Seventh Framework Programme
Theme 6 [SPACE]

Project: 607193 UERRA

Full project title:
Uncertainties in **E**nsembles of **R**egional **R**e-**A**nalyses

Deliverable D2.13:
Ensemble Nudging diagnostics report and documentation

WP no:	2
WP leader:	Met Office
Lead beneficiary for deliverable :	University of Bonn
Name of author/contributors:	Maarit Lockhoff, Michael Borsche and Liselotte Bach
Nature:	Report
Dissemination level:	PU
Deliverable month:	45
Submission date:	24 November 2017

Report for Deliverable 2.13 (D2.13): Scientific report on assessment of ensemble nudging diagnostics

Maarit Lockhoff[1], Michael Borsche[2] and Liselotte Bach[2]

[1] University of Bonn, Gernany
[2] Deutscher Wetterdienst, Offenbach, Germany

1. Introduction

The University of Bonn has completed the production and archiving of five years of the regional ensemble reanalysis system COSMO-EN-REA12. In this report, we present the work on deliverable D2.13 dealing with diagnostics of the ensemble nudging data assimilation system used for the production of the COSMO-EN-REA12 reanalysis. It aims at demonstrating the quality of the production system. This includes an investigation of the observation input stream, the analysis increments and the analysis departure statistics. In order to get a more deep insight into the quality of the uncertainties provided by the ensemble, results from a comparison against observations will be summarized. Section 2 gives details about the system design and development. Observation monitoring, assimilation statistics and verification results are presented in section 3, 4 and 5, respectively. A summary will be given in Section 6.

2. Ensemble reanalysis system

This section is a review of the description of the reanalysis framework outlined in Bach et al. 2016. For more details the interested reader is therefore referred to publication mention beforehand.

The COSMO-EN-REA12 reanalysis system builds upon the deterministic nudging regional reanalysis system COSMO-REA6 that uses the limited-area model COSMO (Consortium for smallscale Modeling, Schättler et al., 2011). The COSMO model is non-hydrostatic and targeted at the representation of meso-alpha and mesobeta processes. The main new development within the COSMO-EN-REA12 encompasses the incorporation of uncertainty estimation capabilities. For the ensemble reanalysis purpose we adapted it to a grid resolution of 12 km. The geographical extension remains the same as compared to COSMO-REA6 covering the CORDEX-EUR11 domain (Giorgi, 2009) as can be seen from Fig. 1. In the employed version, COSMO has 40 hybrid levels in the vertical. The soil model TERRA makes use of 7 vertical layers going down to approximately 14.5m depth. The model equations are solved on a rotated latitude-longitude grid that avoids a convergence of the meridians and allows for equidistant grid points. The applied ensemble data assimilation method is an ensemble nudging scheme (Schraff 1997 and Bach et al. 2016). It includes external analyses of snow, sea surface temperature and a variational soil moisture analysis. In reanalysis mode, ERA-Interim is employed as lateral boundary conditions. The ensemble reanalysis data (comprising 21 members) and corresponding reforecasts are available from the ECMWF MARS archive for a 5 year time period (2006-2010).

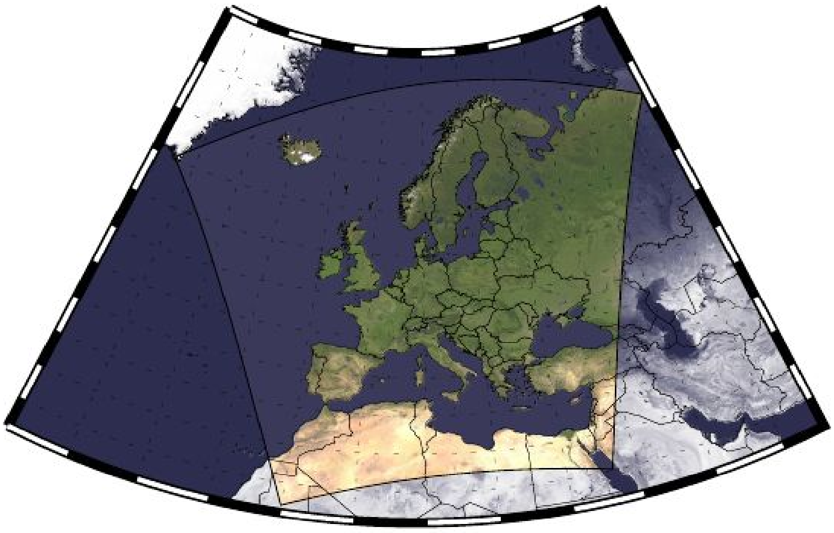

Fig. 1: Domain of COSMO-EN-REA12 (Bach et al. 2016).

Fig. 2. gives an overview on how the individual components described above have been integrated into the processing system used for the COSMO-EN-REA12 production. The initial conditions needed at the beginning of the reanalysis period as well as the 3-hourly boundary conditions are provided by ERA-Interim. Every 6 hours, the nudging runs are interrupted to perform a snow analysis. Once per day at 00 UTC, the sea surface temperature and the soil moisture are updated. This is done for all ensemble members separately. External parameters including leaf area index, plant cover, root depth, carbon dioxide concentrations and an ozone maximum are updated once a day according to a prescribed annual cycle. The three-dimensional fields of the dynamically relevant quantities on model levels are stored in 6-hourly intervals while the surface fields and fields on pressure and height levels are archived at an hourly frequency. The system additionally incorporates reforecasts that are initialized in 6-hourly intervals and have twice a day lead times of 30 hours and twice a day of 6 hours.

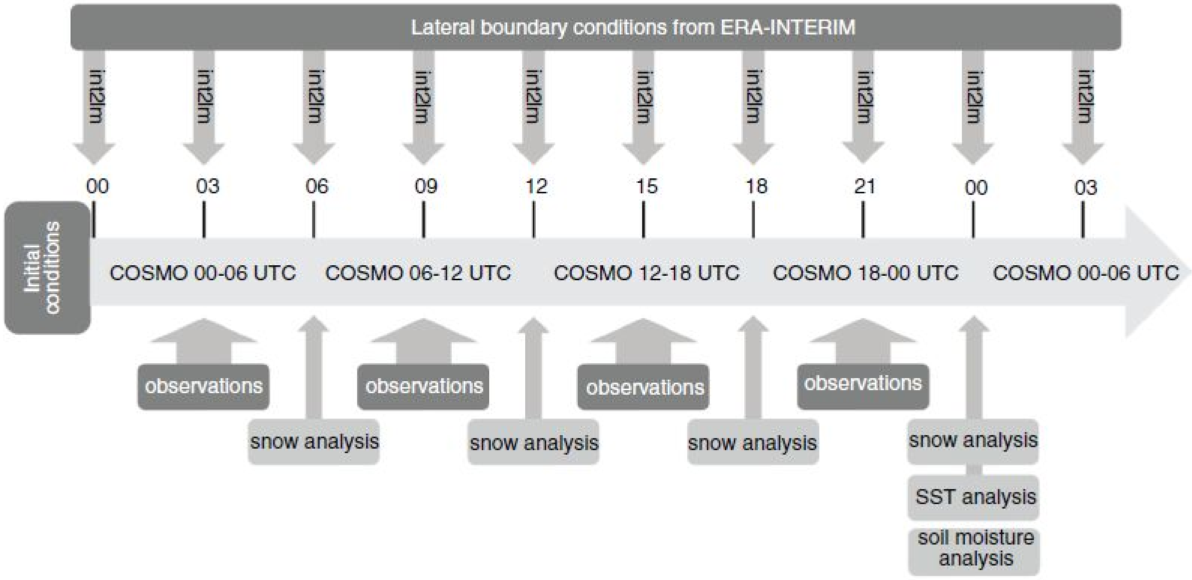

Fig. 2: Process cycle of the ensemble nudging system for one exemplary ensemble member (Bach et al. 2016).

2. Observation monitoring

The observation stream used within the COSMO-EN-REA12 data assimilation consists of conventional observations only. The observation types and corresponding variables that are assimilated are shown in Table 2 together with the mean number of assimilated reports per 6h cycle. Surface level observations include manual and automatic reports from SYNOP stations as well as manual and automatic SHIP reports and drifting buoys (DRIBU). The upper-air observations come from aircraft observations which can be further divided into aircraft reports (AIREP), automatic reports of the type AMDAR (Aircraft Meteorological Data Relay) and reports from ACARS (Aircraft Communication Addressing and Reporting System). More upper-air observations are provided by radiosonde ascents (TEMP) and pilot balloon ascents (PILOT). Moreover, upper-air wind data from wind profiler are assimilated. The most important observation sources are (land surface-level) SYNOP and aircraft reports with appr. 23500 and 18000 reports per 6h cycle respectively. These two sources alone provide already more than 90 % of the observation input.

Fig. 3 shows time-series of total number of observations used by the nudging system for COSMO-EN-REA12 (top) for the years 2006 to 2010 including lines for the active (dark blue), passive (light blue) and rejected (red) observations. It can be seen that the total number of assimilated (active) reports slightly increases over time. The number of rejected reports is very low. This is because the

COSMO-EN-REA12 observation stream uses the information from feedback observation files of the COSMO-REA 6 observation stream as input. In order to get the full picture of the observation monitoring, statistics for REA6 have to be considered as well and are therefore additionally provided in the lower panel of Fig. 3.

Table 1: Observation systems and types as well as corresponding assimilated variables and mean number of assimilated reports per 6 hour cycle used in the ensemble nudging scheme.

Observing system	Observation type	Assimilated variables	Mean number of assimilated reports per 6h
Surface-level observations	SYNOP	surface pressure, wind, humidity	23427
	SHIP	surface pressure, wind, humidity	1321
	DRIBU	surface pressure, wind, humidity	461
Aircraft	AIREP	wind, temperature	522
	AMDAR	wind, temperature	7438
	ACARS	wind, temperature	10308
Radiosondes	PILOT	upper-air wind	8
	TEMP	upper-air wind, temperature, screen-level wind, humidity, geopotential	224
Wind profiler		upper-air wind	370

Concerning the temporal evolution of the number of active reports there are only a few very short-term drops in number of reports most pronounced in January 2009 but the time series does not show any systematic jumps or breaks. This is different for the passive reports for which around July 2008 the number of reports drops to a lower level until end of the year 2010 where the number of reports increase to the level before July 2008. This behavior is also visible in the COSMO-REA6 time series of passive and rejected reports. What causes this feature is yet unclear, at the time of writing. Still these features do not impact the quality of the data assimilation as only passive and rejected reports are affected.

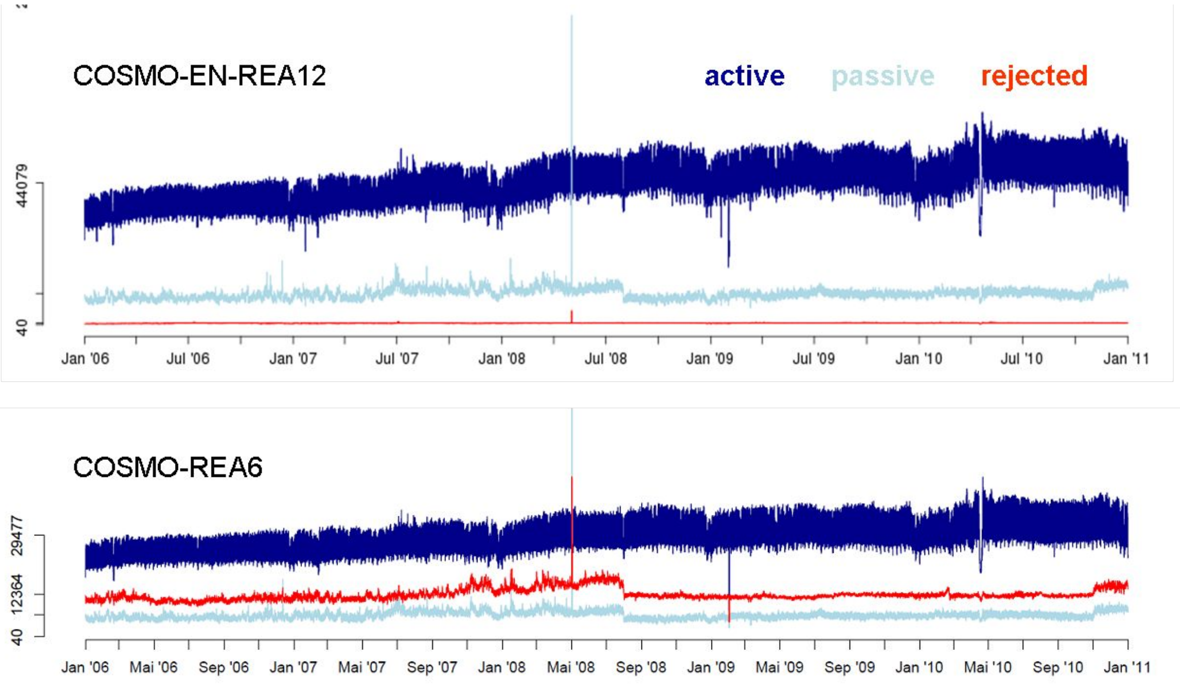

Fig. 3: Time series of number of active (dark blue), passive (light blue) and rejected (red) observations per 6-hour cycle for the entire period from January 2006 to December 2010. Top and bottom panels show total number of observation from COSMO-EN-REA12 and COSMO-REA6 respectively.

Fig. 4 shows time series of number of reports for different observation system (from top to bottom): SYNOP, aircraft (AIREP) and radiosondes (TEMP). Both SYNOP and aircraft observations provide a stable input stream over time with only few short-term drops in the number of observations. For the radiosondes the break in July 2008 is also visible in the active reports and leads to an increase in the number of observations. As the number of radiosonde observations is small compared to the number of aircraft reports the impact is expected to be small, at least as far as temperature and wind are concerned.

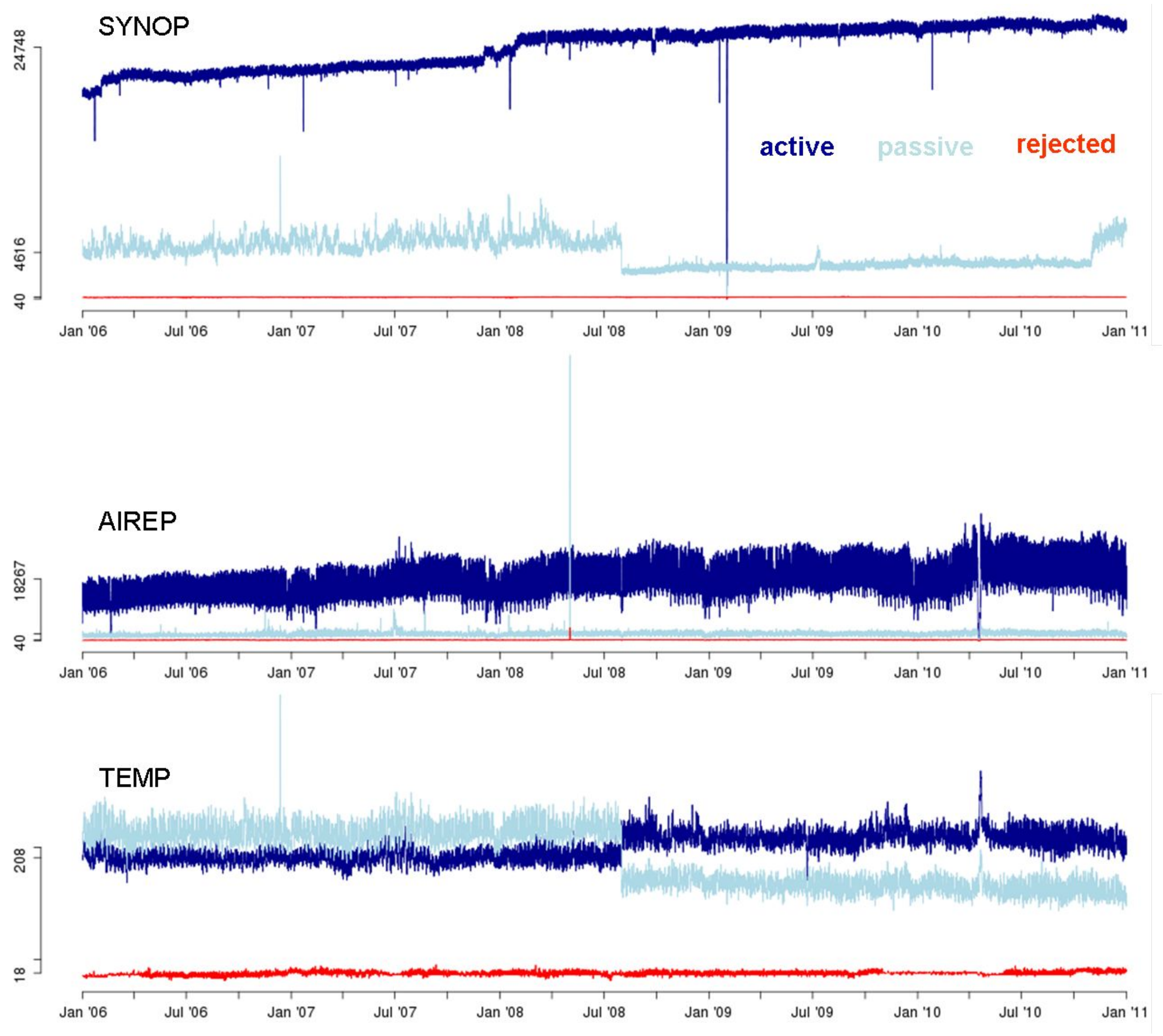

Fig. 4: as Fig. 1 but for SYNOP, aircraft (AIREP) and radiosonde (TEMP) observation systems.

3. Assimilation Statistics

This section includes an investigation of the analysis increments and the analysis departures (difference between observation and analysis values) in order to monitor and assess the quality of the COSMO-EN-REA12 data assimilation.

3.1 Analysis increments

Analysis increments are the adjustments made to the model variables by the data assimilation. They provide important diagnostics of the performance of the data assimilation system. Given a constantly dense observation network, analysis increments can be used to reveal systematic errors, for example biases in the forecast model.

Fig. 5 shows daily averages of the 6-hourly analysis increments for temperature (top) and wind speed (bottom) for all 40 model levels for an arbitrarily chosen member over the full domain. The overall small values, staying within ± 0.2 K and ± 0.2 ms^{-1} for temperature and wind respectively, and the absence of any temporal trends show that the model is performing reasonably well. These findings are in line with results found by Bollmeyer et al. (2015) for COSMO-REA6. Still there are peculiar features visible. These include for temperature the tendency of the model to be too cold during the wintertime at the surface and up to 900 hPa, and to be too warm during the summer (especially in the layer between 900 and 500hPa). Over the whole year, high positive values occur close to the upper boundary (> 150 hPa) indicating too cold model temperatures.

For wind speed there is among others a pronounced band of negative increments (up to -0.40ms^{-1}), that reaches from the surface up to 900 hPa. This feature points at boundary-layer winds being too strong throughout the year with a seasonal cycle in the vertical extension. Largest vertical extensions of the negative band are reached during summer (up to 900hPa) while during summer only height levels up to 950hPa are affected. This feature was as well already described by Bollmeyer et al. 2015 for the COSMO-REA6 and was attributed to problems with boundary layer processes such as e.g. mixing in the boundary layer. The middle troposphere around 400 – 600 hPa is characterized by slightly negative values whereas positive values are found in the upper troposphere/lower stratosphere staying in both cases well below to ± 0.2ms^{-1}.

As expected, for none of the two variables there is an effect of the aforementioned break in July 2008 in the TEMP radiosondes time series visible.

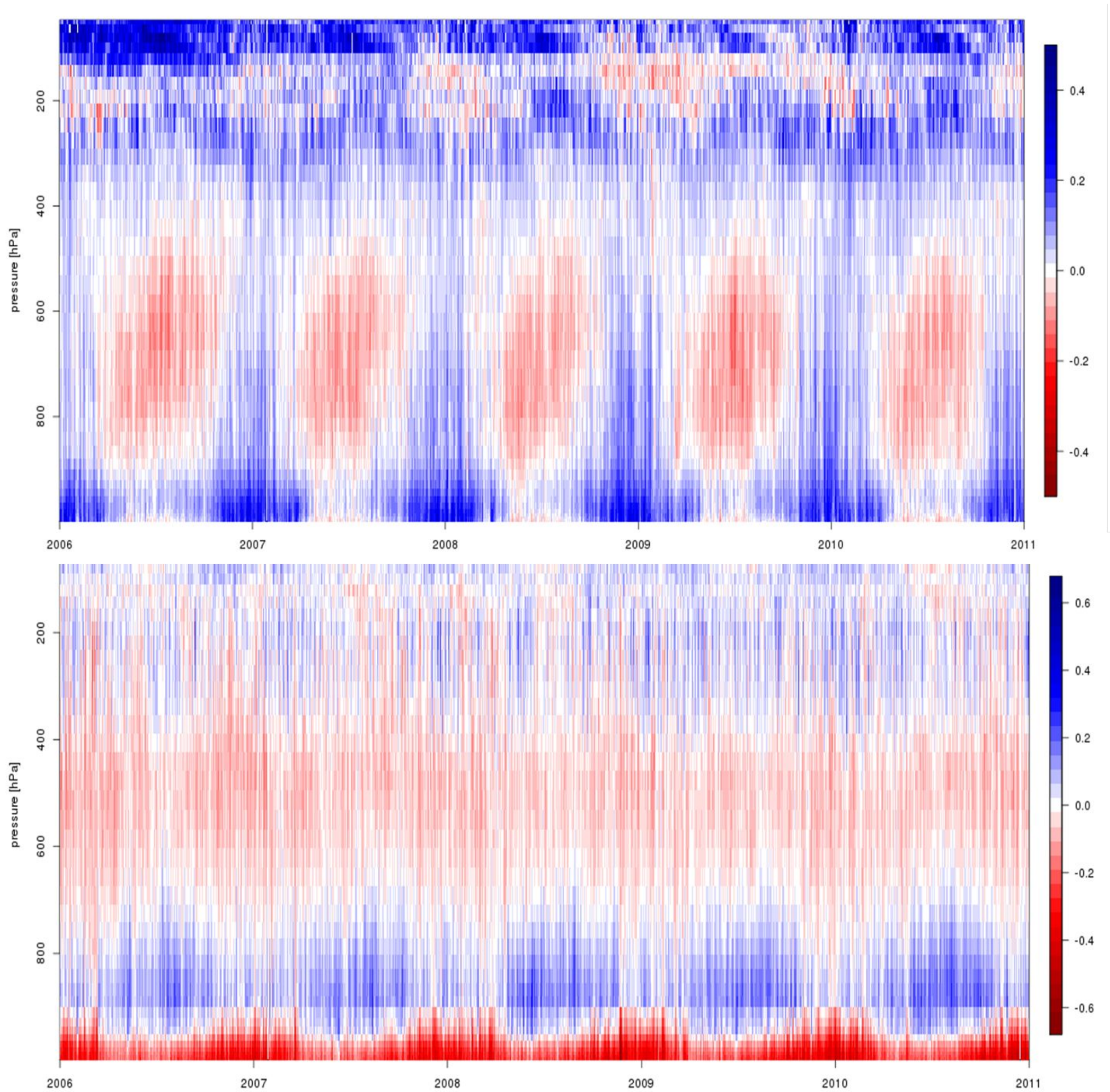

Fig. 5: Daily averages of 6-hourly aggregated area-averaged analysis increments for temperature [K] (top) and wind speed [m/s] (bottom) from January 2006 to December 2010. Red colours for temperature indicate too warm model temperatures and blue colours indicate too cold model temperatures. For wind speed blue colours mean that the model has too low wind speeds while red colour mean that model wind speeds are too high.

3.2 Analysis departures

Fig. 6, 7 and 8 show analysis departures statistics which serve as another tool to monitor and assess the quality of the COSMO-EN-REA12 data assimilation. Analysis departures are the differences between the analysis and the observations values and are hence a measure of how well the analysis fits to the observations. Mean (top) and RMS difference time series are plotted for the year 2007 for 2m temperature (Fig. 6), 2m relative humidity (Fig. 7) and 10m u-wind component (Fig. 8) together with the respective number of available observations (bottom).

For all three variables black lines show 6hourly mean and RMS differences for the ensemble. Blue and green lines show the respective daily means based on the 6 hourly values for the ensemble and the control run (Ctrl) respectively.

For the 2m temperature we find an overall small warm bias (-0.3 K) for both the mean difference of the ensemble and the control. Both lines are very close to each other and show a seasonal dependence with larger values in summer (around 0.5K) and smaller values in winter. As can be seen from the mean 6 hourly ensemble values there is also pronounced diurnal cycle with largest values found for the 00h cycle and smallest in the12h cycle. The RMS differences (middle), which are a measure of the error, are stable over time (2 K) for both the mean difference of the ensemble and the control, with the exception a shot-term peak at the end of December. This short-term increase in error is not related to a change in number of observations (Fig. 6, bottom) which show systematic intra-week fluctuations (with less observations on the weekends and more observation during the working days), but a further stable over time.

The same statistics are shown for 2m relative humidity in Fig. 7. The time series of the number of observations shows the same systematic fluctuations as for 2m temperature observations but is further also stable over time. Time series of control departures and mean departures of the ensemble show consistent behaviour and have small dry bias (0.02), with slightly larger values in summer.

Departure statistics for 10m u-component of wind are shown in Fig. 8. Ctrl and Ens consistently show a small mean differences (overall mean 0.03 ms^{-1}). RMS differences are found to be larger for the ensemble (2.25 ms^{-1}) than for the control (2.05 ms^{-1}). Both the control and the mean ensemble RMS differences show a seasonal dependency with smaller differences in summer and larger differences in winter when higher wind speeds occur

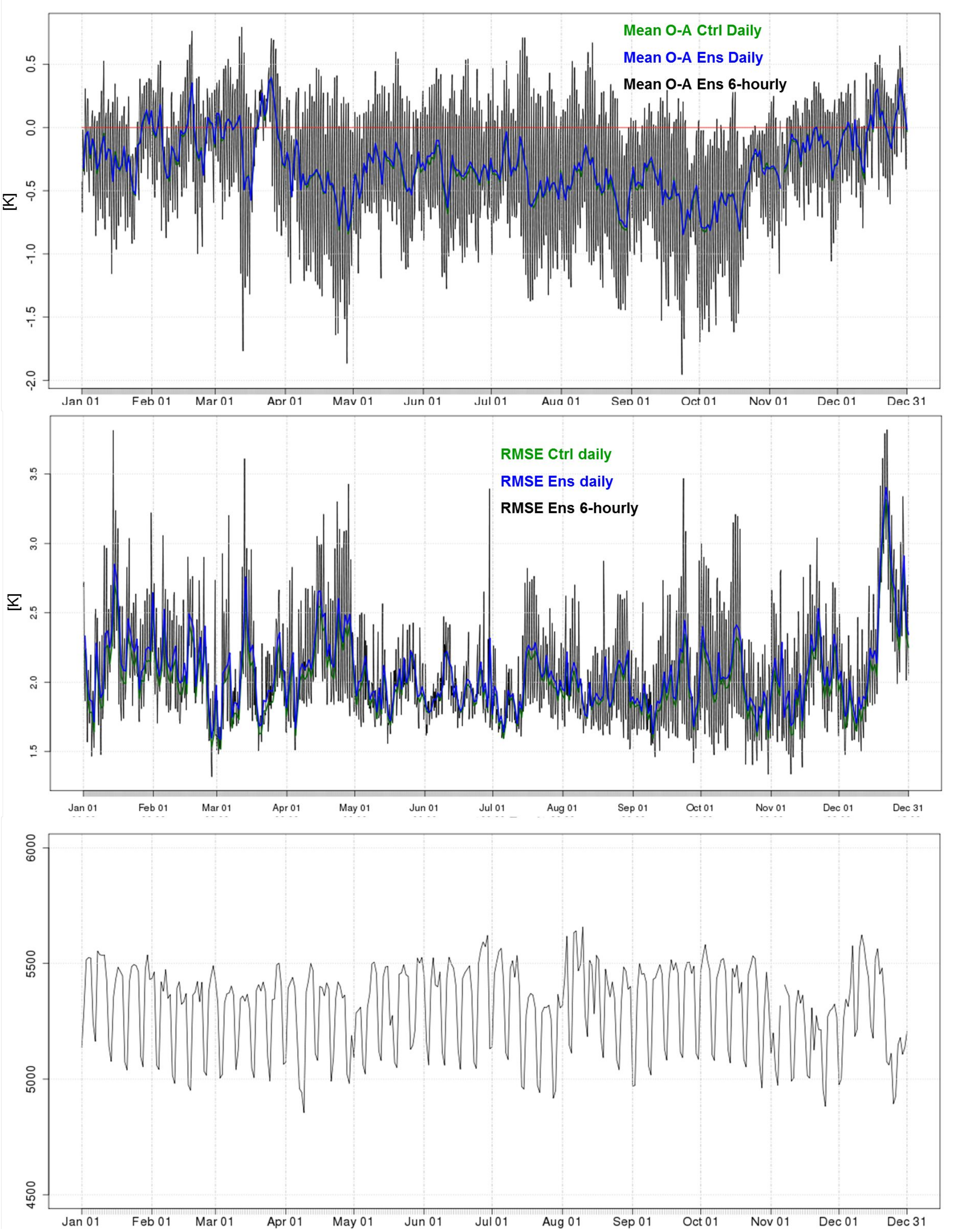

Fig. 6: Mean (top) and RMS (middle) differences between observations and analysis for SYNOP land observations of 2m temperature in 2007. Daily Means for the Control run (Ctrl) as well as daily and 6-hourly values for the ensemble are green, blue and black lines, respectively. The bottom plot shows the count of available observations.

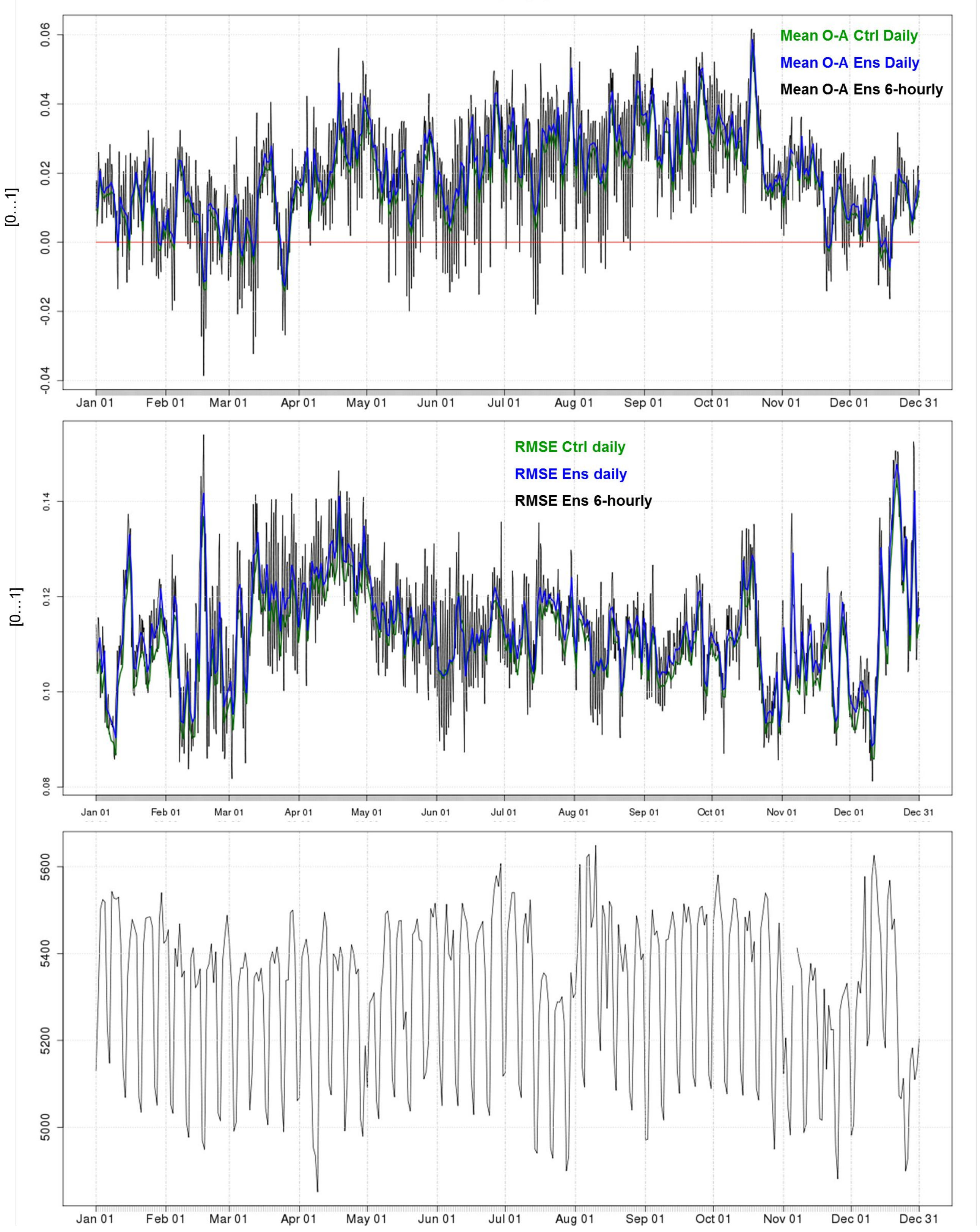

Fig. 7: Same as Fig. 6 but for SYNOP land observations of relative humidity [0…1] in 2007.

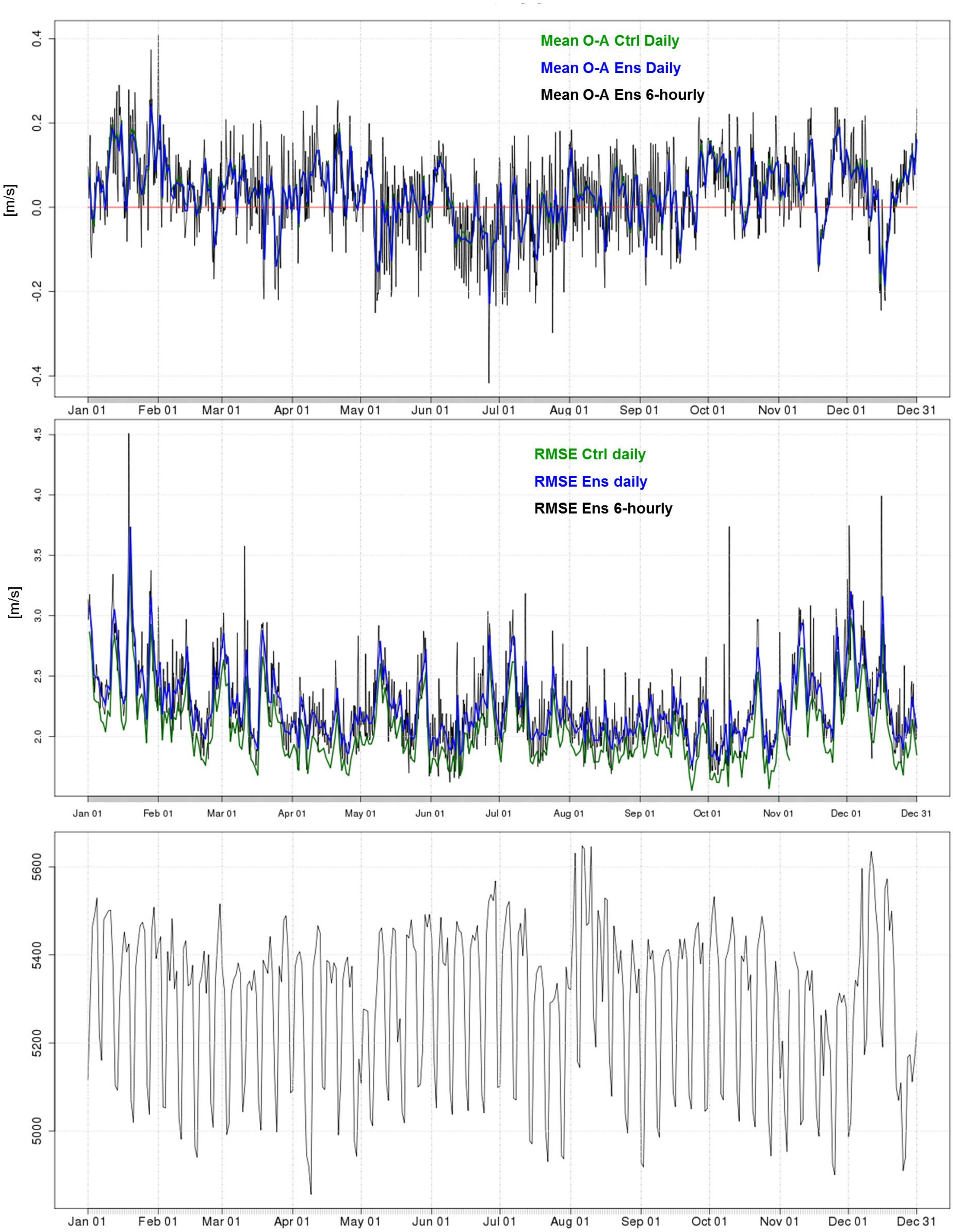

Fig. 8: Same as Fig. 6 but for SYNOP land observations of the 10m u-component of wind [m/s] in 2007

3. Verification statistics

Other studies have used independent observations to evaluate the quality of the COSMO-EN-REA12 especially with respect to the uncertainty estimates that are quantified by an ensemble First results from a pilot study by Bach et al. 2016 using test data from the COSMO-EN-REA12 suite were promising indicating good probabilistic capabilities with respect to precipitation. Data from the final COSMO-EN-REA12 production was recently used by Jermey et al. (2017) within an evaluation of the ensemble reanalysis datasets produced within UERRA. Using daily observations from the European Climate Assessment & Dataset (ECAD) as truth this evaluation put special emphasis on investigating the quality of the uncertainty estimation provided by the reanalyses. The evaluation considered three variables, namely daily means 2m temperature, 10m winds speed and total precipitation. Main findings for COSMO-EN-REA12 include

- good accuracy with respect to ensemble mean; additionally, both accuracy and bias were found to be an improvement compared to the global reanalysis CERA-20C
- improved ensemble accuracy and reliability compared to CERA-20C for all three variables
- improved uncertainty estimation compared to CERA-20C with best uncertainty estimation capabilities found for precipitation

More details on the evaluation can be found in Jermey et al. (2017).

4. Summary

The University of Bonn has completed the production and archiving of five years of the regional ensemble reanalysis system COSMO-EN-REA12 set up during the UERRA project. The COSMO-EN-REA12 reanalysis is produced at 12-km grid spacing with hourly resolution. It includes 21 members and covers the CORDEX-EUR11 domain. This report aimed at demonstrating the quality of the production system and included an investigation of the observation input stream and the data assimilation system using analysis increments and the analysis departure statistics.

Time series of the total number of observation counts revealed an overall stable input stream showing only a small increase over the 5 years period. An abrupt change in number of observations in the radiosonde input stream was found to not have an impact on the quality of the data assimilation. The analysis increments of two main prognostic variables, namely temperature and wind, showed only small changes indicating that the model is reasonably stable. Systematic deviations were found however with respect to the specification of the boundary layer and the tropopause height. Examination of the analysis departure statistics, as an additional tool to monitor and assess the quality of the data assimilation, revealed overall small mean differences for the three variables considered (2m temperature, 2m relative humidity and 10m u-component of wind) and consistency between the control and mean ensemble statistics. RMS differences of the control and of the ensemble were found to be of similar size except for wind where the control had lower errors than the ensemble.

Recent results by Jermey et al. (2017) evaluating the quality of the COSMO-EN-REA12 uncertainty estimation confirm earlier findings of a pilot study by Bach et al. (20016) who found good probabilistic capabilities for COSMO-EN-REA12. Jermey et al. (2017) also found improvements of the COSMO-EN-REA12 (and the regional reanalyses in general) compared a global reanalysis in representing both ensemble uncertainty and accuracy

References

Bach, L., Schraff, C., et. al, Towards a probabilistic regional reanalysis system. Part 1: Evaluation of precipitation from experiments, submitted to Tellus A, 2016.

Bollmeyer, C., Keller, J. D., Ohlwein, C., Wahl, S., Crewell, S. and co-authors. 2015. Towards a high-resolution regional reanalysis for the european CORDEX domain. Q. J. R. Meteorol. Soc. 141(686), 1_15.

Giorgi, F., Jones, C. and Asrar, G. R. 2009. Addressing climate information needs at the regional level: the CORDEX framework. WMO Bull. 58, 175_183.

Jermey, P. (2017): Reanalysis uncertainty evaluation. UERRA deliverable D2.14

Schättler, U., Schraff, C., et al. (2011), A description of the nonhydrostatic regional COSMO-model, Technical report, Deutscher Wetterdienst, Offenbach.

Schraff, C. H. (1997), Mesoscale data assimilation and prediction of low stratus in the Alpine region, Meteorol. Atmos. Phys. 64(1-2), 21–50.