Application and Verification of ECMWF Products 2021

Federal Office of Meteorology and Climatology MeteoSwiss

Pirmin Kaufmann, Christoph Spirig

1. Summary of major highlights

-

2. Use and application of products

-

2.1 Direct Use of ECMWF Products

-

2.2 Other uses of ECMWF output

2.2.1 Post-processing

MeteoSwiss has developed and is currently implementing a NWP postprocessing suite for providing improved automated weather forecasts at any location in Switzerland. The aim is a combined postprocessing of MeteoSwiss' local area model COSMO and IFS-ENS to enable seamless probabilistic forecasts over two weeks leadtime. Both ensemble model output statistics (EMOS) and machine learning (ML) approaches including artificial neural networks (ANN) and generative adversarial networks (GAN) have been applied. Both classical and AI approaches are able to improve the forecasts in terms of CRPS by up to 30% as compared to the direct output of the local area model (LAM). The combined postprocessing of LAM and IFS-ENS not only allows to provide seamless predictions from 0 to 15 days, it also provides better skill than postprocessing the two NWPs individually (Keller et al., 2021), as shown in *Figure 1*.

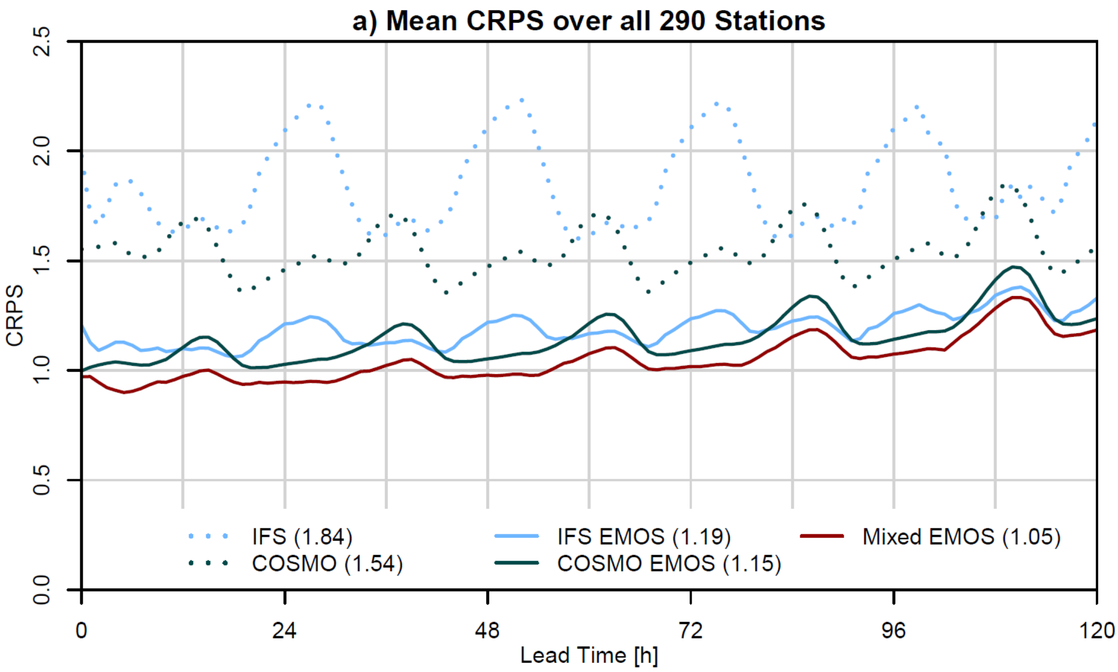

Figure 1: Temperature CRPS of direct model output (dotted) and forecasts postprocessed by EMOS (lines) at forecast lead times covered by both the LAM COSMO-E (2km horizontal resolution) and IFS-ENS. Note that direct model output includes a constant lapse rate correction of temperature for diminishing CRPS effects solely due to different horizontal resolution.

In an explorative ad-hoc study in the framework of a short-term secondment to ECMWF, the potential of using ecPoint (Hewson & Pillosu, 2021) instead of, or in combination with, EMOS to improve precipitation predictions over Switzerland has been assessed. In a nutshell, ecPoint predictions have been compared with LAM COSMO-E and different global EMOS (gEMOS) models, which use statistics like the mean or the fraction of zeros of different sets of pooled predictions from ecPoint, IFS-ENS, and COSMO-E as predictors. While it turned out that ecPoint performed slightly worse than gEMOS based on optimal combinations of IFS-ENS and COSMO-E for low thresholds, it was able to outperform all tested variants of gEMOS for higher thresholds, in particular for longer lead times. *Figure 2* shows ROC curves evaluated over a large set of stations over Switzerland and neighboring areas. Appararently, ecPoints performs quite well for high thresholds of 30 mm for 6 hours accumulated precipitation.

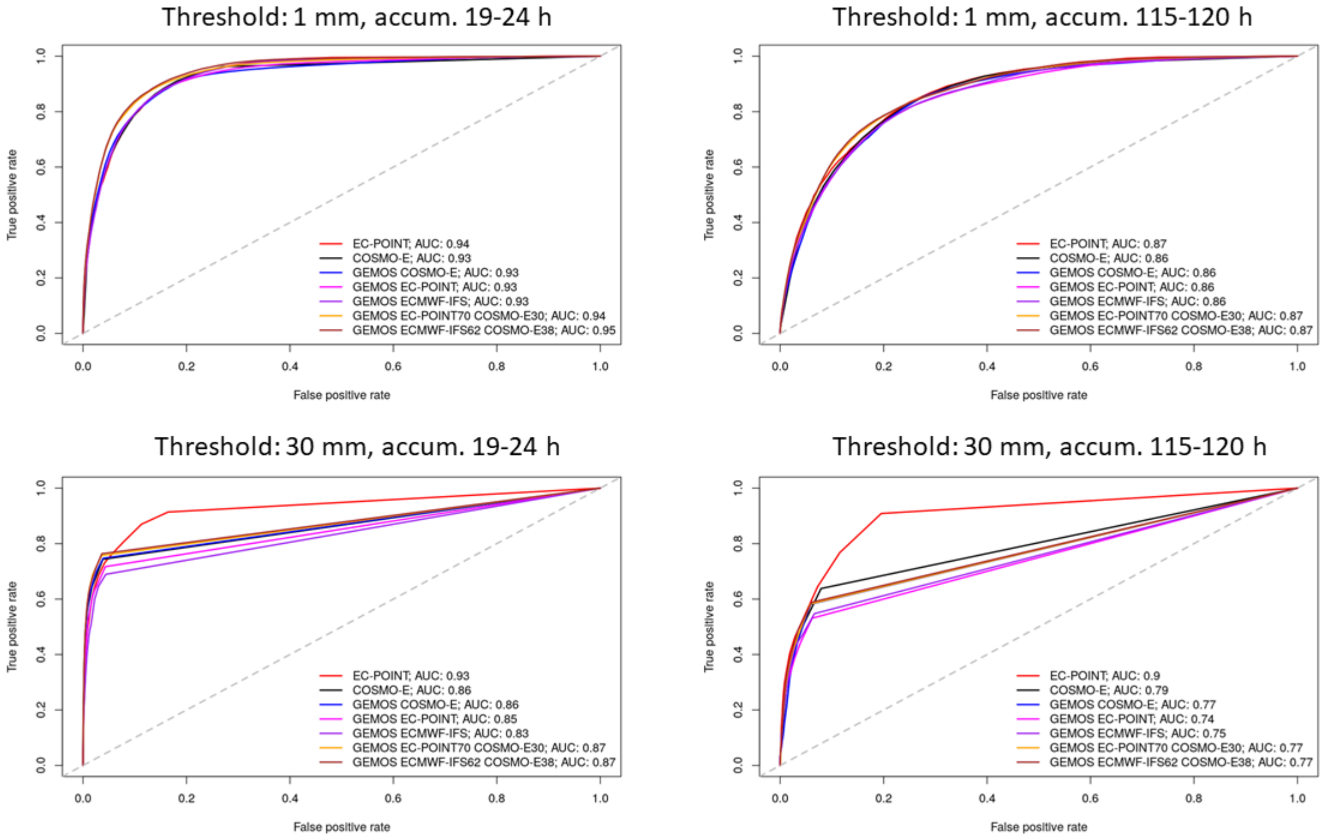

Figure 2: ROC curves comparing ecPoint with COSMO-E, gEMOS based on COSMO-E, gEMOS based on ecPoint forecasts, gEMOS based on IFS ENS, a weighted combination of ecPoint and COSMO-E, and a weighted combination of IFS ENS and COSMO-E. Results shown for 6 h accumulated precipitation and thresholds of 1 and 30 mm.

2.2.2 Derived fields

Include modified ENS output e.g. regimes, clustering, probabilities.

2.2.3 Modelling

ENS boundary conditions are used to run the COSMO limited area model as ensembles in two resolutions. The COSMO-1E ensemble has 10 disturbed members plus one control run that uses HRES instead of ENS at the boundary. The ensemble has a grid spacing of 1.1 km, 80 levels and runs 8 times per day (every 3 hours) out to 33 hours (45 hours in the case of the 03 UTC forecast). The COSMO-2E ensemble has 20 members plus one control run. Its grid spacing is 2.2 km and it has 60 levels. It runs 4 times per day (every 6 hour) out to 120 hours (5 days). The former COSMO-7 deterministic model has been phased out in 2020. The domains with the respective topography are illustrated in *Figure 3*. Both ensembles start from a common assimilation cycle KENDA-1, which implements a 40 member ensemble transform Kalman filter.

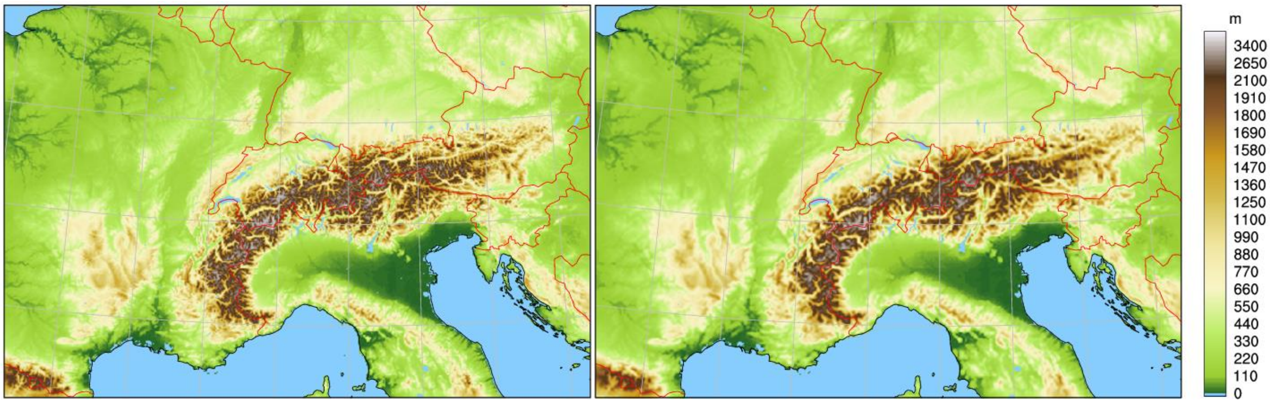

Figure 3: Model domains and orography of COSMO-1E (left) and COSMO-2E (right)

Trajectories are calculated from HRES forecasts with the Lagrangian tool LAGRANTO (Sprenger and Wernli 2015). It uses HRES forecasts in 0.25° resolution for worldwide calculations and 0.1° resolution for calculations in Europe. The trajectories are started from predefined locations in routine production and can be calculated on-demand for arbitrary location for emergency response.The dispersion of air contaminants is simulated with the Lagrangian particle dispersion model FLEXPART (Pisso et al. 2019) based on HRES forecasts in 0.5° resolution worldwide and 0.1° resolution in Europe. The dispersion from a normed source at predefinded locations is calculated routinely. For emergency response purposes, dispersion from arbitrary locations can be simulated on demand.

3. Verification of ECMWF products

3.1 Objective verification

3.1.1 Direct ECMWF model output (both HRES and ENS), and other NWP models

The routine seasonal model verification at MeteoSwiss includes both HRES and ENS. The surface parameters verified on an hourly basis are precipitation (in addition to hourly sums also 6-hourly and 12-hourly sums), total cloud cover, global radiation, sunshine duration (also 12-hourly), 2 m temperature, dewpoint temperature, relative humidity, 10 m wind speed, gusts (also 6-hourly), wind direction, station pressure, and pressure reduced to sea level. These parameters are available at about 160 SwissMetNet stations in Switzerland. Only a subset of these parameters are available from SYNOP stations outside Switzerland within the domain of the COSMO models (approx. 620 stations). *Figure 4* shows the performance of the precipitation forecast of the HRES in comparison to the COSMO-2E model of MeteoSwiss as a performance diagram (Roebber 2009). The measurements are 6-hourly sums from the Swiss automatic rain gauge network, the forecasted value is the sum from lead time 66 to lead time 72 from all 4 model runs per day, for spring 2021 (MAM). COSMO-2E is smoothed by a 3 by 3 spatial averaging filter to somewhat alleviate the double penalty effect. Note that more spatial averaging would be needed for a fair comparison on the scale of HRES. For COSMO-2E, the control run (orange) and the ensemble-median (pink) are shown. All three forecasts show an overestimation of the frequency of occurrence for low thresholds (symbols are above diagonal), with HRES overestimating the most, as has to be expected for a coarser model. For higher thresholds, there is still some overestimation by both HRES and the control run of COSMO-2E. In contrast, the ensemble-median of COSMO-2E shows an increasing underestimation with increasing threshold. This is the effect of taking the median of an increasingly skewed distribution. In terms of critical success index (CSI, also called threat score; curved lines depict equal CSI), the HRES is better than the COSMO-2E control run for several medium thresholds. This is probably due to a double penalty effect, as COSMO-2E shows more structure in its fields even after a 3 by 3 averaging (5 by 5 averaging would be needed to obtain the same spatial smoothness as HRES). For the highest threshold however, the success ratio (x-axis) and the probability of detection (y-axis) of the COSMO-2E control run are both clearly higher than those of HRES, leading to a better CSI.

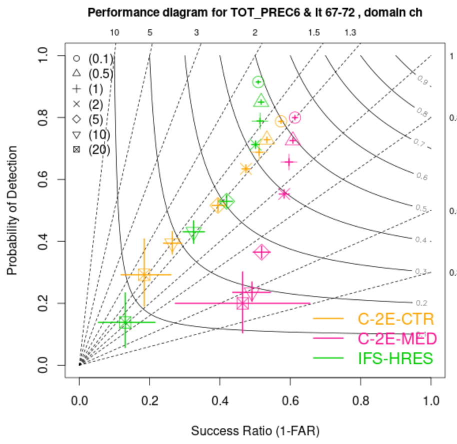

Figure 4: Performance Diagram for the forecasts of COSMO-2E control run (orange), COSMO-2E ensemble median (pink), and HRES (green), for 7 thresholds (different symbols). Forecast value: 6-h sum of precipitation from lead time 66 to 72 from 4 runs per day at Swiss stations.

3.1.2 Post-processed products and end products delivered to users

ERA-Interim forecasts as a benchmark for MeteoSwiss public weather forecasts

Since 2010 MeteoSwiss computes a score for reporting the quality of public weather forecasts. This score is an aggregate of mean absolute error-like scores of temperature, precipitation, sunshine duration and wind speed at selected reference sites in Switzerland. In order to assess how much of the intra- and interannual variability in this score can be attributed to changes in weather predictability on a larger scale, an analysis on the variability of ERA Interim forecast quality in Switzerland was carried out. The rationale was to take advantage of ERA Interim as an essentially constant forecast system to serve as a benchmark for investigating the MeteoSwiss forecast quality subject to continuous changes and developments over the same time period.

ERA Interim forecasts were first postprocessed to remove the expected systematic biases of a rather coarse resolution model in a topographically complex region such as Switzerland. Then the same aggregated score was computed for these postprocessed ERA Interim forecasts and the official MeteoSwiss forecasts.

Figure5 shows these scores for two postprocessing variants of ERA Interim and MeteoSwiss (MCH) forecasts. Considering yearly averaged scores it can be seen that ERA Interim shows no trend in quality over these 10 years whereas MeteoSwiss forecasts show an overall quality increase. Still, significant deviations from this general trend over periods of 1-2 years can be identified. While some of these, e.g. increased performance in 2011, are apparent in both ERAInterim and MeteoSwiss forecasts, others only appear in MeteoSwiss forecasts. The latter can be related to changes in the forecasting procedure.

The findings underline the importance of analysing sufficiently long records of verification scores when assessing improvements of forecast systems. Only this long-term picture allows to distinguish a momentary increase of score from a positive trend. An exploratory analysis of weather type frequency (using objective classification) during this period could not identify a direct link between forecast quality anomalies and occurrence of certain weather types. While there was positive correlation of dry anomalies and forecast performance, the inability to identify a link to weather types may be due to the limited sample size covering only 10 years.

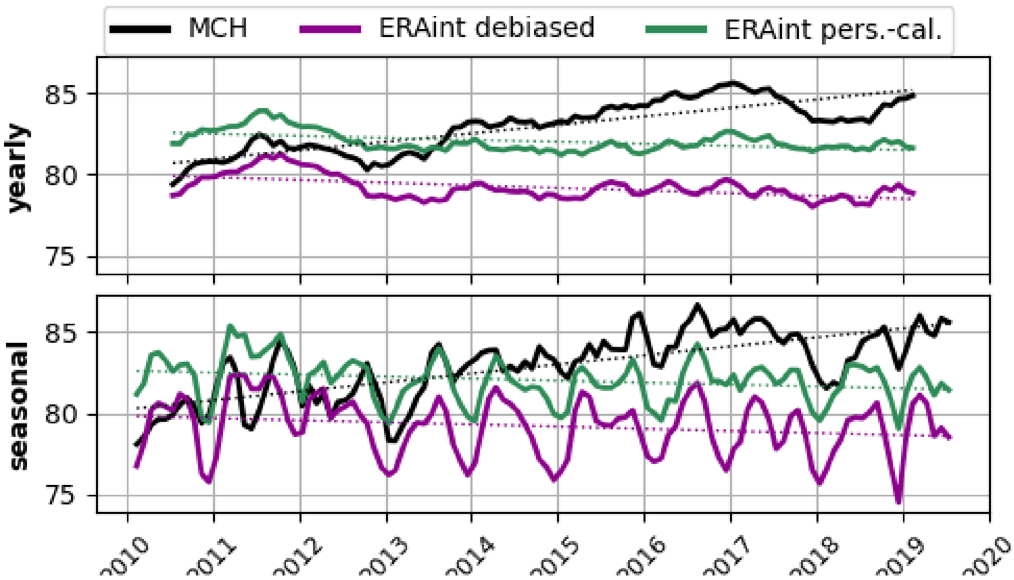

Figure 5: Comarison of MeteoSwiss public weather forecast (MCH) and two variants of postprocessed ERA-Interim scores from 2010 to2020. The score is positively oriented and saturates at 100 for perfect forecasts. Top: Yearly moving average window, bottom: seasonally moving average window.

3.1.3 Monthly and Seasonal forecasts

-

3.2 Subjective verification

3.2.1 Subjective scores

-

3.2.2 Case studies

-

4. Requests for additional output

-

5. References to relevant publications

(Copies of relevant internal papers may be attached)

Pisso, I., Sollum, E., Grythe, H., Kristiansen, N. I., Cassiani, M., Eckhardt, S., Arnold, D., Morton, D., Thompson, R. L., Groot Zwaaftink, C. D., Evangeliou, N., Sodemann, H., Haimberger, L., Henne, S., Brunner, D., Burkhart, J. F., Fouilloux, A., Brioude, J., Philipp, A., Seibert, P., and **Stohl, A.**, 2019: The Lagrangian particle dispersion model FLEXPART version 10.4. *Geosci. Model Dev.*, **12**, 4955–4997, https://doi.org/10.5194/gmd-12-4955-2019.

Keller, R., J. Rajczak, J. Bhend, C. Spirig, S. Hemri, M. A. Liniger, and **H. Wernli, 2021:** Seamless Multimodel Postprocessing for Air Temperature Forecasts in Complex Topography. *Wea. Foreacasting*, 36 (3), 1031 – 1042, https://doi.org/10.1175/WAF-D-20-0141.1

Roebber, P.J., 2009: Visualizing multiple measures of forecast quality. Wea. Forecasting, 24, 601-608, https://doi.org/10.1175/2008WAF2222159.1 .

Sprenger, M. and **H. Wernli**, 2015: *Geosc. Model Dev*., **8**, 2569–2586, DOI: 10.5194/gmd-8-2569-2015

Authors sections 2.2.1 and 3.1.2: *C. Spirig[1], R. Keller[1,2] , S. Hemri [1,3], J. Rajczak[1], M. Gerber[2], J. Bhend[1]*

1: MeteoSwiss, Development of Forecasting
2: Centre for Climate Systems Modelling, C2SM , ETH Zurich
3: Department of Mathematics, University of Zurich