SPECIAL PROJECT FINAL REPORT

All the following mandatory information needs to be provided.

Project Title:	Testbed for the Evaluation of COSMO Model Versions
Computer Project Account:	spitrasp
Start Year - End Year :	2018-2020
Principal Investigator(s)	Amalia Iriza (NMA,Romania) [1] Andrea Montani (Arpae-SIMC, Italy) [2] Ines Cerenzia (Arpae-SIMC, Italy)[2] Enrico Minguzzi (Arpae-SIMC, Italy)[2]
Affiliation/Address:	[1]National Meteorological Administration (NMA) Sos. Bucuresti-Ploiesti nr.97, 013686 Bucuresti, Romania [2]Environmental Agency of Emilia-Romagna – Hydro-Meteo-Climate Viale Silvani, 6, 40122, Bologna, Italy
Other Researchers (Name/Affiliation):	Francesco Batignani (COMET, Italy) Flora Gofa (HNMS, Greece), Rodica Dumitrache (NMA, Romania) Andrzej Wyszogrodzki (IMGW, Poland)

The following should cover the entire project duration.

Summary of project objectives

(10 lines max)

The main objective of the **"Testbed for the Evaluation of COSMO Model Versions"** Special Project is to perform testing of new COSMO model versions prior to their official release using the software environment built on the ECMWF platform during previous SPITRASP projects (2013-2015, 2016-2018). This evaluation of new model versions carried out according to source code management procedures and using the Test Suite platform is taken into account before any operational implementation and release of an official model version. The NWP test suite currently represents a benchmark for rigorous testing of all new model features and allows the model developers to produce guidelines for the selection of a new operational implementation of the model. Several model versions and configurations have been installed and tested up to now in the framework of the SPITRASP special projects, while more are expected to be evaluated using this platform.

Summary of problems encountered

(If you encountered any problems of a more technical nature, please describe them here.)

No problems encountered.

Experience with the Special Project framework

(Please let us know about your experience with administrative aspects like the application procedure, progress reporting etc.)

We consider the procedures used for application and progress reporting are clear and straightforward. Periodic reminders of resource usage and reporting deadlines are very useful. Collaboration with the administrative and support team from ECMWF was very good for the entire duration of the project.

Summary of results

(This section should comprise up to 10 pages, reflecting the complexity and duration of the project, and can be replaced by a short summary plus an existing scientific report on the project.)

First activities were performed during the frame of the COSMO Priority Task NWP Test Suite (2013-2015) and as part of previous special projects at ECMWF (2013-2015, 2016-2017).
The NWP test suite procedure was adopted by COSMO in order to perform carefully-controlled and rigorous testing, including the calculation of verification statistics, for any COSMO model test-version. Following the source code management procedure, this testing phase should offer the necessary information on the model forecasting performance, in order to determine whether the upgrade of a model test-version to a new release version is possible. For previous testing procedures, the VERSUS system was used to perform verification. For the newer simulations, the MEC (Model Equivalent Calculator) software and Rfdbk (DWD developed) system was employed for verification procedures. Model output obtained from the experiments is locally stored in the ECFS system. All the necessary software (MEC, Rfdbk) used for NWP Test suite purposes are also implemented on ecgate.

Phases I & II: Model set-up & Model Configuration and Execution of Runs

In the frame of the present Special Project, 4 model versions were employed for testing, either as operational or new releases (5.03, 5.05, 5.05_1 and 5.06). Starting with version 5.03 of the COSMO model, tests were performed on the Cray HPC available, using ECMWF computer resources both for numerical simulations and for archiving procedures. As previous versions 5.04x of the model (quasi 5.05) were only meant to test various new developments later available in the official 5.05 version, version 5.03 was also used as reference (operational) against COSMO-5.05. For all tests, int2lm version 2.05 was used for the interpolation of initial and lateral boundary conditions.

Version 5.03 was previously implemented and tested for evaluation against COSMO version 5.01 during the previous special project “COSMO NWP meteorological test suite” (2016-2018).

As a consequence, versions 5.05, 5.05_1 (version 5.05 with a bug fix) and 5.06 of the COSMO model were implemented on the Cray HPC following the procedures established during the previous special projects. Billing units were provided by the members as part of the SPITRASP special project previously registered.

Tests of model versions up to 5.03 of the model were only been performed for the 7 km horizontal resolution of the COSMO model. Starting from version 5.04a of the COSMO model, the 2.8km horizontal resolution of the model was also tested using the NWP Suite.

Starting from version 5.05, the model at 7 km horizontal resolution was integrated in two configurations, double precision (DP) and single precision (SP). Only double precision was employed for the 2.8km resolution. 7 km single precision runs are also compared to double precision runs from the same model versions (test version). During the testing of model version 5.05, procedures were also developed to move from forecast mode to hindacst mode, in order to reduce computational costs. For this purpose, the operational 5.03 version of the model was also integrated at both resolutions, in double and single precision for the 7 km horizontal resolution and in both forecast and hindcast mode, in order to be used as reference for the verification of version 5.05. Starting from version 5.06, all testing is performed in hindcast mode only.

A summary of the configurations used for testing of the different model versions is presented below:

➔ COSMO 5.05 (new test version) against 5.03 (reference/operational):

- runs at 7.0 km, 40 model levels; 72h forecast range, forecast mode, DP
- runs at 7.0 km, 40 model levels; 72h forecast range, forecast mode, SP
- runs at 2.8 km, 50 model levels; 48h forecast range, forecast mode, DP
- runs at 7.0 km, 40 model levels; hindcast mode (30 days forecast range), DP
- runs at 7.0 km, 40 model levels, hindcast mode (30 days forecast range), SP
- runs at 2.8 km, 50 model levels; hindcast mode (30 days forecast range), DP

➔ COSMO 5.06 (new test version) against 5.05_1 (reference/operational):

- runs at 7.0 km, 40 model levels; hindcast mode, DP
- runs at 7.0 km, 40 model levels; hindcast mode, SP
- runs at 2.8 km, 50 model levels; hindcast mode, DP

While previous tests and evaluations of model runs were performed for January and July 2013, starting with version 5.05 simulations are now performed for July 2017 and December 2017, 2 months in total (for each model resolution, each model version and configuration).

Configuration for COSMO 5.05 (new test version) against 5.03 (reference/operational):

For the modifications to the test suite and testing of version 5.05, initial and lateral boundary conditions were obtained from the ICON global model for the 7 km resolution. For the 2.8 km resolution, initial and lateral boundary conditions were interpolated from the 7 km model output. The forecast period of each daily run was 72 hours for 7km and 48 hours for 2.8 km. A series of new tests were performed for version 5.03 of the COSMO model to prepare for new adaptations of the test suite. The verification procedures for this model version were performed both with VERSUS and the new R-based DWD software in order to obtain an objective comparison of the results from the two verification solutions.

Configuration for COSMO 5.06 (new test version) against 5.05_1 (reference/operational):

Starting from version 5.06 of the model, evaluations are performed only in hindcast mode in both double and single precision versions for the 7 km horizontal resolution setup and only in double precision for the 2.8 km horizontal resolution configuration . For the hindcast mode, initial conditions are provided by ECMWF HRES analysis, whereas lateral boundary conditions are introduced with a 3 hourly frequency and they include the ECMWF HRES analyses (at hours 00, 06, 12 and 18UTC) and short cut off analyses (at hours 03, 09, 15 and 21UTC) with soil initialized from ICON-EU, then free soil (both model resolutions). The main features of the models used in the testing procedure are presented in Table 1.

For the simulations performed in hindcast mode, additional post-processing of the model output files was necessary in order to process the available information:

- grib1 files were remapped by changing the timeRangeIndicator, while the corresponding files for 00 UTC were used twice: as analysis for one day and as forecast from the previous day with step 24
- precipitation files were additionally processed due to the requirements of the verifications system, that uses as input accumulated fields for this parameter.

Table 1: Main features of the models used in the testing procedure.

	ECMWF HRES	**COSMO 7p0**	**COSMO 2p8**
Grid points (nx x ny)	901 x 501	661 x 471	1587 x 1147
Model levels	137	40	50
Resolution (dx x dy)	0.1 x 0.1	0.0625 x 0.0625	0.025 x 0.025

Model integration domains for all versions were the same, as can be seen from Figure 1.

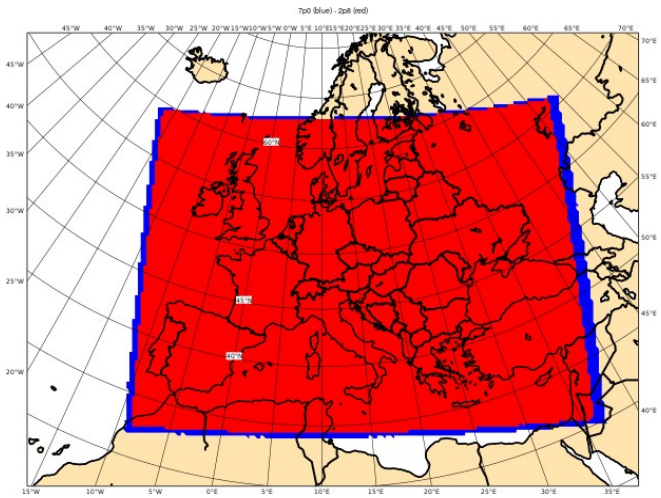

Fig. 1 Integration domain for the COSMO model at 7 km of horizontal resolution (blue) below the domain for 2.8 km of horizontal resolution (red).

The hindcast mode costs less in terms of BU and simulation time compared to the forecast mode (about one third and a half for running respectively each model configuration). Differences are mainly due to reduced time range of hindcast simulations, compared to the 72h or 48h forecast ranges set for COSMO in forecast mode at 7km and 2.8km respectively. Examples of computational costs for the different configurations of the Test Suite are presented in Tables 2 and 3.

Table 2 Cost of the suite for processing of intial and lateral boundary conditions.

INT2LM from IFS to COSMO-7km (forecast)	***INT2LM from IFS to COSMO-7km (hindcast)***
HRES → 7p0 ~ 47 BU, ~ 5min, EC_total_tasks=36, EC_nodes=1	**HRES → 7p0** ~ 17 BU, ~ 1.5min, EC_total_tasks=72, EC_nodes=1
INT2LM from ICON to COSMO-7km	
ICON → 7p0 ~ 40 BU, ~ 6min EC_total_tasks=24, EC_nodes=1	

Table 3 Cost of the suite for different configurations.

COSMO-5.03	***COSMO-5.05***	***COSMO-5.05_1***	***COSMO-5.06***
7p0_DP ~ 3000 BU ~ 28min EC_total_tasks=480, EC_nodes=20	**7p0_DP** ~ 2500 BU ~ 13min EC_total_tasks=720, EC_nodes=20	**7p0_DP** ~ 890 BU ~ 5 min	**7p0_DP** ~ 800 BU ~ 5min
7p0_DP→2p8_DP ~ 278 BU ~ 14 min 24 sec EC_total_tasks=72, EC_nodes=2	**7p0_DP→2p8_DP** ~ 278 BU ~ 14 min 24 sec	**7p0_DP→2p8_DP** ~ 115 BU ~ 6 min	**7p0_DP→2p8_DP** ~ 121 BU ~ 6 min
2p8_DP ~ 36500 BU, ~ 1h 50min 16sec EC_total_tasks=1296, EC_nodes=36	**2p8_DP** ~ 27250, ~ 1h 45 EC_total_tasks=972, EC_nodes=27	**2p8_DP** ~ 13.535 BU, ~ 53 min	**2p8_DP** ~ 12.708 BU, ~ 50 min
	7p0_SP ~ 1650 BU, ~ 9min EC_total_tasks=720, EC_nodes=20	**7p0_SP** ~ 712 BU, ~ 4 min	**7p0_SP** ~ 761 BU, ~ 4 min

Phase III: Model Output Verification

The procedure based on the Rfdbk/MEC system was been completed and has become the main verification tool employed for model evaluations. The system is made up of two components: MEC (The Model Equivalent Calculator) and Rfdbk. MEC applies the observation operators from the data assimilation scheme to model forecasts and stores the results in NetCDF feedback file format used for verification. Rfdbk is an R-based code developed to exploit the information contained in feedback files and allows loading of their content, calculation of basic verification scores and performing various convenience functions such as data adjustment and binning.

Grid-to-point comparisons were employed to compare gridded surface and upper-air model data to point observations, taking into account around 3600 selected NWP suite stations situated in an area covering -25/24/65/65 (W/S/E/N), as can be seen in Figure 2. Suspect observation values (forecast-observation greater than a specific limit are excluded) and included in the verification test to eliminate errors connected to observations. Due to the requirements of the MEC+Rfdbk software, all observations were converted in netcdf format with the bufr2netcdf software.

The verification modules for all model versions evaluated were the following:

- **surface continuous parameters** 2m temperature (T2M), 2m dew point (TD2m), wind speed (FF), total cloud cover (N), surface pressure (PS): ME (mean error), RMSE (root mean square error), SD (standard deviation), R^2, TCC (tendency correlation), LEN (number of observations), OMEAN, FMEAN (observed and forecast mean);

- **precipitation verification** (6h, 12h) for selected thresholds (greater than 0.2, 0.4, 0.6, 0.8, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 14, 16, 18, 20, 25, 30): ETS, FBI, POD, FAR, LEN etc.;

- **upper air verification (TEMP based)** – Temperature (T), dew point (TD), relative humidity (RH), wind speed (FF) and wind direction (DD) for selected pressure levels (250., 500., 700., 850., 925., 1000.): ME, MAE, RMSE, SD, etc.

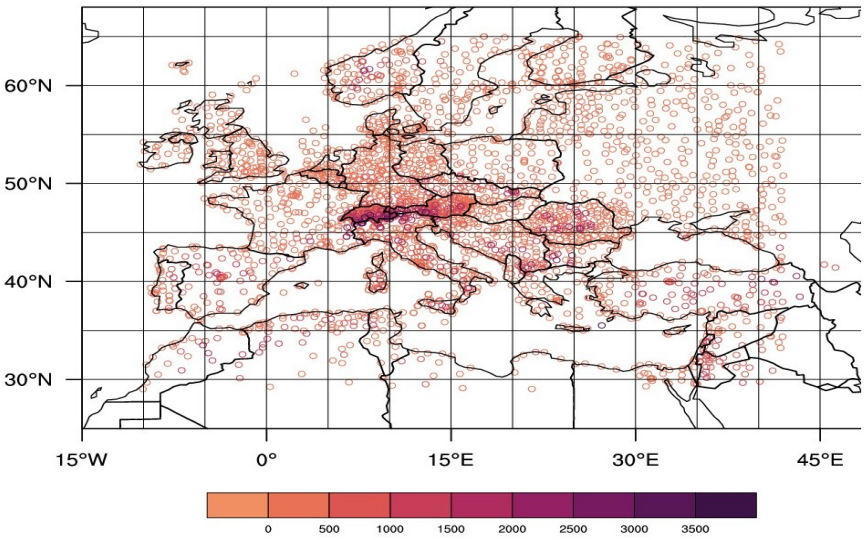

Fig. 2 Location of meteorological stations used for the verification.

Phase III: Model Output Verification

Selective verification results for the evaluated model versions are presented in the following section (Figures 3 – 11). These represent only a summary of the derived statistics. The complete overview of all the statistical analysis is available on the COSMO web-site, with detailed descriptions of the procedures and analysis of results in Cerenzia et al. (2020).

Selective results for COSMO 5.05 (new test version) single precision runs

For the single precision run, the Rfdbk based verification was performed for the 5.05 version of the COSMO model at 7 km resolution, against the double precision run of the same version, for the months of July and December 2017. The same verification procedures as for the comparison of 5.05 against 5.03 (double precision, forecast mode) were employed.

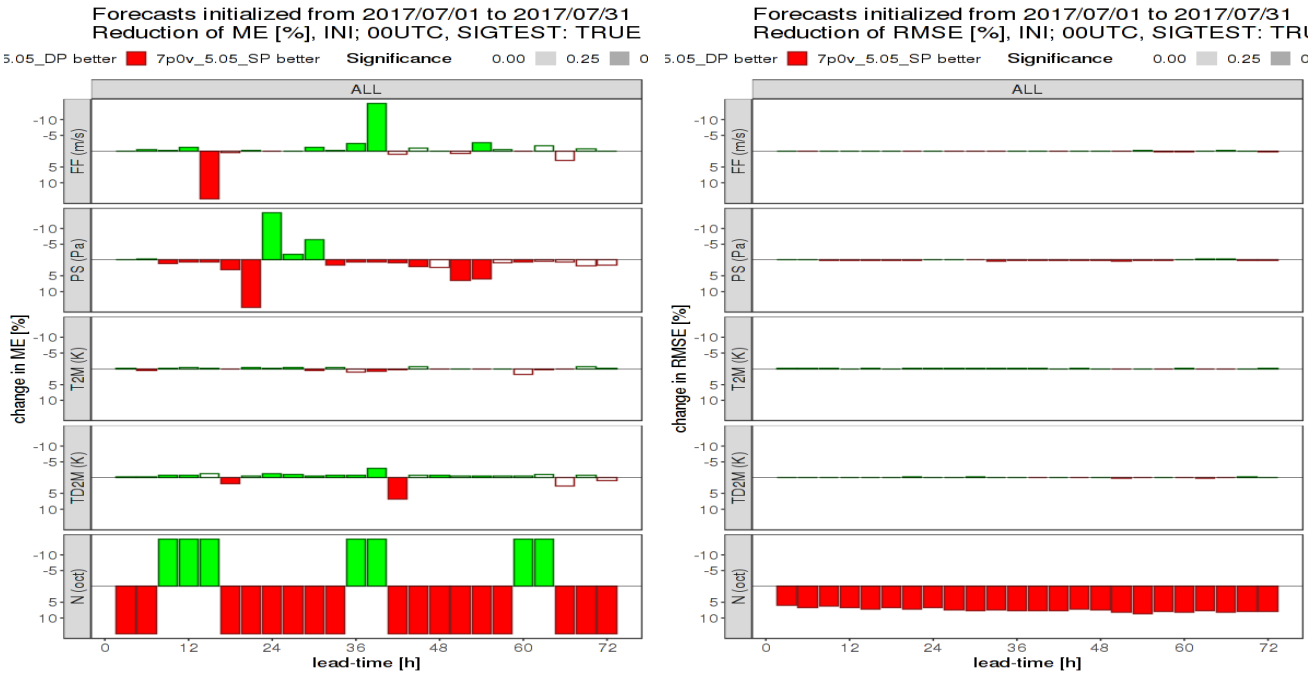

Fig. 3 COSMO-7km Continuous parameters verification results (00UTC run) for version 5.05; colors denote: SP version better (red), DP version better (green); reduction in ME (left) and RMSE (right) for July 2017. Parameters (top to bottom): FF (m/s), Ps (Pa), T2M (K), TD2M (K) and N (octa).

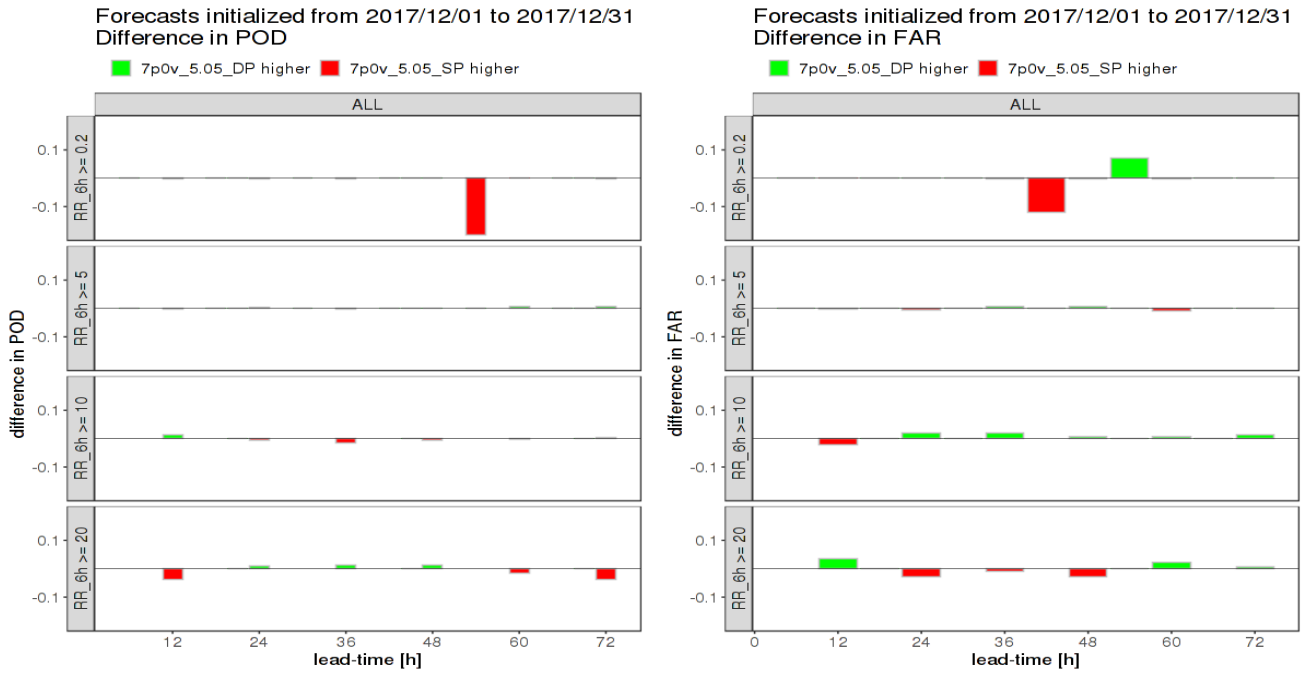

Fig. 4 COSMO-7km Verification results for 6-hour accumulated precipitation (00UTC run) for version 5.05; colors denote: values for SP version higher (red), values for DP version higher (green); reduction in POD (left) and FAR (right) for December 2017. Thresholds (top to bottom): 0.2mm/6h, 5mm/6h, 10mm/6h and 20mm/6h.

Differences between the DP and SP runs for version 5.05 were mostly insignificant, especially with regards to root mean square error, as can be seen from Figure 3. Results for the Total Could Cover needed some further investigation due to differences in the DP verification compared to the previous results, which led to the identification of a bug in the single precision version 5.05 of the model, which was later on corrected in version 5.05_1.

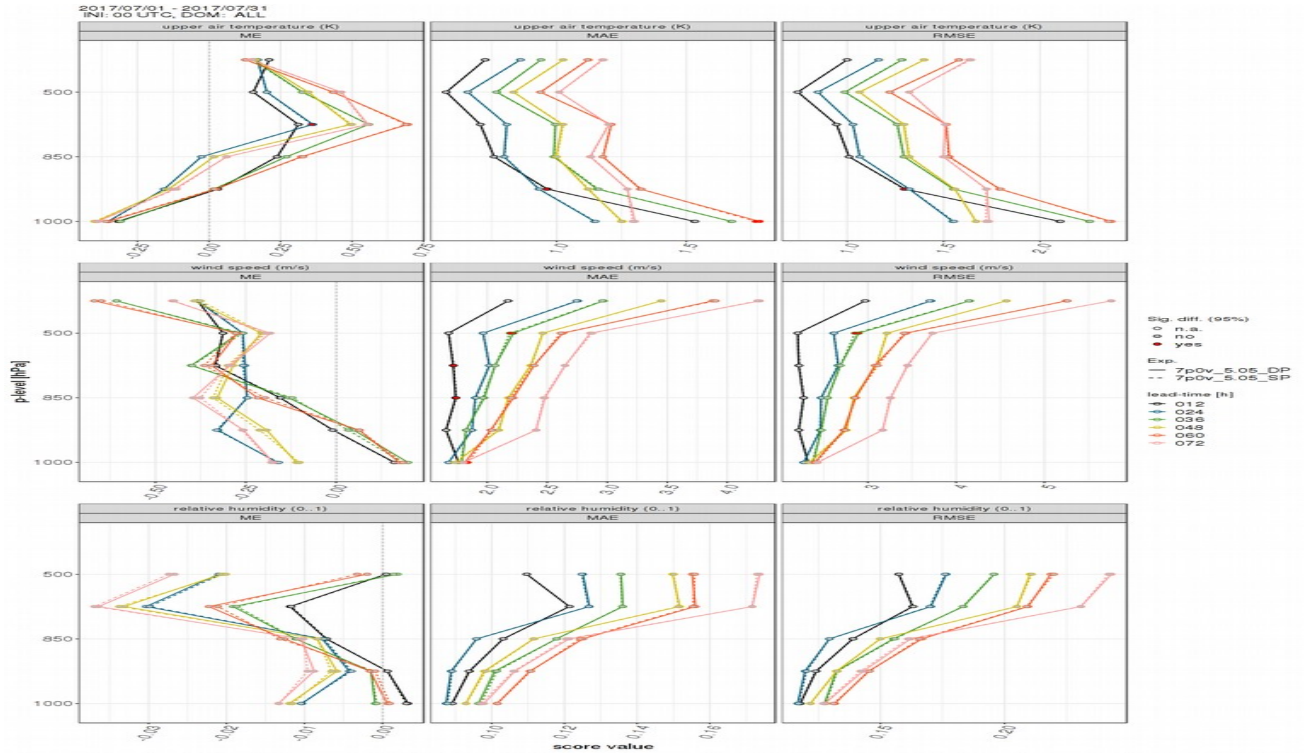

Fig. 5 COSMO-7km Upper air parameters verification results (00UTC run) for version 5.05; SP version (dashed line), DP version (solid line); scores (left to right): ME, MAE and RMSE for July 2017. Parameters (top to bottom): temperature, wind speed and relative humidity. Colors denote anticipations: +12 hours (black), +24 hours (blue), +36 hours (green), +48 hours (yellow), +60 hours (red) and +78 hours (pink).

The statistics for precipitation (6h and 12h accumulation) were quite similar, while some problematic behavior observed for some timesteps, which could not be attributed to the model performance as it was not systematic. For the upper air parameters as well, differences between the DP and the SP runs of the 5.05 version were insignificant, both for the summer and the winter months.

Selective results for COSMO 5.06 (new test version) in hindcast mode

For the DP runs (v5.05_1 against v5.06), T2M differences were insignificant for both seasons between the two model versions and the two resolutions (Figure 6). Both COSMO 5.05_1 and COSMO 5.06 underestimate the values forecasted for 2m dew point temperature for both seasons, while an increased underestimation from v5.06 is visible in July for both model resolutions. For PS, a reduction in both ME and RMSE values is obtained with the use of v5.06 (7km) compared to v5.05_1. With respect to 10 meter wind speed, behavior of NWP test results exhibit almost identical results for both seasons and resolutions for model versions v5.05_1 and v5.06, mainly underestimating observed values with the 7km resolution model and overestimating with the 2.8km one. Total Could Cover did not exhibit any change in the results between 5.05_1 and 5.06 versions for both resolutions. On the following graphs (Figures 6 - 11), version 5.05_1 of the model is denoted v5.05.

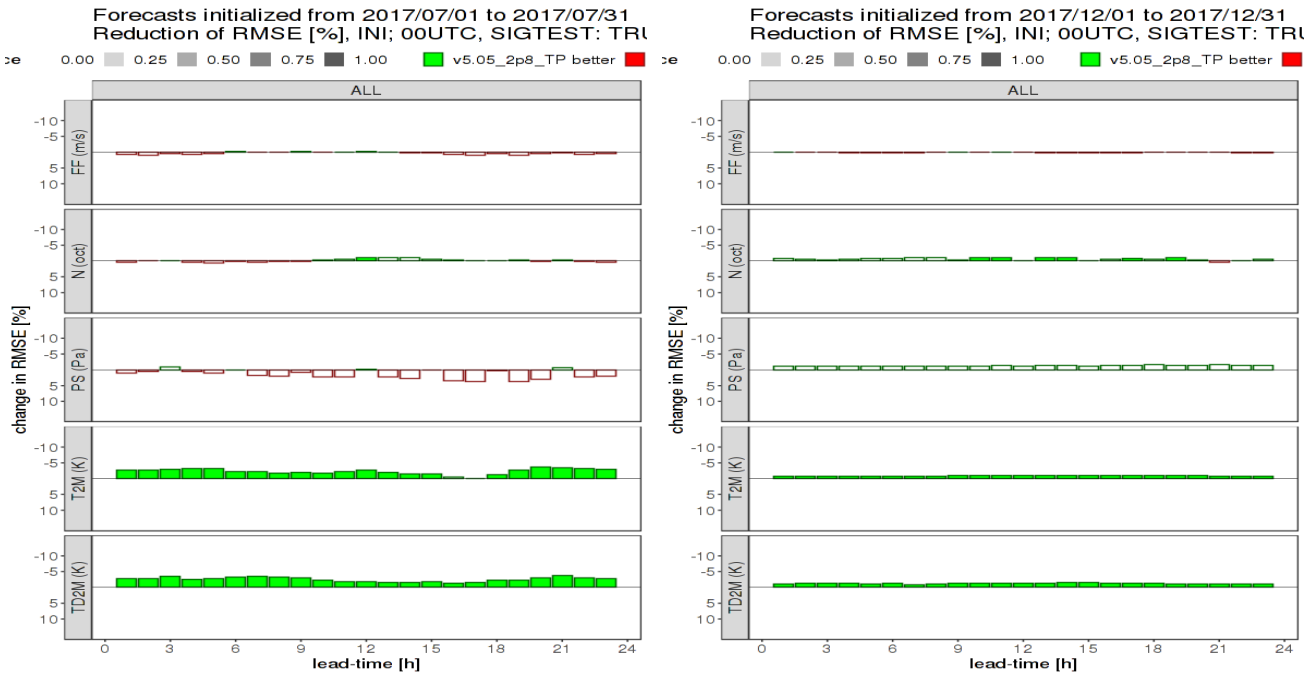

Fig. 6 COSMO-2.8km Continuous parameters verification results (00 UTC, DP runs) – reduction in RMSE for version 5.06; colors denote: v5.05_1 better (green), v5.06 better (red) for July 2017 (left) and December 2017 (right). Parameters (top to bottom): FF (m/s), N (octa), Ps (Pa), T2M (K) and TD2M (K).

The scores for the upper air parameters (Figure 7) in general also show similar behaviour for both models, with small differences between the two model versions with the 7km resolution and the 2.8km one for RH. During summer, version 5.06 performs slightly better in forecasting RH values, with reduced overestimation of RH from the surface up to almost 500mb. Overall, there is a small positive impact of version 5.06. Upper air temperature comparison for the two model versions gave insignificant differences in most cases. The outcome from FF performance comparison is similar, with no strong trend in the impact on the performance for the 7km version model. For the summer period, there is no difference with the coarse resolution models, while for the 2.8km resolution there is an increase in the overestimation almost for all levels during summer, while there is no difference in the performance during winter.

For the single precision runs, the verification was performed for the 5.06 version of the COSMO model at 7 km resolution, against the double precision run of the same version. The performance between the two precision schemes is almost identical for all surface parameters examined. For PS, a very small increase during summer for all hours is shown in the ME and RMSE values when the SP scheme is adopted. Small changes in the performance were noticed between the two precision schemes (v5.06) with regards to upper air parameters. The more noticeable differences are exhibited with upper air wind speed (Figure 7), where there are small differences during the hours of the day for both months, more obvious during the winter with respect to the ME values in upper atmosphere. For upper air relative humidity, small changes are shown in the middle atmosphere, with no specific tendency in the error, while for upper air temperature, the differences between the two precision schemes for 5.06 were shown in the upper atmospheric levels, with a very small reduction during summer.

An important observation was that the difference in the comparison between the two precision schemes for model version 5.06 were almost always smaller than the differences between model versions (v5.05_1 vs. v5.06) in DP mode.

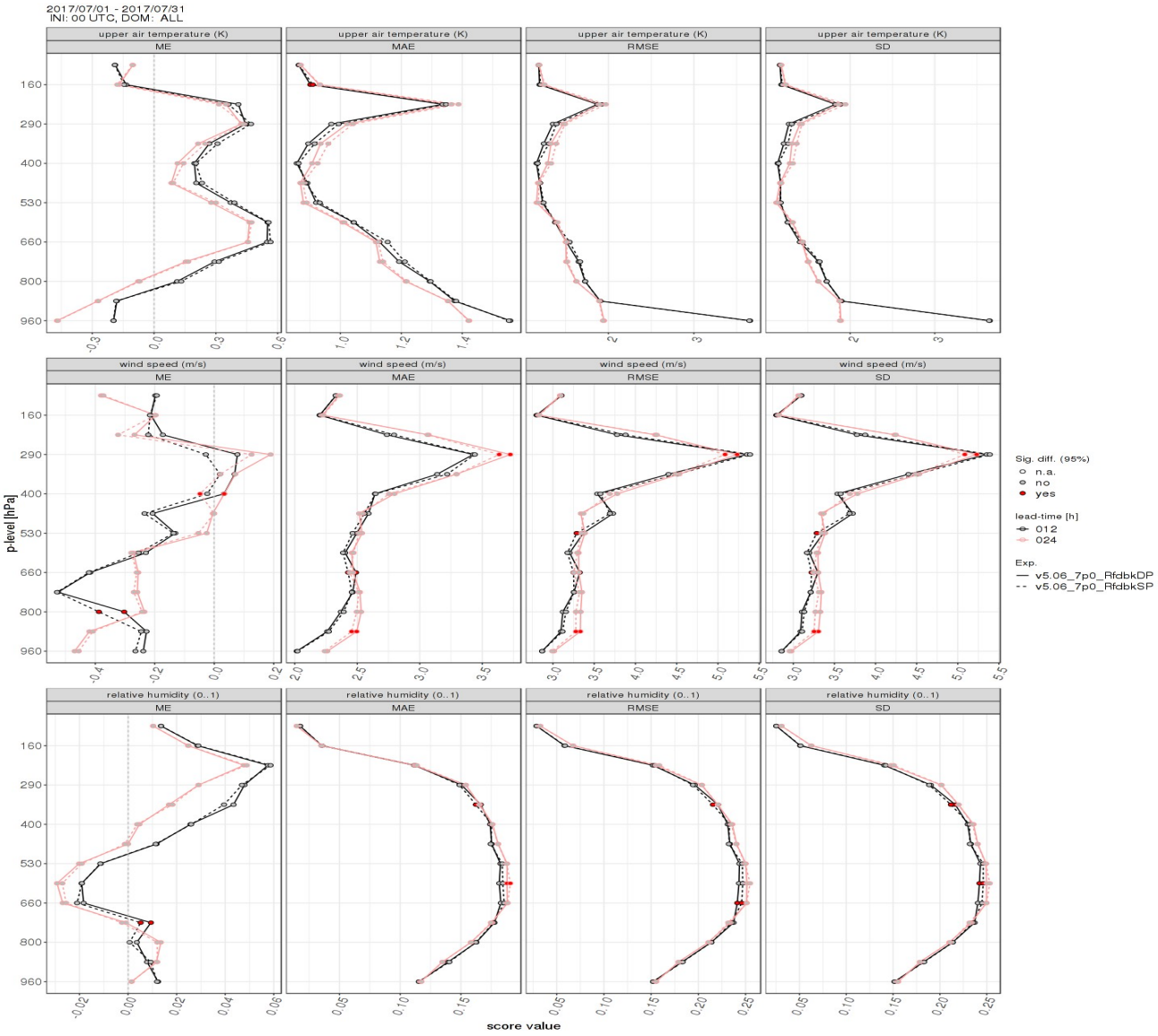

Fig. 7 COSMO-7km Upper air parameters verification results (00UTC run) for version 5.06; SP version (dashed line), DP version (solid line); scores (left to right): ME, MAE, RMSE and SD for July 2017. Parameters (top to bottom): temperature, wind speed and relative humidity. Colors denote anticipations: +12 hours (black), +24 hours (pink).

The impact of small sample size for summer was takes into account when evaluating the results for 6 hour and 12 hour accumulated precipitation. on the calculations has to be examined. As mentioned before, due to the configuration of the hindcast mode files, post processing for precipitation was necessary. As a consequence, the statistical indices could not be compared to the optimum values that would reveal the true performance of the model. Instead, the comparison was focused on the relative performance of the two model versions.

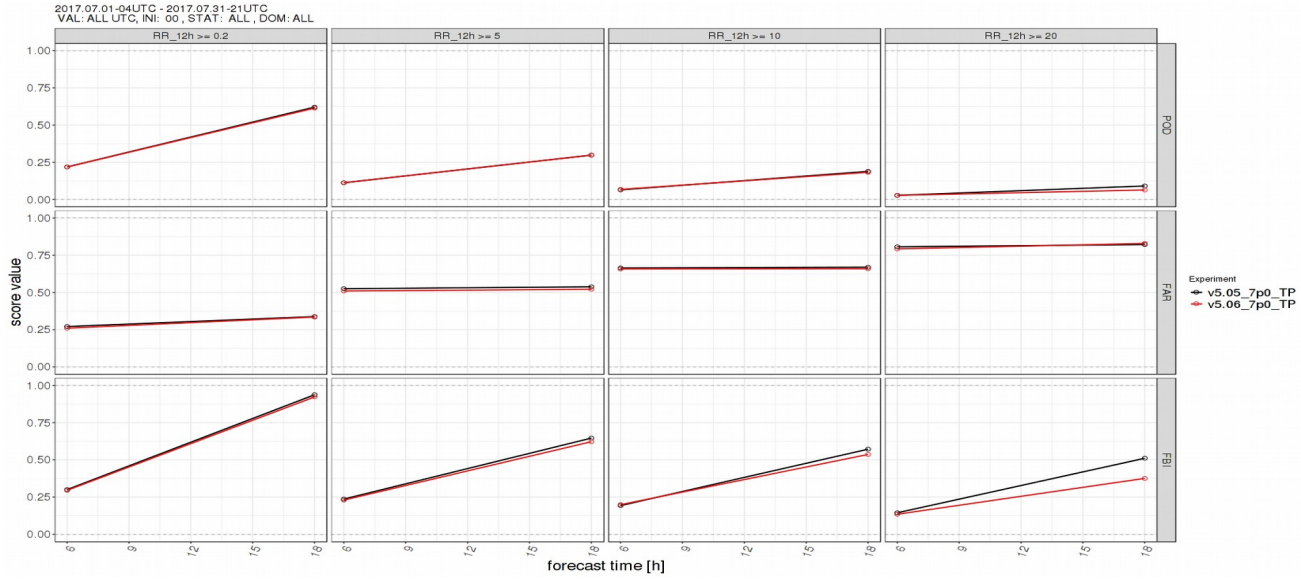

Fig. 8 COSMO-7.0km Verification results for 12-hour accumulated precipitation (00UTC run) for v5.06_1 (black) and v5.06 (red); scores (top to bootm): POD, FAR and FBI for July 2017. Thresholds (left to right): 0.2mm/12h, 5mm/12h, 10mm/12h and 20mm/12h.

For the forecast of 12 hour accumulated precipitation (Figures 8-9), the statistics between versions 5.05_1 and 5.06 of the model were also quite similar, with underestimation of precipitation amounts for all thresholds.

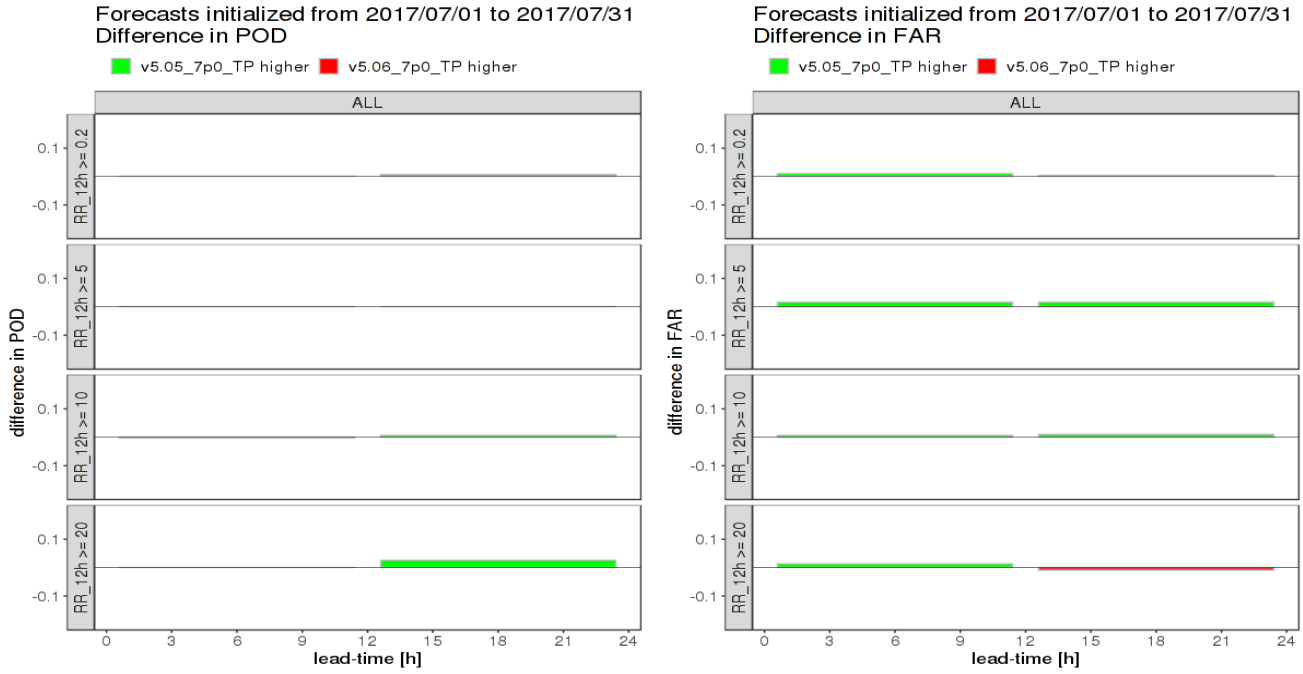

Fig. 9 COSMO-7.0km Verification results for 12-hour accumulated precipitation (00UTC run, DP runs); colors denote: values for v5.05_1 higher (green), values for v5.06 higher (red); reduction in POD (left) and FAR (right) for July 2017. Thresholds (top to bottom): 0.2mm/12h, 5mm/12h, 10mm/12h and 20mm/12h.

Generally, higher FAR and lower POD with increasing threshold can were obtained. A small deterioration of the FBI score for high thresholds was visible during summer for the 7km resolution model implementation. An opposite behaviour was observed for the 2.8km resolution configurations. The results were almost identical during winter for both model versions (5.05_1 and 5.06) at 7km resolution, with slightly more significant differences in FAR and FBI for the 20mm threshold. For the 2.8km resolution slightly larger differences were observed only in FAR, again for the 20mm threshold. The results were also similar for the DP against SP comparison of the 5.06 version of the model, with some differences in all scores in the higher threshold category.

As for previous parameters, the statistics for 6h accumulated precipitation (Figure 10 - 11) for the two versions of the model were quite similar, with slightly larger differences for the FBI score. These differences were mostly visible in the higher threshold category for both periods and resolutions. Small differences can be noticed in all scores in the higher threshold category for the winter period, for the 7km resolution of the model, while for the 2.8km resolution some differences are visible in FAR for the highest threshold and in FBI for the 10mm and 20mm thresholds.

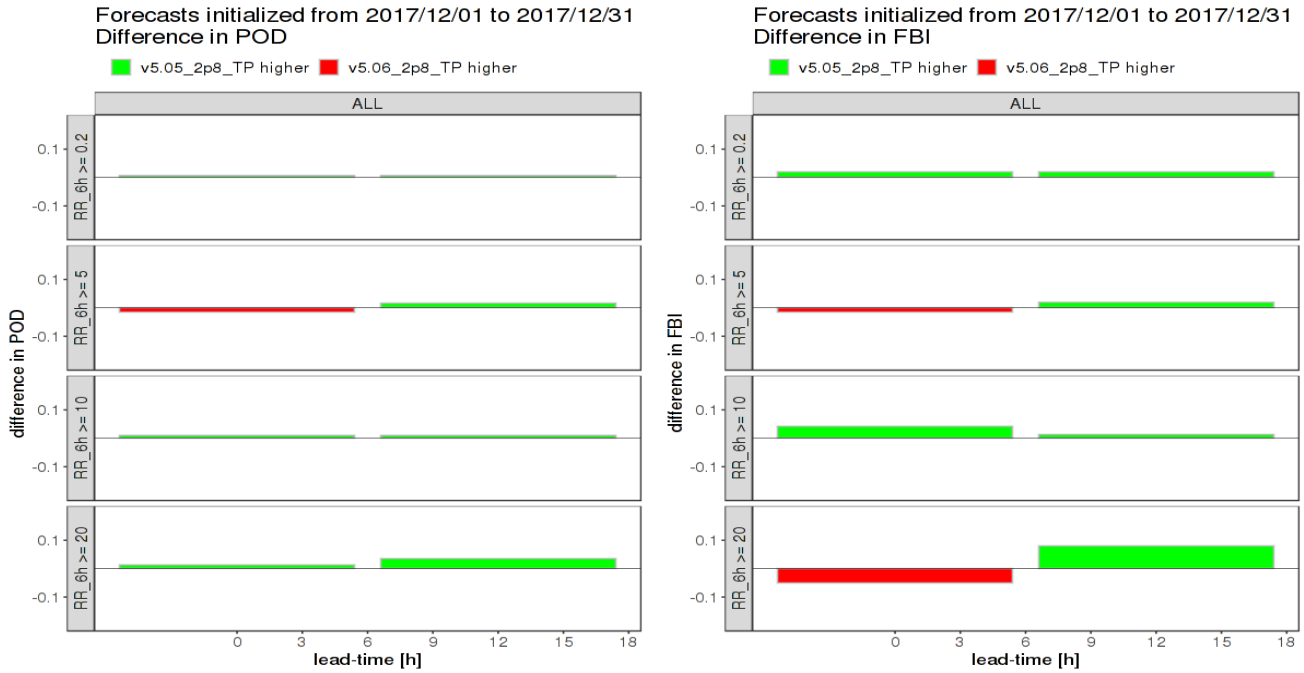

Fig. 10 COSMO-2.8km Verification results for 6-hour accumulated precipitation (00UTC run, DP runs); colors denote: values for v5.05_1 higher (green), values for v5.06 higher (red); reduction in POD (left) and FBI (right) for December 2017. Thresholds (top to bottom): 0.2mm/6h, 5mm/6h, 10mm/6h and 20mm/6h.

With regards to the DP versus SP comparison (Figure 11), results were almost identical also for December, with some differences in FBI for the higher threshold categories.

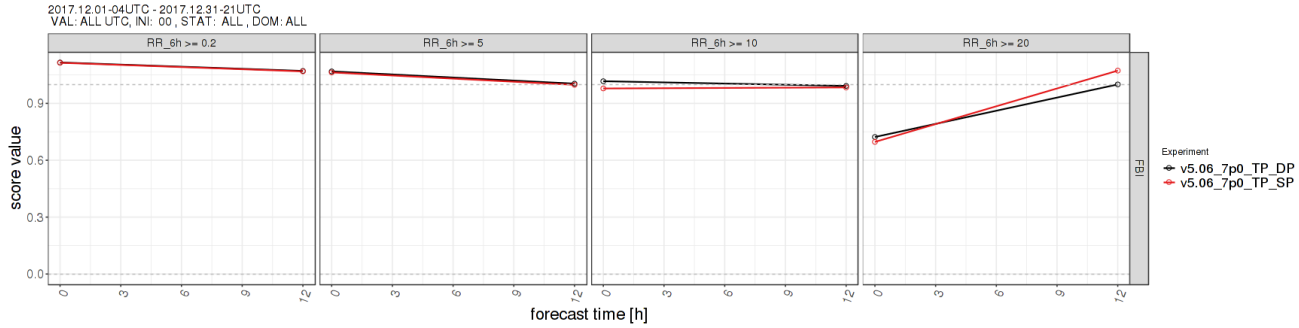

Fig. 11 COSMO-7km Upper air parameters verification results (00UTC run) for version 5.06; SP version (dashed line), DP version (solid line); scores (left to right): ME, MAE, RMSE and SD for July 2017. Parameters (top to bottom): temperature, wind speed and relative humidity. Colors denote anticipations: +12 hours (black), +24 hours (pink).

Following the analysis of the relative performance of the two new model versions: 5.05 and 5.06 against the operational ones (5.03 and 5.05_1), the newer versions of the model, on the whole, presented no significant changes in performance, while in a few cases outperformed their respective predecessors, or were slightly worse.

For a more detailed descriptions of model configurations and a thorough analysis of the results, we refer to Cerenzia et. al (2020), while the complete overview of all the statistical analysis is available on the COSMO web-site.

List of publications/reports from the project with complete references

M. Milelli: “WG6 Activities (+ PP CEL-ACCEL + PP IMPACT + COSMO LEPS + COSMO Web)”, The 22st COSMO General Meeting, 1-11 September 2020, teleconferences

I. Cerenzia: “NWP Test Suite”, The 22st COSMO General Meeting, 1-11 September 2020, teleconferences

A. Iriza-Burca: “NWP Test Suite update”, The 22st COSMO General Meeting, 1-11 September 2020, teleconferences

I. Cerenzia, A. Iriza-Burca, M. Bogdan, F. Gofa, F. Fundel, H. Reich (contributors) - *“Numerical Weather Prediction Meteorological Test Suite”: COSMO 5.06 vs. 5.05_1*, COSMO-Model Report, June 2020

M. Milelli: “Other WG6 Activities”, The 21st COSMO General Meeting, Rome, Italy, 9 - 13 September 2019

F. Gofa: "Verification and Case Studies. overview of Activities", The 21st COSMO General Meeting, Rome, Italy, 9 - 13 September 2019

I. Cerenzia: “NWP Test Suite”, The 21st COSMO General Meeting, Rome, Italy, 9 - 13 September 2019

A. Iriza-Burca: “NWP Test Suite Suggestions”, The 21st COSMO General Meeting, Rome, Italy, 9 - 13 September 2019

Future plans

(Please let us know of any imminent plans regarding a continuation of this research activity, in particular if they are linked to another/new Special Project.)

The current research activity which includes the evaluation of COSMO versions through a defined procedure (the NWP test suite) will be continued for the final model versions (and added configurations). These activities, together with the extension of the Test Suite to evaluate ICON model official releases in the context of the COSMO to ICON-LAM – migration are currently carried on in the frame of the „COSMO and ICON Numerical Weather Prediction Test Suite" special project approved for 2021-2023.