Neighborhood verification of convection in the Swiss COSMO models with radar and satellite measurements

Thomas Leutert, Daniel Leuenberger and Francis Schubiger

MeteoSwiss, Operation Center 1, CH-8058 Zurich Airport

1 Introduction

Increasing the resolution of numerical weather prediction (NWP) models is a promising approach to improve weather forecasts on local and regional scales: High-resolution models better represent topography, boundary layer and land surface processes, and avoid uncertainties due to convection parameterizations (e.g., Ban et al., 2014). Several studies confirm a benefit in simulating precipitation for high-resolution models at convection-resolving scales in comparison to coarser-resolution models (e.g., Langhans et al., 2013; Prein et al., 2013; Kendon et al., 2012; Langhans et al., 2012; Knote et al., 2010; Hohenegger et al., 2008). However, conventional skill scores often do not capture the higher realism from increased resolution due to small spatial or temporal shifts of the forecast (double penalty effect). Therefore, a neighborhood verification approach which does not only take into account individual grid points, but also rewards closeness in space, time and other relevant aspects (e.g., Ebert, 2008) has to be used for the evaluation of high-resolution forecasts.

Assessing forecasts of three configurations of the non-hydrostatic limited-area atmospheric prediction model from the Consortium for Small-Scale Modelling (COSMO) with grid sizes of 1.1 km (COSMO-1), 2.2 km (COSMO-2) and 6.6 km (COSMO-7), we evaluate the role of model resolution for the quality of convection simulations for summer 2014 in Switzerland. The main objectives of this study are as follows:

- Compare the diurnal cycle of convective precipitation in COSMO-1, COSMO-2 and COSMO-7 against gridded precipitation estimates from combined radar and rain gauge observations. Is there a significant difference between convection-permitting and convec-tion-parametrizing models?
- Extend the neighborhood verification framework from precipitation data to brightness temperatures measured by satellites. What are reasonable brightness temperature threshold values to verify convective clouds in NWP models?
- Assess the NWP models of MeteoSwiss against radar and satellite observations. How well does COSMO-1 simulate clouds in comparison to COSMO-2 and COSMO-7? On which spatial scales does model performance benefit from increased resolution?

2 Verification approach

A better representation of reality by high-resolution models in comparison to models with a lower resolution does not necessarily imply greater accuracy. In case of high-resolution forecasts, traditional verification methods tend to overemphasize errors on small spatial scales, leading to an unfair double penalty effect (Roberts and Lean, 2008). Taking into account more than one grid point helps to reduce this double penalty effect. Neighborhood verification assesses forecast skill scores for different spatial windows and thresholds, allowing for the identification of the scales and thresholds where model quality reaches highest values.

Ebert (2008) summarizes the main neighborhood verification methods. They represent different decision models to assess the usefulness of a forecast. The Fractions Skill Score (FSS) is well suited for the verification of high-resolution forecasts (Eckert, 2009) and evaluates the simulated fraction exceeding/falling below a certain threshold. It is calculated via Fractions Brier Score (FBS) from the fractions of observed (P_{obs}) and forecast (P_{fcst}) grid points which exceed/fall below a certain threshold (e.g., Weusthoff et al., 2010):

$$FBS = \frac{1}{N} \sum_{N} (P_{fcst} - P_{obs})^2 \tag{1}$$

$$FSS = 1 - \frac{FBS}{\frac{1}{N}\left(\sum_N P_{fcst}^2 + \sum_N P_{obs}^2\right)} \tag{2}$$

For precipitation data, threshold exceedances are assessed, whereas we use threshold undercuts for brightness temperature. The FSS asymptotically approaches a value that depends on the frequency bias (1 if no bias). According to Roberts and Lean (2008), useful spatial scales of a forecast are characterized by FSS > FSS_{useful} ($\mathrm{FSS}_{useful} = 0.5 + \mathrm{f}_{obs}/2$ where f_{obs} is the observed fractional coverage for a given threshold over the domain). A simulation with a useful FSS provides additional benefit in comparison to the forecast of a constant ratio of events to non-events. The skill scores were calculated with the neighborhood verification package from Ebert (2008) for a domain covering Switzerland.

In order to assess precipitation, the thresholds 0.1 mm/h, 0.2 mm/h, 0.5 mm/h, 1 mm/h, 2 mm/h, 5 mm/h, 10 mm/h and 20 mm/h have been used. We chose the following brightness temperature thresholds: 210 K, 220 K, 230 K, 240 K, 250 K and 260 K. This brightness temperature range prevents an influence by surface temperatures in Swiss summers and at the same time covers the relevant spectrum of tropospheric cloud temperatures. All skill scores were calculated for square neighborhoods with different sizes (COSMO-1: 1 × 1, 2 × 2, 6 × 6, 18 × 18, 30 × 30, 54 × 54 and 90 × 90 grid points; COSMO-2: 1 × 1, 3 × 3, 9 × 9, 15 × 15, 27 × 27 and 45 × 45 grid points; COSMO-7: 1 × 1, 3 × 3, 5 × 5, 9 × 9 and 15 × 15 grid points).

3 Observational data

CombiPrecip is a suitable product to verify quantitative precipitation forecasts. It is produced with a mesh size of 1 km and combines information from the Swiss radar composite based on 4 Doppler radars and ∼180 automatic rain gauges measuring each 10 minutes. The large observation errors of the spatially dense radar precipitation estimates are reduced by rain gauge point measurements by means of co-kriging with external drift (Sideris et al., 2014). Hourly accumulations from CombiPrecip are used to assess convective precipitation in the complex Alpine topography. The CombiPrecip data have been interpolated to the grids of COSMO-1, COSMO-2 and COSMO-7 by averaging over all grid points lying inside a model grid box.

In order to estimate the skill of forecasting convective clouds, observational satellite imagery from the Spinning Enhanced Visible and Infrared Imager (SEVIRI) located onboard Meteosat Second Generation (MSG) (Schmetz et al., 2002) is compared to simulated satellite images from the COSMO models using the NWP-SAF RTTOV forward operator (Keil et al., 2006; Keil and Tafferner, 2003). The focus lies on brightness temperatures computed from Channel 9 (10.8 μm). They are closely correlated with the target temperatures and thus can be used to distinguish between (cold) clouds and the warmer Earth's surface (e.g., Böhme et al., 2011; Schmetz et al., 2002). For the verification, we assume an inverse relationship between brightness temperature and the top height of convective clouds: High, convective clouds are cold, low clouds are warmer. The satellite data set with a resolution of 3.2 km (west-east) × 5.4 km (north-south) over Switzerland (Schmetz et al., 2002) has been interpolated to the COSMO model grid by means of bilinear interpolation. Spatial scales below ∼5 km are ignored in the interpretation of the brightness temperature verification results.

4 Model data

We validated and verified three Swiss COSMO configurations: COSMO-1 (1.1 km), COSMO-2 (2.2 km) and COSMO-7 (6.6 km). The configurations differ in horizontal resolution and in physical parameterizations: COSMO-7 uses a mass-flux convection scheme with equilibrium closure based on moisture convergence for shallow to deep convective clouds (Tiedtke, 1989). COSMO-1 and COSMO-2 are assumed to explicitly resolve deep convection, but rely on a parameterization for shallow convection. COSMO-7 is driven by the Integrated Forecasting System (IFS) of the European Center for Medium-Range Weather Forecasts (ECMWF). COSMO-1 and COSMO-2 are nested into COSMO-7. The model domain for the assessment of convection has been chosen to fully cover the domain of Switzerland. COSMO-2 and COSMO-7 are run in operational mode, while COSMO-1 still runs in experimental mode. Here, all 00 UTC model runs of the verification period are considered up to a lead time of 24 h.

5 Verification of the 12th June 2014 case

First, a typical convection day is analyzed. A flat pressure distribution over central Europe, warm temperatures and moist air led to favourable conditions for deep convection over Switzerland on 12th June 2014. It was a summer day characterized by a strong diurnal cycle of convection.

At the beginning of the day (01 UTC), the satellite (Fig. 1) and radar (Fig. 2) observations (interpolated to a 1.1 km grid) show cloud leftovers from the previous day (Fig. 1), but no precipitation (Fig. 2). The sky has cleared up at 10 UTC. At 11 UTC, convection is initiated over the Alps and reaches its maximum intensity around 19 UTC. Afterwards, convective activity decreases again (not shown).

The models overestimate cloudiness at 01 UTC and 10 UTC (Fig. 1). At 01 UTC, COSMO-1 and COSMO-2 predict too much clouds over the western part of Switzerland and in the southeastern corner of the domain. The overprediction is less pronounced in COSMO-7. 9 hours later, low brightness temperatures have almost completely disappeared in the observations. The models show a patchy pattern of cold temperatures over Eastern Switzerland, indicating a too early onset of convection. The granular structure in case of COSMO-7 is likely caused by the parameterization scheme for deep convection. At 19 UTC, observed brightness temperatures over Switzerland have decreased significantly. Although not perfect, the overall convective structure is well captured by COSMO-1 and COSMO-2. COSMO-7 misses the centre of convection over the northern part of Switzerland.

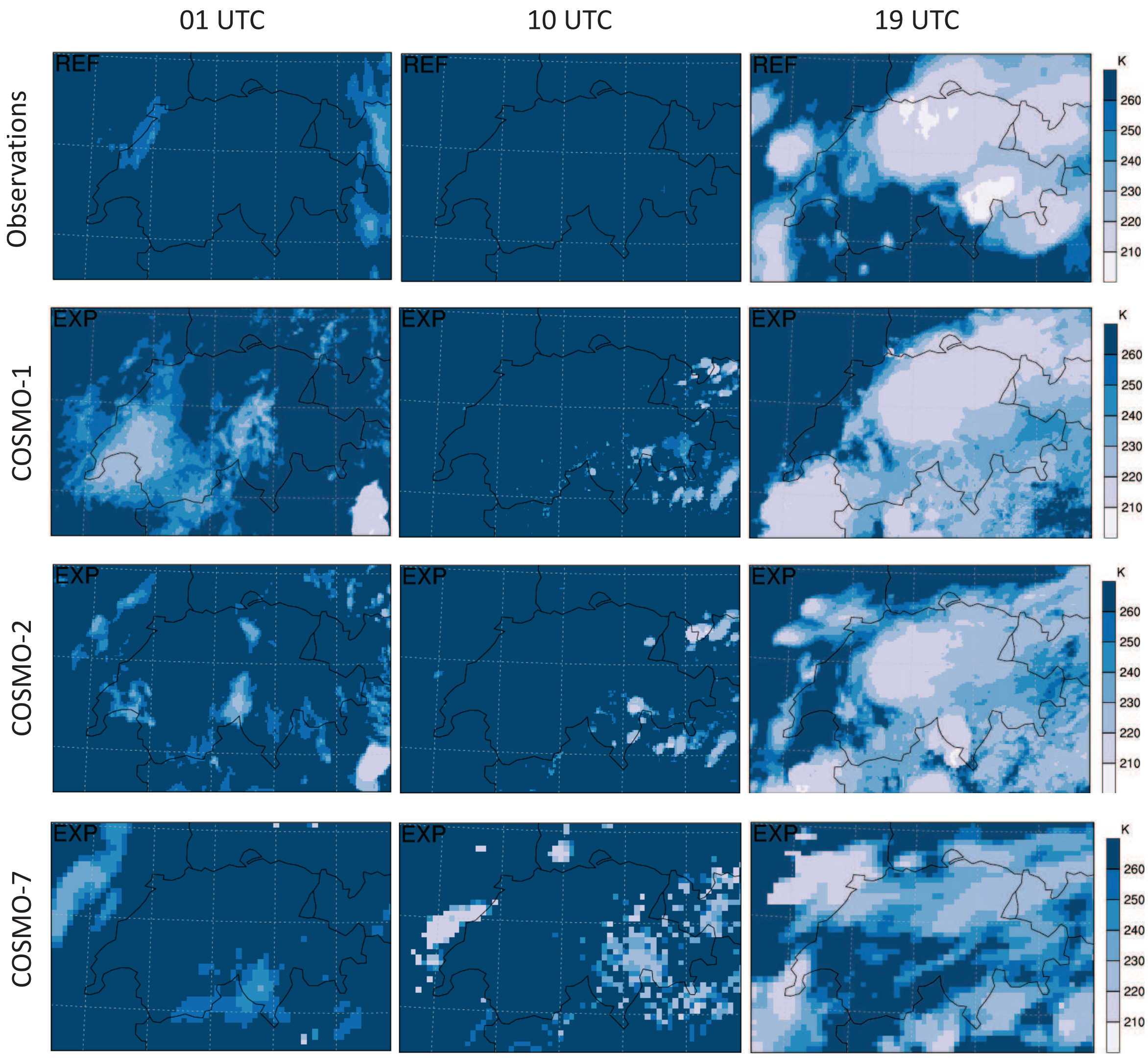

Figure 1: Brightness temperature during 12th June 2014 over Switzerland. In addition to the satellite data (interpolated to a grid with 1.1 km mesh size; upper row), the model output of COSMO-1 (middle row), COSMO-2 and COSMO-7 (lower row) is shown at 01 UTC (left column), 10 UTC (middle column) and 19 UTC (right column). Light colors indicate low brightness temperatures linked to high-altitude clouds.

Apart from the region in the northeast of Lake Constance, simulated precipitation in COSMO-1 and COSMO-2 is negligible at 01 UTC (Fig. 2). The observations and COSMO-7 show no rain at all. At 10 UTC, COSMO-1 and COSMO-2 show slight precipitation in the eastern part of the domain.

COSMO-7 overforecasts precipitation in the same zone. This overestimation of precipitation seems to spatially coincide with the granular pattern of brightness temperature (Fig. 1). As expected from the simulated brightness temperatures, convective precipitation over the Alps increases towards 19 UTC in COSMO-1 and COSMO-2. In contrast to COSMO-1 and COSMO-2, COSMO-7 heavily underforecasts precipitation over the whole domain at 19 UTC.

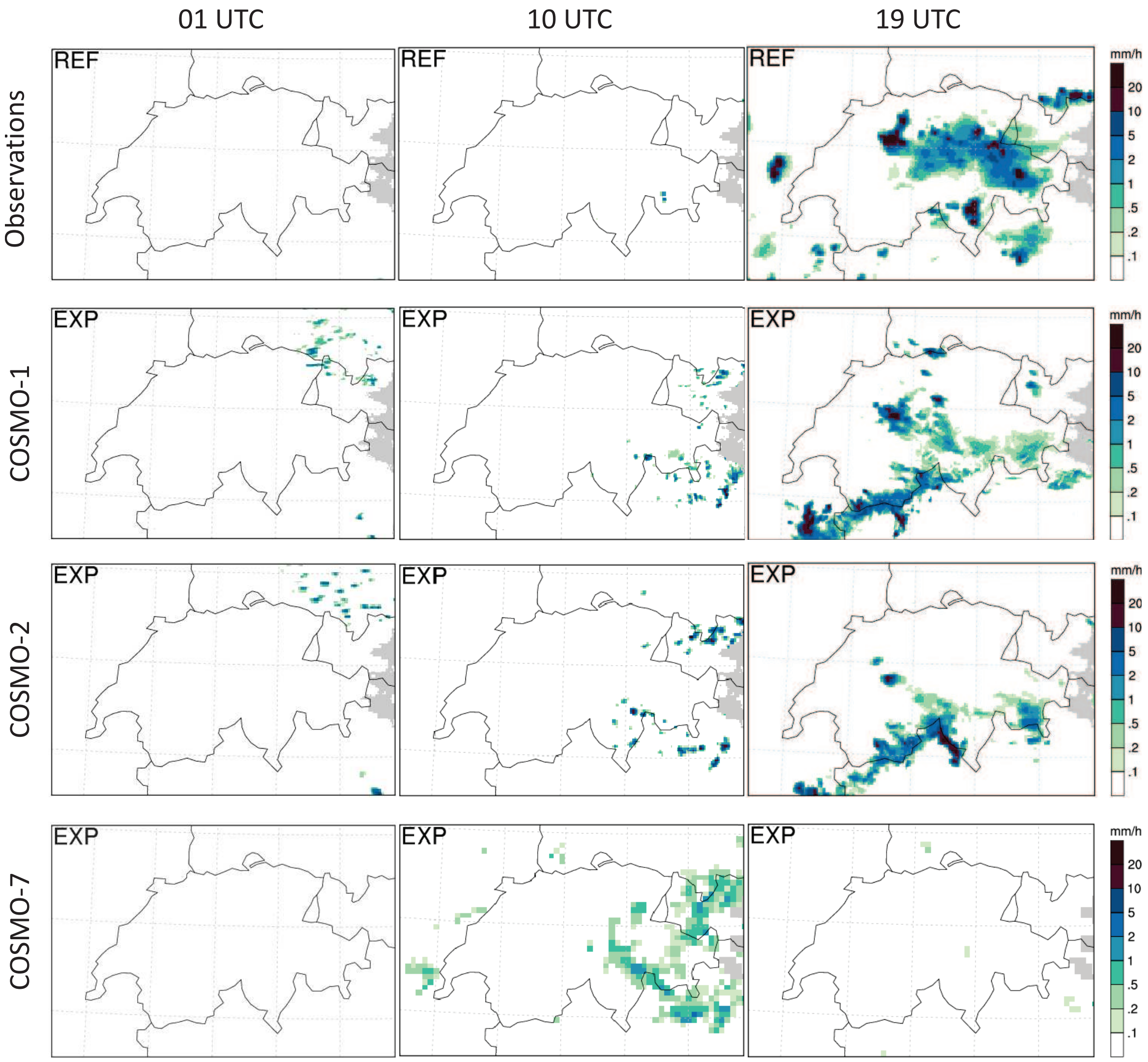

Figure 2: Precipitation during 12th June 2014 over Switzerland. In addition to the radar data (interpolated to a grid with 1.1 km mesh size; upper row), the model output of COSMO-1 (middle row), COSMO-2 and COSMO-7 (lower row) is shown for hourly sums ending at 01 UTC (left column), 10 UTC (middle column) and 19 UTC (right column). Dark colors indicate high precipitation amounts.

In all following neighborhood verification plots (Figs. 3, 4 and 7), higher skill score values indicate better performance. High skill is visualized by warm colors. The useful scales are marked by bold numbers.

Fig. 3 shows neighbourhood verification results for brightness temperature. The FSS quantifies the skill of forecasting a value below a certain threshold. Generally, the FSS tends to increase with increasing spatial scale and with increasing threshold. The best skill is obtained for high thresholds on large scales. This makes sense: (1) The requirements for a good match between model and observation are relaxed more heavily on large than on small spatial scales (e.g., Ebert, 2008). (2) A mixture of many different brightness temperature values falls below high thresholds, but only a small number of specific extreme values falls below low thresholds. Therefore, the skill at low thresholds (e.g., high clouds) is reduced in comparison to high thresholds (e.g., low clouds, no clouds).

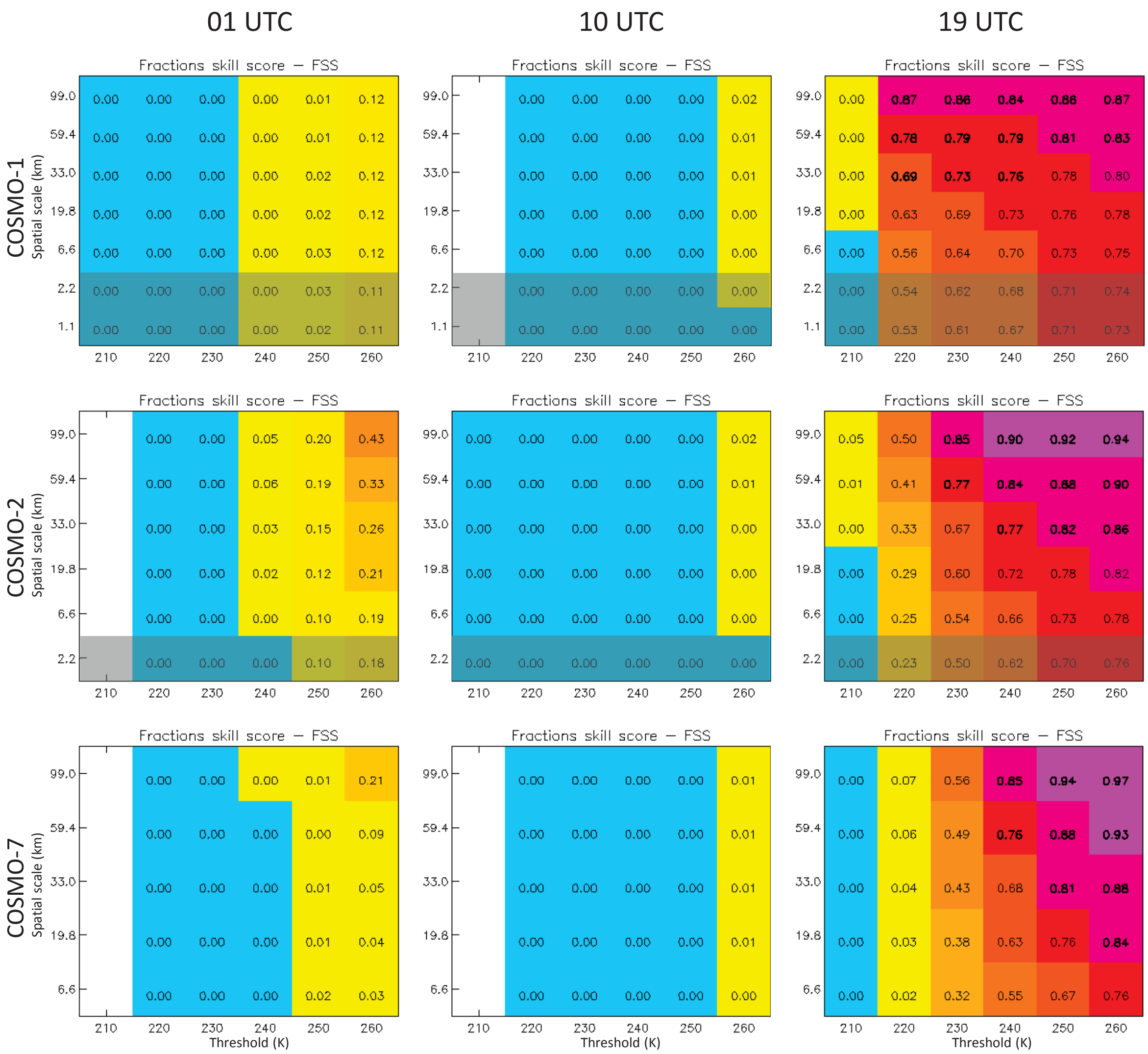

Figure 3: FSS of brightness temperatures from COSMO-1 (upper row), COSMO-2 (middle row) and COSMO-7 (lower row) at 01 UTC (left column), 10 UTC (middle column) and 19 UTC (right column) for 12th June 2014. Values shaded in gray should be disregarded since the spatial resolution of the satellite observations is about 5 km.

At 01 UTC, all models show low values of FSS in terms of brightness temperature forecasts without any useful scales at all (Fig. 3). COSMO-2 gives the best scores. 9 hours later, the FSS values are zero for most scales and thresholds, since almost no clouds are observed. In the evening (19 UTC), COSMO-1 and COSMO-2 simulate the overall brightness temperature structure better than COSMO-7, especially for low thresholds (high clouds). COSMO-7 can best discriminate between clouds/no clouds (threshold 260 K). COSMO-1 exhibits most useful scales.

In contrast to brightness temperature, threshold exceedances and not threshold undercuts are evaluated for the precipitation forecasts (Fig. 4). The neighborhood verification table of COSMO-7 is empty at 01 UTC. At this time, COSMO-7 and CombiPrecip do not show any precipitation for all spatial scales (Fig. 2). A FSS cannot be computed when the fraction of grid points exceeding a threshold is 0 in both observation and forecast (Eq. 2). Consequently, COSMO-7 correctly forecasts no rain at 01 UTC. In the morning (10 UTC) and evening (19 UTC), the precipitation forecast performance of COSMO-7 is very low. The high brightness temperature skill scores of COSMO-7 at 19 UTC are not at all reflected in the precipitation forecasts of COSMO-7. This model completely fails to predict the observed rainfall over Switzerland at 19 UTC.

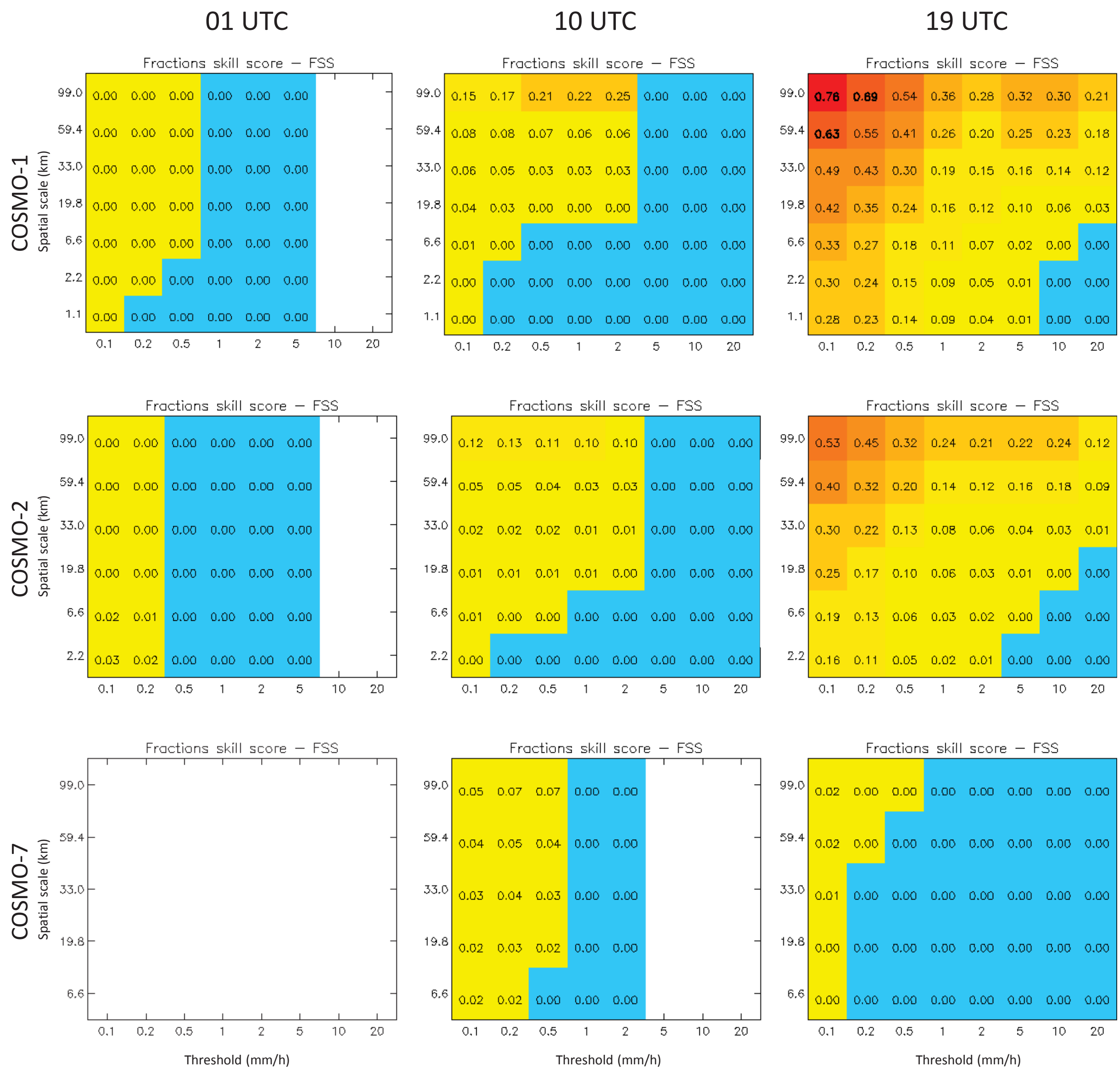

Figure 4: FSS of hourly precipitation of 12th June 2014 from 00 UTC to 01 UTC (left column), 09 UTC to 10 UTC (middle column) and 18 UTC to 19 UTC (right column) for the forecasts of COSMO-1 (upper row), COSMO-2 (middle row) and COSMO-7 (lower row).

The FSS of precipitation from COSMO-1 and COSMO-2 are low at 01 UTC and 10 UTC and rise thereafter (Fig. 4). Only COSMO-1 produces useful precipitation forecasts at spatial scales of 59.4 km and 99.0 km. It outperforms the other models for most scales and thresholds.

To sum up, the analysis of the forecasts of 12th June 2014 reveals that both COSMO-1 and COSMO-2 produce results which are closer to the observed values than those of COSMO-7. Neighborhood verification generally confirms the better skill of the convection-permitting models, as seen from visual inspection of the spatial brightness temperature and precipitation distributions. In the next chapter, the evaluation of brightness temperature and precipitation is extended to the period from June to August 2014.

6 Verification of June, July and August 2014

The end of May and the beginning of June were characterized by several cold front passages and moist conditions. After some cool days in the beginning of June, a heatwave affected Switzerland for approximately one week. Following the warm temperatures, heavy thunderstorms brought precipitation to Switzerland. Towards the end of June, a cold front and inflowing unstable air masses further intensified thunderstorm and rainfall activity. Generally, June was very warm and characterized by a high potential for convection. In contrast to June, July was cool and wet. Apart from a warm and stable weather phase (15th July to 19th July), advancing depressions and thunderstorms led to record precipitation values. Flooding and landslides caused widespread damage in Switzerland. The cool conditions over Switzerland persisted in August. Especially during the first half of the month, numerous thunderstorms associated with atmospheric disturbances led to heavy rainfall events (MeteoSwiss, 2014). Prevalent wet and convective conditions during June, July and August 2014 make this period suitable for the verification of precipitation and cloud cover forecasts.

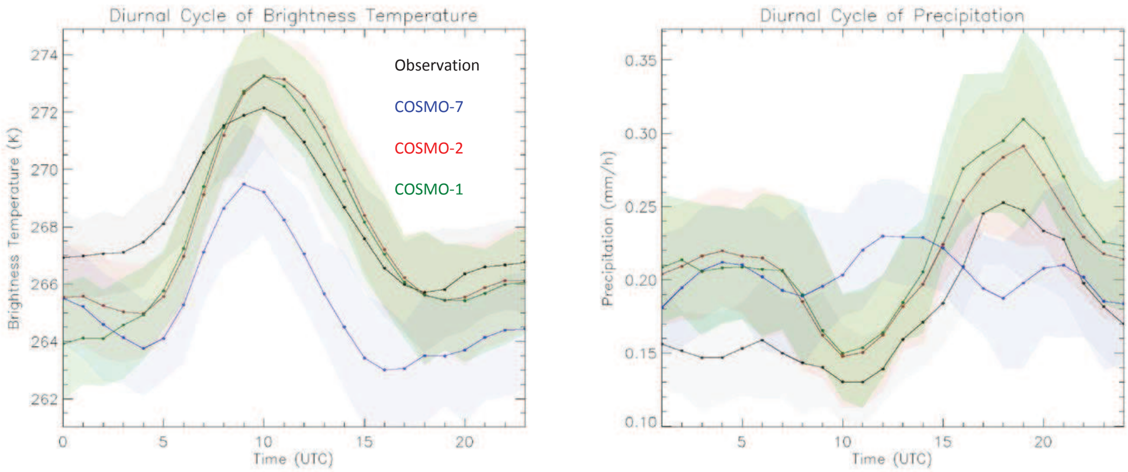

Figure 5: Mean diurnal cycles of brightness temperature (left) and precipitation (right) averaged over June, July and August 2014. Black curves denote observations, blue curves COSMO-7, red curves COSMO-2 and green curves COSMO-1 forecasts. The uncertainty estimates (shaded areas) are based on bootstrapping and represent the range between the 10th and 90th percentile.

Fig. 5 shows the mean diurnal cycles of brightness temperature and precipitation. The shaded areas denote the uncertainty represented by the range between the 10 % and the 90 % quantile of a distribution computed using bootstrapping. To this end, 100 equally sized random samples of the days were drawn from the original data of the three-month period (e.g., Boos, 2003).

In contrast to the precipitation values accumulated over 1 hour, the brightness temperature values are instantaneous. All diurnal cycles of average precipitation and brightness temperature have been computed in intervals of 1 hour from 01 UTC to 24 UTC for precipitation and from 00 UTC to 23 UTC for brightness temperature.

The observed brightness temperature values are characterized by a maximum (minimum cloud cover) around 10 UTC and a minimum (maximum high cloud cover) around 18 UTC (Fig. 5, left). COSMO-1 and COSMO-2 capture the amplitude and the timing of the observed diurnal cycle well and clearly exceed the performance of COSMO-7. Despite their good overall performance, COSMO-1 and COSMO-2 underestimate nighttime and early morning average brightness temperature. In the first half of the day, the observations indicate low rainfall (Fig. 5, right). Then, rainfall starts to increase and peaks around 18 UTC. After reaching its maximum value, precipitation decays much faster than the clouds. This is physically realistic, as clouds which have rained out tend to persist without further precipitation. COSMO-7 agrees substantially worse with the observations than COSMO-1 and COSMO-2.

The poor representation of the diurnal cycle of precipitation in COSMO-7 (too early maximum) is probably linked to the convection scheme (e.g., Dai et al., 1999). This parameterization scheme is not designed to represent individual convective showers (Kendon et al., 2012). The convection-permitting models, i.e. without parameterized deep convection, capture the diurnal cycle of precipitation much better.

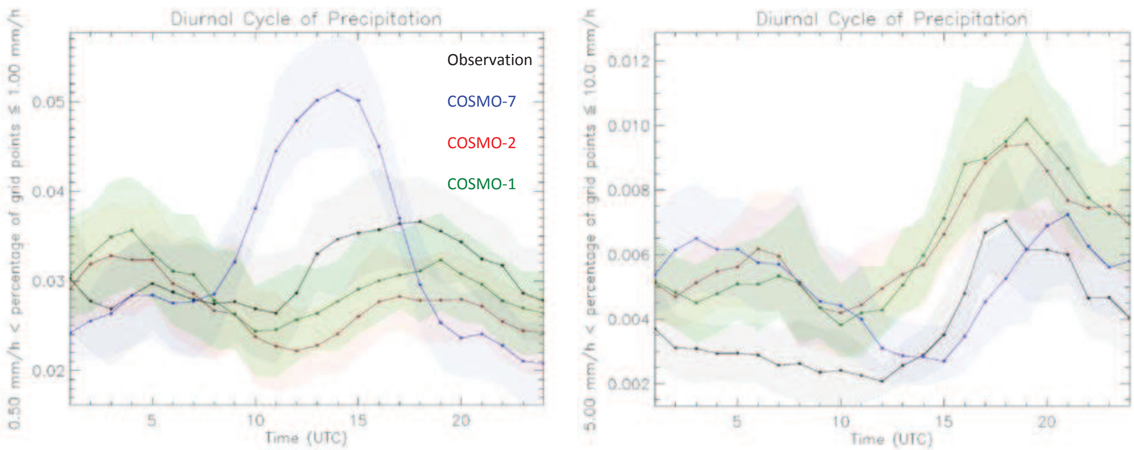

Figure 6: Diurnal cycles (June, July and August 2014) of the relative amount of grid cells with precipitation higher than 0.5 mm/h but lower/equal 1.0 mm/h (left) and precipitation higher than 5.0 mm/h but lower/equal 10.0 mm/h (right). Black curves denote observations, blue curves COSMO-7, red curves COSMO-2 and green curves COSMO-1 forecasts. The uncertainty estimates (based on bootstrapping; shaded areas) represent the range between the 10th and 90th percentile.

The too early peak of average precipitation in COSMO-7 (Fig. 5, right) is predominantly caused by light precipitation events (e.g., values between 0.5 mm/h and 1.0 mm/h; Fig. 6, left). In contrast, the maximum of heavy precipitation (e.g., values between 5.0 mm/h and 10.0 mm/h; Fig. 6, right) in COSMO-7 even occurs later than observed. Generally, the timing of the diurnal cycle of precipitation is well captured by the convection-permitting models (COSMO-1 and COSMO-2).

Fig. 7 shows neighbourhood results aggregated over June, July and August 2014. The best brightness temperature forecasts are produced by COSMO-1 and COSMO-2 (left panel). Their skill scores are similar for most spatial scales and thresholds. At spatial scales $<$ 33 km, COSMO-1 performs marginally better than COSMO-2. COSMO-7 is beaten on all scales and for all thresholds. Useful spatial scales are observed for high brightness temperature forecasts on large scales during June, July and August 2014.

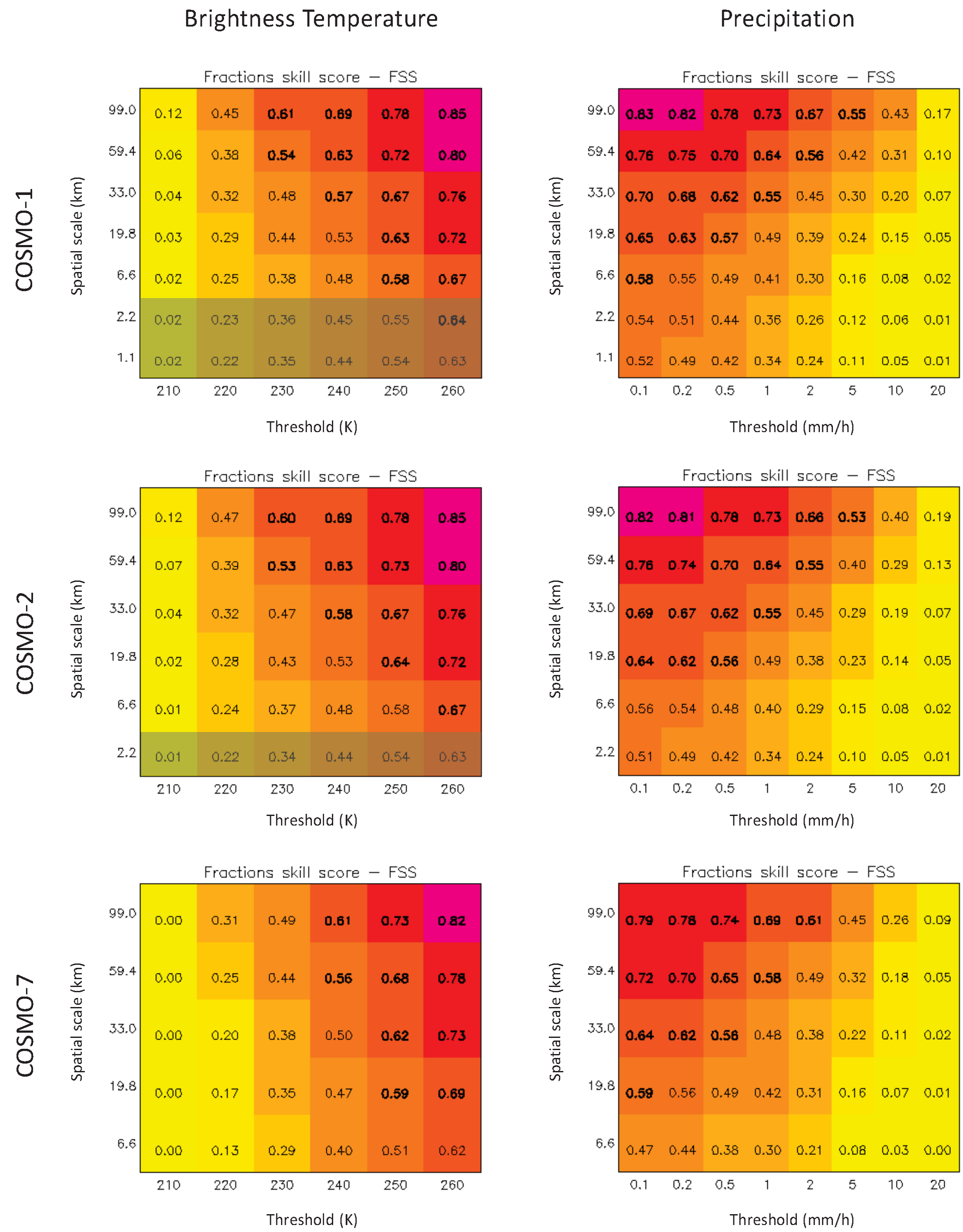

Figure 7: FSS of brightness temperature (left) and precipitation (right) as simulated by COSMO-1 (upper row), COSMO-2 (middle row) and COSMO-7 (lower row) for all 00 UTC runs from 1st June to 31st August 2014 for all hourly forecasts (hourly precipitation sums). The brightness temperature values in gray should be disregarded since the spatial resolution of the satellite observations is about 5 km.

Similar to brightness temperature, we find useful scales for all models in case of precipitation forecasts (Fig. 7, right panel). COSMO-1 outperforms COSMO-2 and COSMO-7. The simulations of COSMO-7 are clearly worst for all scales and thresholds.

We address the question whether the models forecast the correct number of grid point values within a certain brightness temperature range by means of a frequency-intensity distribution (Fig. 8). The frequency-intensity distribution is discretized into bins, using the threshold values from the neighborhood verification. The number of grid points in a bin is always computed relative to the total number of grid points in the domain in order to compare models that differ in resolution. Values above 260 K are not shown in the frequency-intensity distribution. Therefore, the values of a curve do not add up to 1. The observation values are those interpolated to the COSMO-7 grid with a resolution of 6.6 km. An overestimation of low brightness temperature values (< 250 K) occurs in all models, especially in COSMO-7, indicating too much high cloudiness. It may be caused by the parametrization of ice nucleation and/or the neglected sedimentation of cloud ice (Pfeifer et al., 2010). This positive bias of high clouds is smaller for the high-resolution models COSMO-1 and COSMO-2, confirming a benefit of the explicit resolution of convection.

Additionally, the Upscaling (UP) method was calculated using the Equitable Skill Score (ETS) (e.g., Ebert, 2008). The UP/ETS results qualitatively agree with the FSS in most cases (not shown).

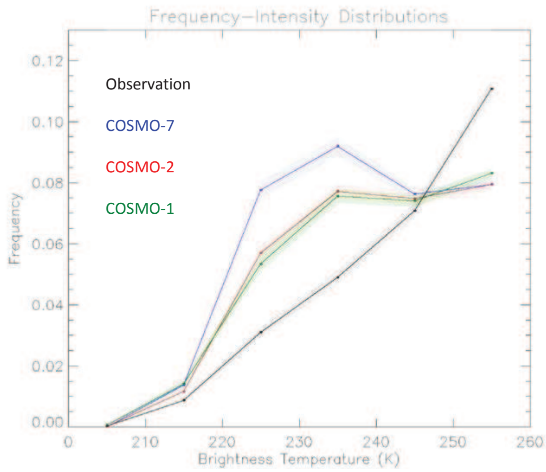

Figure 8: Frequency-intensity distributions of brightness temperature for the observations (black curve), COSMO-1 (green curve), COSMO-2 (red curve) and COSMO-7 (blue curve) during June, July and August 2014. The shaded areas represent the 80 % confidence intervals based on bootstrapping.

6 Conclusions

Precipitation and brightness temperature forecasts of COSMO-1, COSMO-2 and COSMO-7 have been assessed for a typical day with strong convective activity (12th June 2014) and for the summer 2014 (June, July and August) by means of the neighborhood verification technique for the domain of Switzerland. Thresholds of 0.1 mm/h, 0.2 mm/h, 0.5 mm/h, 1 mm/h, 2 mm/h, 5 mm/h, 10 mm/h and 20 mm/h have been used for precipitation and 210 K, 220 K, 230 K, 240 K, 250 K and 260 K for brightness temperature.

To provide a holistic view of model performance, our evaluation focused on the Fractions Skill Score (FSS) method of the neighborhood verification package, but also includes spatial distributions, frequency-intensity distributions and average diurnal cycles. Brightness temperature serves as a proxy for convective cloud activity which was intense during this period, especially in June 2014, and posed a great challenge for all NWP models.

The spatial distributions of precipitation and brightness temperature for a single convective day reveal that the high-resolution models COSMO-1 and COSMO-2 exhibit more realistic patterns than COSMO-7. This is also reflected in the results of the FSS. At 19 UTC (maximum in convective activity), the FSS indicates highest skill in COSMO-1 both for precipitation (for nearly all scales and thresholds) and for brightness temperature (for most scales and thresholds below 240 K, e.g. for deep convective clouds).

For June, July and August 2014 and averaged over all forecast hours up to +24 h, the results show the best scores for COSMO-1. No clear scale dependency of the improvements of COSMO-1 in comparison to COSMO-2 is observed. The mean daily cycles of the convection-permitting COSMO-1 and COSMO-2 models are much better in timing and amplitude than COSMO-7. However, they still overestimate nighttime and morning clouds and precipitation. The too early maximum of precipitation in COSMO-7 is caused by light precipitation events. The frequency-intensity distributions reveal that all three models tend to overestimate the number of brightness temperature values below 250 K.

The results of this study suggest that brightness temperatures of the 10.8 μm channel observed by satellites and simulated by the COSMO model can well serve as a proxy for convective clouds and have the potential to complement precipitation for the spatial verification of convective processes. It remains to be explored if brightness temperature as a verification proxy is also suited for other types of clouds like frontal or stratus clouds. All in all, our verification results confirm additional benefit for the use of convection-permitting models in comparison to coarser models with parameterized convection, further increasing confidence in the high-resolution modelling approach. A decrease in horizontal grid size from 7 km to 2 km seems to improve forecast performance more than for convection-permitting models from 2 km to 1 km.

7 Outlook

The subsequent steps could reduce the uncertainty in the assessment of the COSMO models to better understand the role of model resolution in terms of forecast skill.

- An additional temporal dimension in the neighborhood verification will allow for a relaxation of the requirements for exact matches in time (especially for hourly precipitation sums).
- The impact of using data from more than only the 00 UTC runs could be quantified for each model. Furthermore, the relationship between model skill and lead time may be assessed using all model runs.
- It would be interesting to evaluate brightness temperature and precipitation forecast performance for specific meteorological conditions and forcing factors (weather type dependent verification).
- With regard to the relation between brightness temperature and cloud top height, it has to be kept in mind that brightness temperature at 10.8 μm as recorded by satellites does not provide complete information about cloud cover in the atmosphere: Medium brightness temperatures may stem from mid-level clouds, but can also be caused by semi-transparent high-level clouds. Therefore, the analysis of satellite data could be extended to the (water vapour) channels 5 (6.2 μm) and 6 (7.3 μm).

References

[1] Ban, N., Schmidli, J., Schär, C., 2014: Evaluation of the convection-resolving regional climate modeling approach in decade-long simulations. *Journal of Geophysical Research: Atmospheres*, 13, **7889-7907**

[2] Böhme, T., Stapelberg, S., Akkermans, T., Crewell, S., Fischer, J., Reinhardt, T., Seifert, A., Selbach, C., Van Lipzig, N., 2011: Long-term evaluation of COSMO forecasting using combined observational data of the GOP period. *Meteorologische Zeitschrift*, 2, **119-132**

[3] Boos, D.D., 2003: Introduction to the bootstrap world *Statistical Science*, 2, **168-174**

[4] Dai, A., Giorgi, F., Trenberth, K.E., 1999: Observed and model-simulated diurnal cycles of precipitation over the contiguous United States. *Journal of Geophysical Research-Atmospheres*, 104, **6377-6402**

[5] Ebert, E.E., 2008: Fuzzy verification of high-resolution gridded forecasts: a review and proposed framework. *Meteorological Applications* , 1, **51-64**

[6] Eckert, P., 2009: COSMO Priority Project "INTERP"; Final Report. Technical Report, No. 16, **1-34**

[7] Hohenegger, C., Brockhaus, P., Schär, C., 2008: Towards climate simulations at cloud-resolving scales. *Meteorologische Zeitschrift*. 4, **383-394**

[8] Keil, C., Tafferner, A., 2003: LMSynSat User's Guide. Unpublished Manuscript, Institute of Atmospheric Physics, DLR-Oberpfaffenhofen, Germany

[9] Keil, C., Tafferner, A., Reinhardt, T., 2006: Synthetic satellite imagery in the Lokal-Modell. *Atmospheric Research* , 82, **19–25**

[10] Kendon, E.J., Roberts, N.M., Senior, C.A., and Roberts, M.J., 2012: Realism of Rainfall in a Very High-Resolution Regional Climate Model. *Journal of Climate*, 17 **5791-5806**

[11] Knote, C., Heinemann, G., Rockel, B., 2010: Changes in weather extremes: Assessment of return values using high resolution climate simulations at convection-resolving scale.*Meteorologische Zeitschrift* , 1, **11-23**

[12] Langhans, W., Schmidli, J., Fuhrer, O., Bieri, S., Schär, C., 2013: Long-Term Simulations of Thermally Driven Flows and Orographic Convection at Convection-Parameterizing and Cloud-Resolving Resolutions. *Journal of Applied Meteorology and Climatology* , 6, **1490-1510**

[13] Langhans, W., Schmidli, J., Schär, C., 2012: Bulk Convergence of Cloud-Resolving Simulations of Moist Convection over Complex Terrain. *Journal of the Atmospheric Sciences* , 7, **2207-2228**

[14] Pfeifer, M., Yen, W., Baldauf, M., Craig, G., Crewell, S., Fischer, J., Hagen, M., Hühnerbein, A., Mech, M., Reinhardt, T., Schröder, M., Seifert, A., 2010: Validating precipitation forecasts using remote sensor synergy: A case study approach. *Meteorologische Zeitschrift* , 6, **601-617**

[15] Prein, A.F., Gobiet, A., Suklitsch, M., Truhetz, H., Awan, N.K., Keuler, K., Georgievski, G., 2013: Added value of convection permitting seasonal simulations. *Climate Dynamics* , 9-10, **2655-2677**

[16] Roberts, N.M., Lean, H.W., 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Monthly Weather Review* , 136, **78-97**

[17] Schmetz, J., Pili, P., Tjemkes, S., Just, D., Kerkmann, J., Rota, S., Ratier, A., 2002: An introduction to Meteosat Second Generation (MSG). *Bulletin of the American Meteorological Society* , 7, **977-992**

[18] Sideris, I.V., Gabella, M., Erdin, R., Germann, U., 2014: Real-time radar-rain-gauge merging using spatio-temporal co-kriging with external drift in the alpine terrain of Switzerland. *Quarterly Journal of the Royal Meteorological Society* , 680, **1097-1111**

[19] Tiedtke, M., 1989: A Comprehensive Mass Flux Scheme for Cumulus Parameterization in Large-Scale Models. *Monthly Weather Review* , 8, **1779-1800**

[20] Weusthoff, T., Ament, F., Arpagaus, M., Rotach, M.W., 2010: Assessing the Benefits of Convection-Permitting Models by Neighborhood Verification: Examples from MAP D-PHASE, *Monthly Weather Review* , 9, **3418-3433**