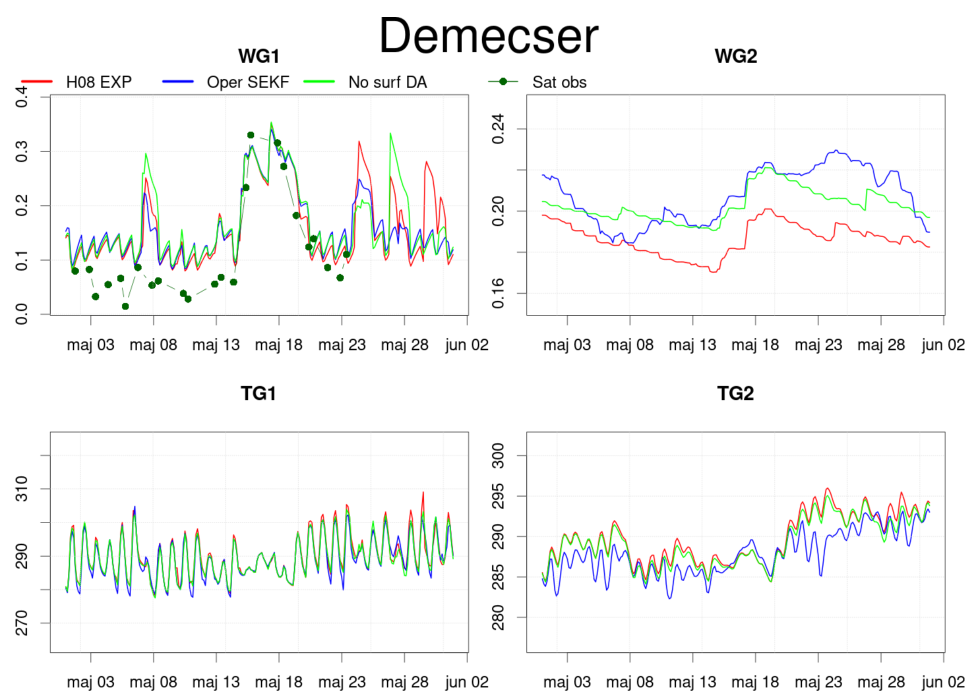
We have shown that it is feasible to perform inline data assimilation of satellite surface soil moisture observations in an operational manner for cy43t2 version of the AROME model. The results however are not satisfactory enough for operational use. Further experiments are needed to assess the reasons for somewhat poor performance. For the most part, default values and namelist settings were used. This includes the default observation error value for WG1 observations equal to 0.4. Helga Tóth reran the experiment after the conclusion of the stay with the value for observation error equal to 0.1. Some basic results are shown in appendix B. In this case the lowering of the observation error in fact degrades the results, for which a possible explanation is the poor quality of the observations themselves. This should be investigated further, by additional validation of the satellite product, reexamination of the calibration process and assessing possible discrepancies between observations and SURFEX. A run for a longer period would also be desirable to observe the behaviour of the assimilation system for other seasons, complementary to the one month in spring shown in this report. The 7 day spinup period was chosen mostly due to time constraints, however a longer spinup period might be necessary. Experiment runs with different number of control variables, and/or different values for observation and model errors should also be performed. For an operational setting it would be desirable to perform WG1 assimilation along side existing CANARI T2m and RH2m assimilation. It should be investigated if this presents technical difficulties in the current version of the code. A possible simple experiment would be to perform the assimilation of T2m and RH2m as is for all periods except 09h and 21h, for which we perform the WG1 assimilation.

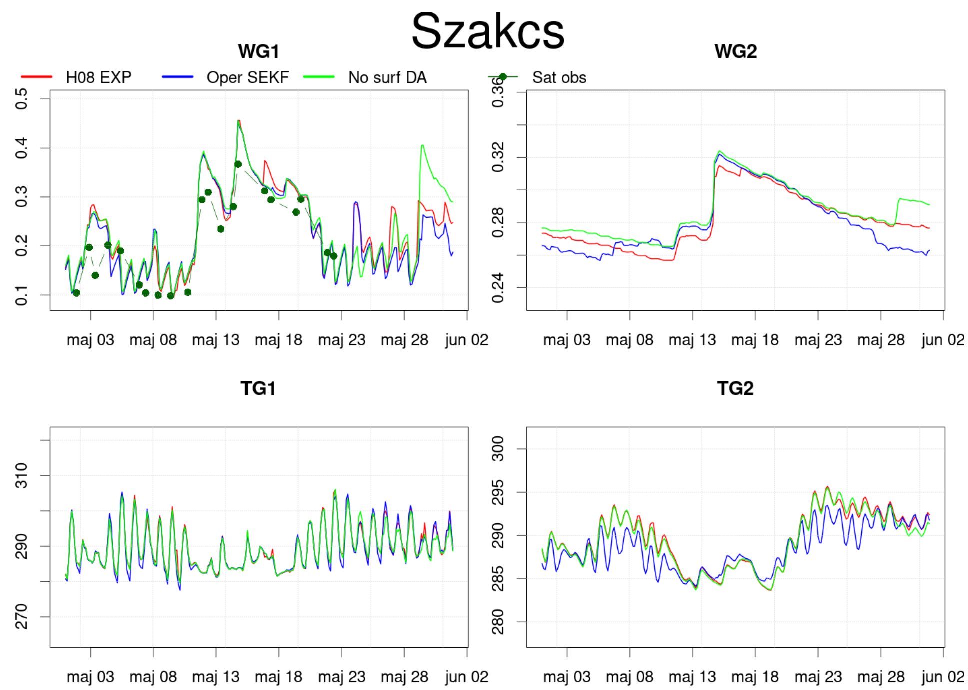
## Appendix A SURFEX analyses for different experiments and satellite observations at selected locations

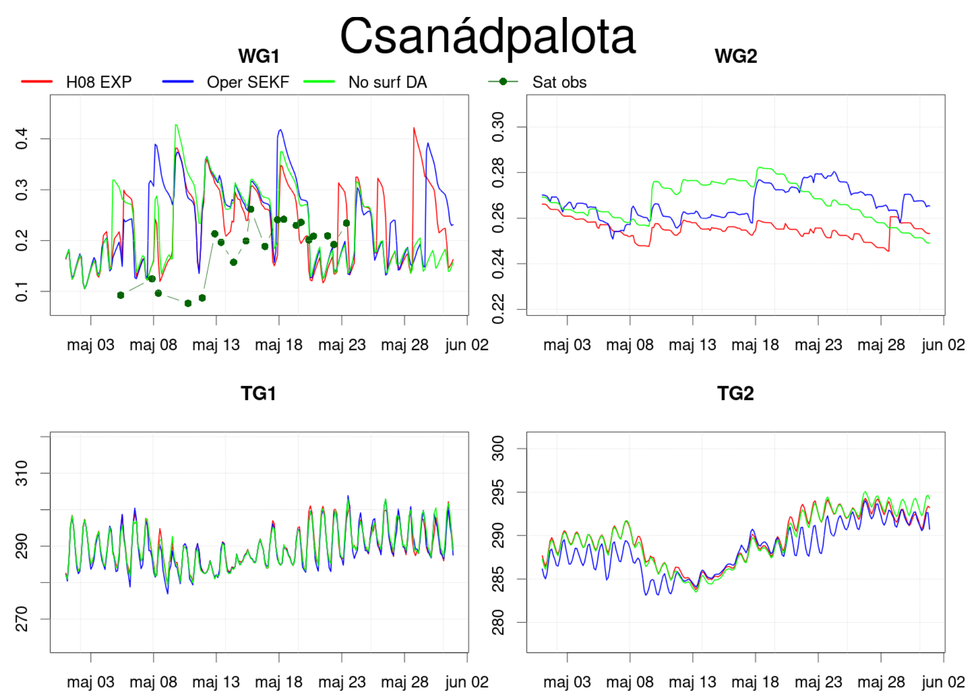
In this appendix we show the evolution of SURFEX variables for different experiments at selected points. Note the different levels of agreement between experiments and observations.

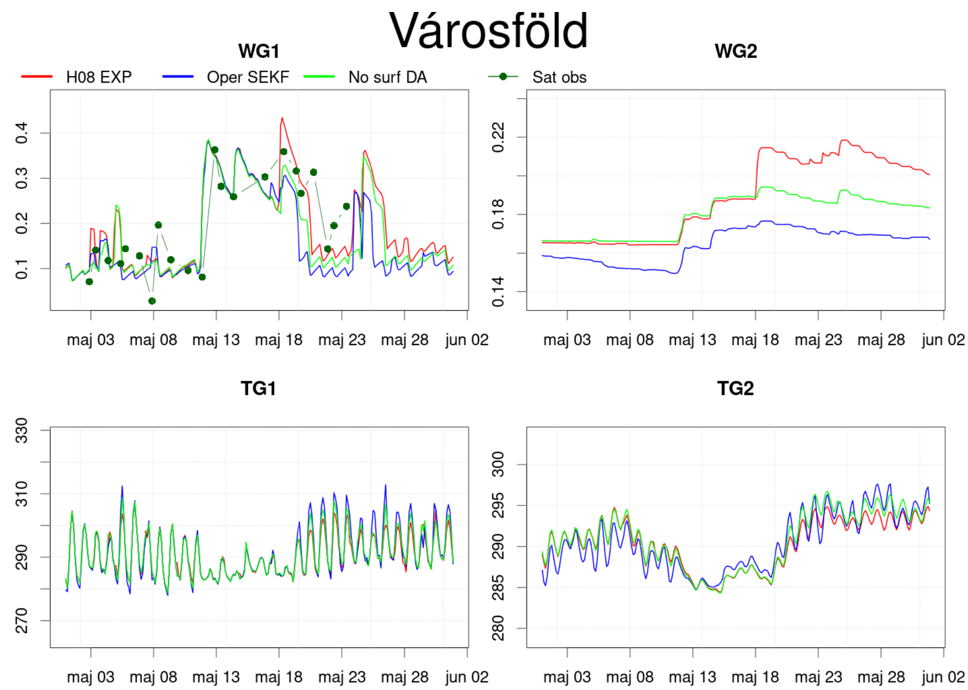



Gyomaendrőd
WG1
WG2
H08 EXP
Oper SEKF
No surf DA
Sat obs
TG1
TG2
maj 03
maj 08
maj 13
maj 18
maj 23
maj 28
jun 02


Demecser
WG1
WG2
H08 EXP
Oper SEKF
No surf DA
Sat obs
TG1
TG2
maj 03
maj 08
maj 13
maj 18
maj 23
maj 28
jun 02


Szakcs
WG1
WG2
H08 EXP
Oper SEKF
No surf DA
Sat obs
TG1
TG2
maj 03
maj 08
maj 13
maj 18
maj 23
maj 28
jun 02


Csanádpalota
WG1
WG2
H08 EXP
Oper SEKF
No surf DA
Sat obs
TG1
TG2
maj 03
maj 08
maj 13
maj 18
maj 23
maj 28
jun 02




## Appendix B Rerun of the satellite observation assimilation experiment with XERROBS_M=0.1

The default value for the surface soil moisture observation error is equal to 0.4, which is a very large value. It was discovered only after the conclusion of the stay that a more reasonable value had not been used. The experiment with an observation error of 0.1 has since been rerun by Helga Tóth. We briefly present the results on figure (18). Contrary to expectations, choosing a lower value for observation error degrades the results. We see, that compared to the old experiments, the analyses for WG1 of the rerun are closer to observations. The forecasts however show an underestimation of RH2m and overestimation of T2m compared to the other experiments. The WG1 analysis is, compared to observations again overestimated in the new experiment. Having the system account more for the observations would likely degrade the results further. This indicates a possible issue with observations themselves.

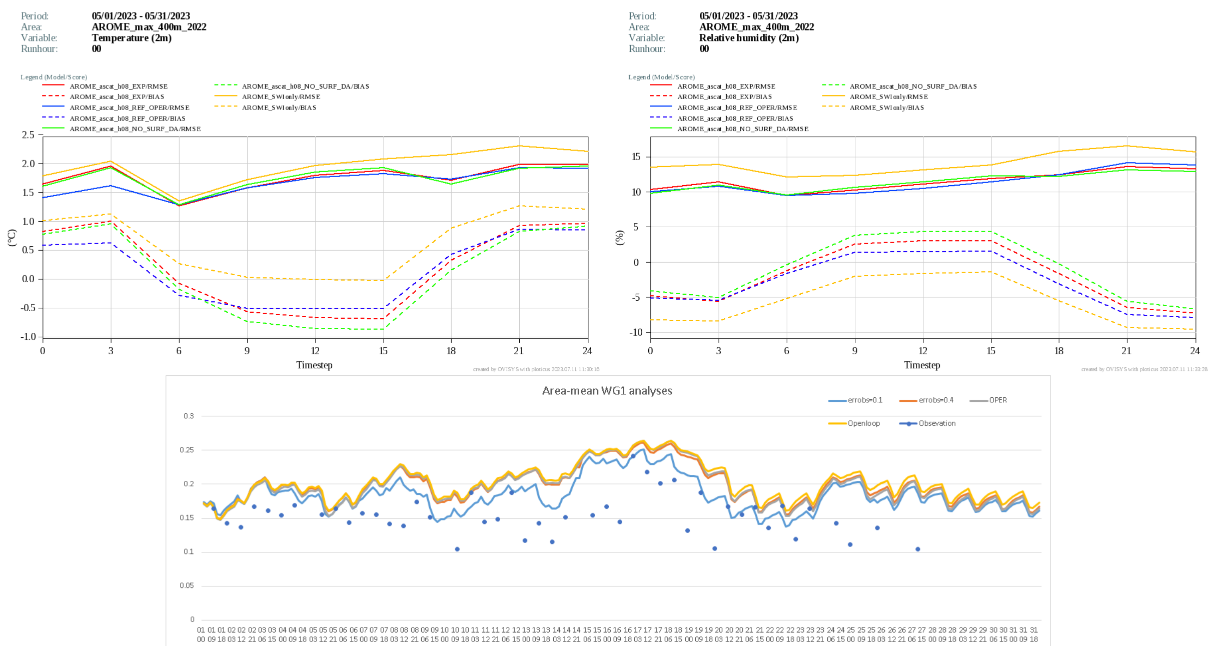



FIGURE 18. Rerun of the satellite observation experiment with XERROROBS_M=0.1. Top: Bias and RMSE of (left) T2m and (right) RH2m with respect to synop observations for 00 forecast runs. Bottom: WG1 analysis for all of the experiments including rerun along with observations.

## REFERENCES

[1] M. Ličar, 2022/09: Assimilation of Satellite Based SWI Observations using a Simplified Extended Kalman Filter Algorithm in SURFEX, RC LACE stay report

[2] EUMETSAT Satellite Application Facility on Support to Operational Hydrology and Water Management (H SAF). (2015). Product User Manual (PUM) for product H08 – SM-OBS-2, SM-OBS-2 - Small-scale surface soil moisture by radar scatterometer

[3] Reichle, R., & Koster R. (2004). Bias reduction in short records of satellite soil moisture. Geophysical Research Letters, 31(19).

[4] Operational Drought and Water Scarcity Management System homepage, https://aszalymonitoring.vizugy.hu

[5] Le Moigne, P. (2018). SURFEX scientific documentation

[6] SURFEX. n.d. SURFEX homepage. (2022). www.umr-cnrm.fr. https://www.umr-cnrm.fr/surfex/ .