Application and verification of ECMWF products 2014

MeteoSwiss

1. Summary of major highlights

2. Use and application of products

Include medium-range deterministic and ensemble forecasts, monthly forecast, seasonal forecast

2.1 Post-processing of model output

2.1.1 Statistical adaptation

Extended and long range forecasts

In an ongoing effort of improving the usability of extended and long range forecasts, MeteoSwiss currently explores long range forecasting of application-specific climate indices. Typically such indices are based on daily values of meteorological variables and involve the use of thresholds. MeteoSwiss therefore investigates the use of ECMWF's extended range and long range forecasts at daily resolution and the application of post-processing techniques for deriving forecasts of such indices. Bias-correction is essential for indices involving thresholds, but estimating biases of daily data proved to be difficult. There are only about thirty years of observations to derive a daily climatology. Our analyses based on System 4 hindcasts and using the perfect model approach allowed us to develop an appropriate post-processing method for reducing temperature biases of System 4. A temporal smoothing can improve the estimates substantially as illustrated in Figure 1 and documented in Mahlstein et al. (submitted).

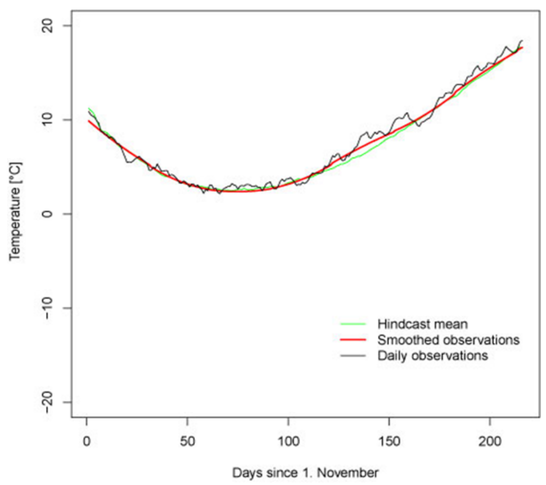


Figure 1: Example of the hindcast mean (green), the mean daily observations (black), and the fitted observations (red) for one specific grid cell in Western Europe.

At the example of forecasting heating degree days (HDD), an index considering days with temperatures below 17° serving as a proxy of heating energy demand, the benefit of an appropriate bias-correcation can be illustrated. Skill of forecasting HDD terciles (verified against ERA-Interim) can be improved by bias-correction (Figure 2).

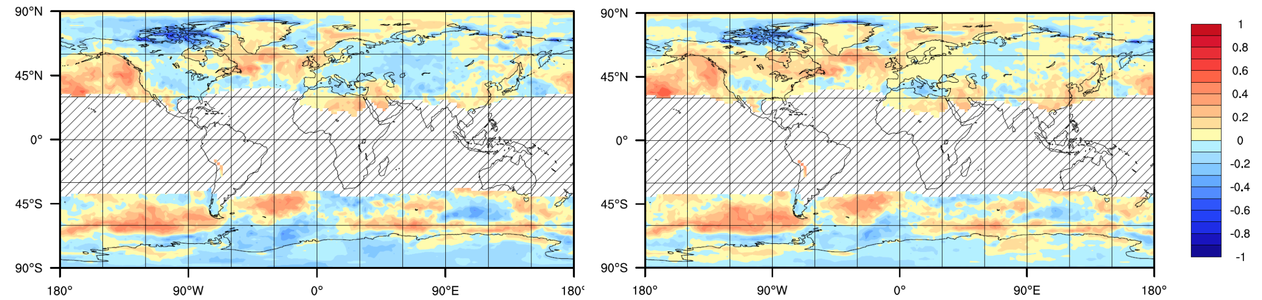


Figure 2: Ranked probability skill scores for predicting HDD terciles, left: November forecasts for DJF based on raw System 4 output, right: same forecasts using bias-corrected System 4 output.

2.1.2 Physical adaptation

MeteoSwiss runs its own short-range forecasting system. The core of this system is the non-hydrostatic model COSMO (www.cosmo-model.org). It is running operationally at two spatial scales: The regional model COSMO-7 with a horizontal resolution of about 6.6 km is driven by the ECMWF global model IFS. The local model COSMO-2, having a horizontal grid spacing of about 2.2 km, is nested in COSMO-7. The nesting of NWP models is illustrated in Fig. 1.

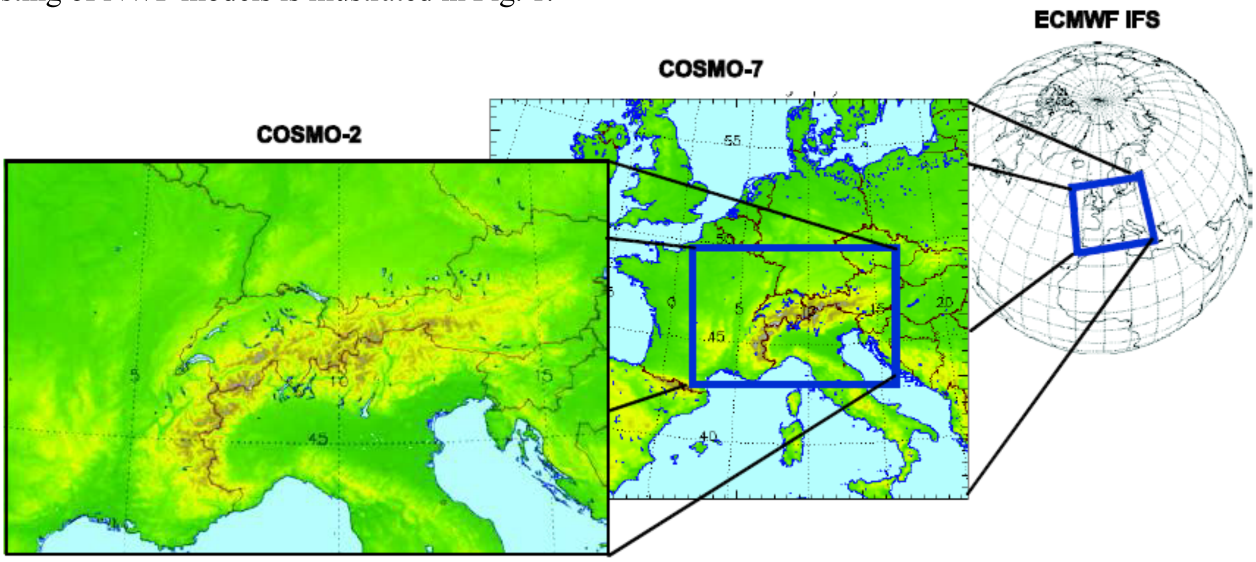


Fig. 1 NWP System of MeteoSwiss

Both COSMO-7 and COSMO-2 have their own assimilation cycle, which is updated in intervals of 3 hours. Three daily 72 hours COSMO-7 forecasts are calculated, based on the 00, 06 and the 12 UTC IFS (main or boundary conditions) runs. One COSMO-2 forecast is computed every 3 hours just after the computation of the necessary COSMO-7 boundary conditions. The lead time of the COSMO-2 forecast starting at 03 UTC is 45 hours, and 33 hours otherwise. The cut-off time for all forecasts is 45 minutes.

An on-demand mode can be activated, e.g. in case of an incident in nuclear power plants. COSMO-2 is then computed hourly with at least 4 hours assimilation and 6 hours forecast.

A sophisticated set of scripts controls the whole operational suite, and allows for a very high reliability of the system, with less than 2% of the forecasts requiring manual intervention. This same environment is also used to run parallel suites to validate proposed modifications to the system and to facilitate experimentation by the modelling group.

The computing resources and expertise are provided by the Swiss National Supercomputing Centre (CSCS, www.cscs.ch). COSMO-2 and COSMO-7 are calculated in parallel on a Cray XE6 equipped with 12-cores Opteron Magny-Cours with 1728 computational cores and achieve a sustained performance of 280 GFlops on 1079 computational Cores for COSMO-2. Pre- and post-processing run on the service nodes of the machine. A similar machine with 4032 computational Cores is available as fail-over and used for research and development. A large multi-terabytes long-term storage is used for archiving purposes and a 1 GBit/s link connects the MeteoSwiss main building with the CSCS (on the other side of the Alps!).

2.1.3 Derived fields

Including post-processing of EPS output e.g. clustering, probabilities

2.2 Use of products

Use of ECMWF products in operational duties, in particular use in severe weather situations

3. Verification of products

Include medium-range deterministic and ensemble forecasts, monthly forecast, seasonal forecast. ECMWF does extensive verification of its products in the free atmosphere. However, verification of surface parameters is in general limited to using synoptic observations.
More detailed verification of weather parameters by national Services is particularly valuable.

3.1 Objective verification

3.1.1 Direct ECMWF model output (both deterministic and EPS)

We performed a verification of the deterministic precipitation forecasts for selected regions of Switzerland issued by the bench forecasters and compared them to ECMWF's IFS outputs (12h00 UTC runs of HRES and ENS). Our main aim was to identify in which contexts the forecaster can bring a significant additional value to the NWP output. An assessment of the skill of the forecasters against NWP as well as of the skill of the ENS against HRES can be useful to choose the best forecast guidance.
Precipitation amounts are predicted by forecasters up to seven days ahead for regions and represent spatial averages. The verified forecasts were: VAL: daily regional mean amounts predicted by forecasters; IFS HRES: spatial average over same regions than VAL; IFS ENS: spatial average over same regions than VAL of the ensemble median. In this exploratory study, the focus was on three regions with different climatologies: Plateau West (lowland area north of the Alps), Alps West (mountains in the Alps) and Ticino South (lowland south of the Alps). The verification period runs from 2010 to 2013. Forecasts were compared with regional average observations provided by a tool recently developed at MeteoSwiss which combines high resolution radar images with raingauge measurements[1]. The scoring rule used for the verification is a measure of accuracy based on the error between forecast and observation in a suitably rescaled space. It is a part of the global quality score[2] used at MeteoSwiss since 2013.

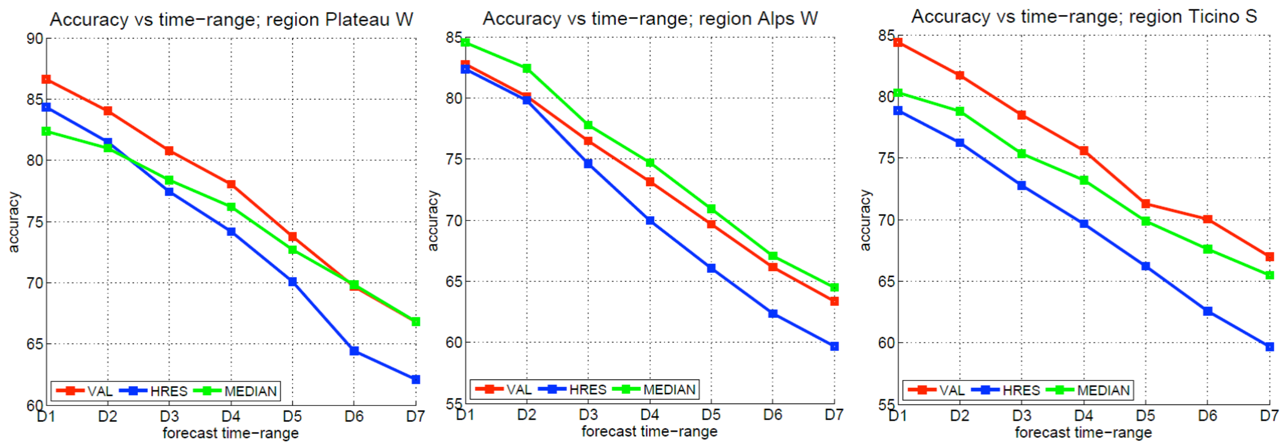


Figure 3 Forecasts accuracy against time-range for the forecasts: 1) issued by forecasters (VAL), 2) IFS HRES, 3) median of IFS ENS, for three regions of Switzerland. The verification period runs from 2010 to 2013.

For the verified regions, the ENS median exhibits on the average a better accuracy than HRES, even for short-term forecasts (except for Plateau West). When looking at the bias, for short-term forecasts (D1-D2) HRES exhibits a significant overestimation of quantities for regions Alps West and Ticino South. On the other hand, for the same regions, the ENS median underestimates precipitation amounts for long-term forecasts (D6-D7); for the region Alps West, this is especially the case during Summer and Autumn.
This preliminary study indicates that the ENS median is a good candidate to serve as guidance for forecasters and should be considered more frequently, taking care of compensating the underestimating bias whenever necessary. A more detailed analysis including all regions of Switzerland will be performed in

[1] I. V. Sideris, M. Gabella, R. Erdin and U. Germann, “Real-time radar-raingauge merging using spatiotemporal co-kriging with external drift in the alpine terrain of Switzerland”, *Quaterly Journal of the Royal Meteorological Society*, 00: 1-22, 2011.
[2] D. Cattani, A. Faes, M. Giroud Gaillard and M. Matter, “COMFORT: continuous MeteoSwiss forecast quality score”, submitted.

order to confirm and extend the previous results. As further work, it might be also interesting to consider other quantiles from the ENS, depending on region/season/time-range, in order to determine the best "first guess".

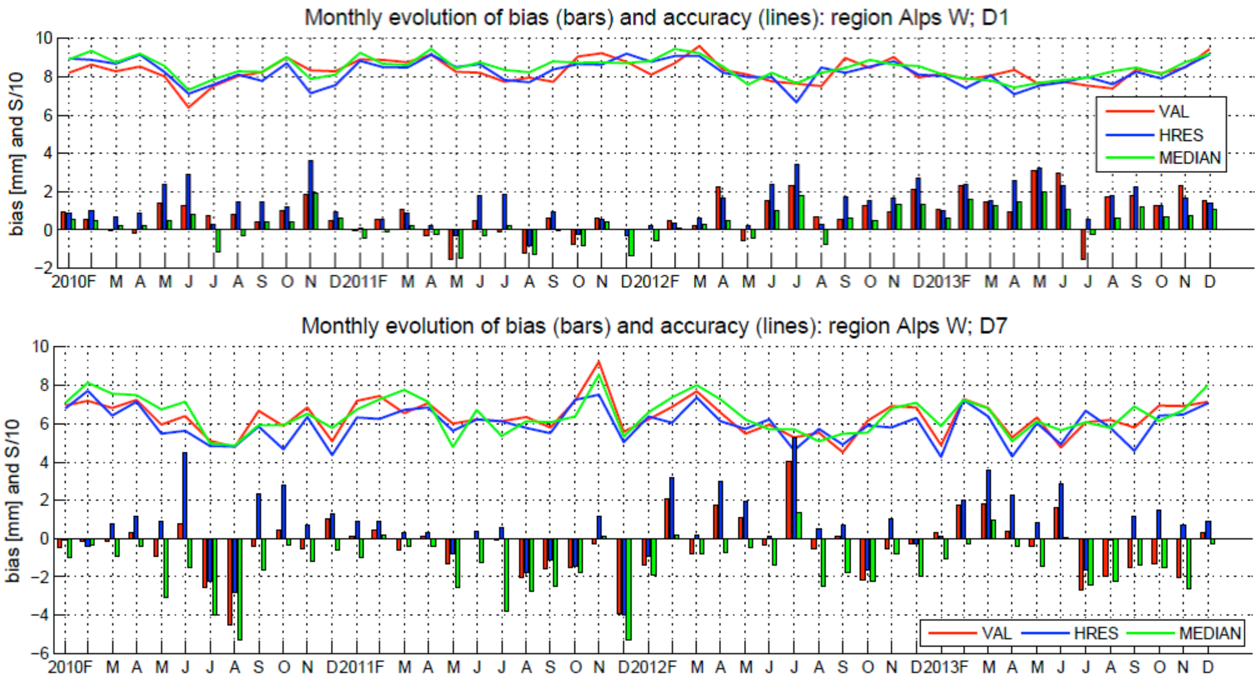
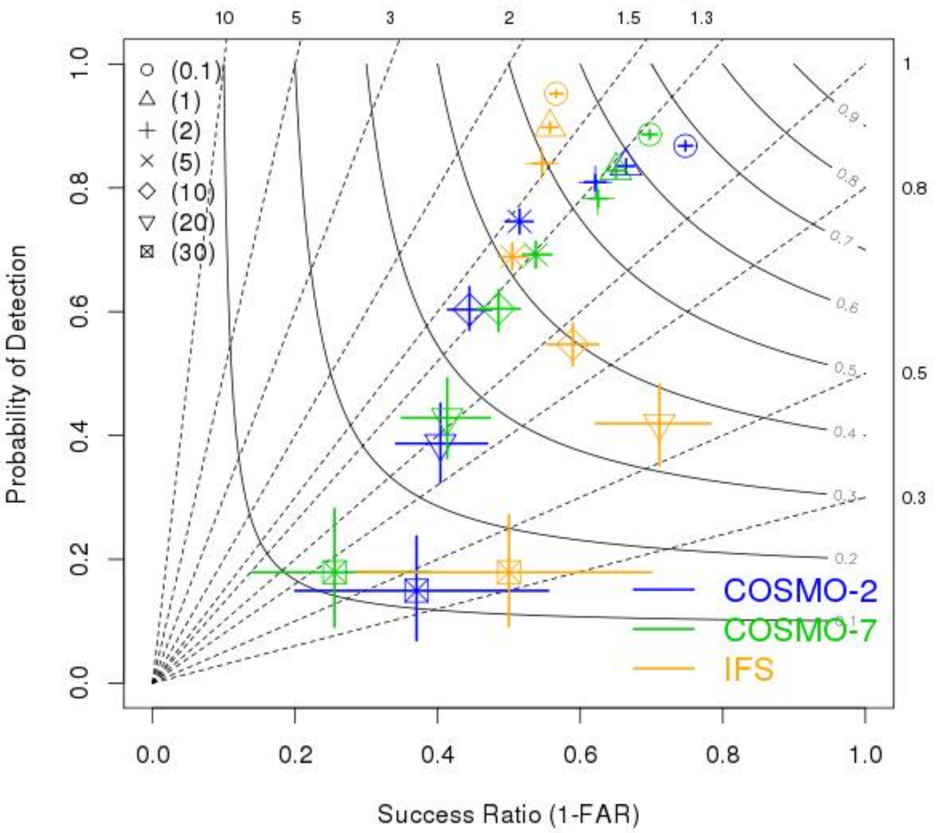


Figure 4 Monthly evolution of bias and accuracy for the forecasts: 1) issued by forecasters (VAL), 2) IFS HRES, 3) median of IFS EPS, for the region *Alps West.* The time-ranges are D1 (top plot) and D7 (bottom plot).

3.1.2 ECMWF model output compared to other NWP models

As part of the operational seasonal verification with SYNOP observations IFS forecasts are regularly compared to the COSMO models operated at MeteoSwiss. For this, parameters such as pressure, 2 m temperature, 2 m dewpoint, 10 m wind speed, cloud cover, 12-hourly precipitation, and 10 m wind gusts



are compared to associated observations. For precipitation as one of the most important parameters, an example of a verification summary is shown in Fig. 2. Compared to COSMO-2 and COSMO-7, IFS shows a stronger overestimation of low precipitation amounts and also a stronger underestimation of high amounts. For thresholds of 5 mm/12 hours the values of all three models are quite similar.

Fig. 2 Comparison of the 12 hours accumulated precipitation forecast performance for Spring 2014 of all 00 and 12 UTC forecasts of COSMO-2, COSMO-7 and IFS for the 117 Swiss stations and the lead time range 13-24 hours for 7 different thresholds in a performance diagram derived from the geometrical relationship between several scores based on the contingency table (Roebber, 2009). Plotted is the success ratio (1-false alarm ratio) against the probability of detection (POD) with the dotted auxiliary lines indicating the frequency bias (FBI) with the values at the margin axes of the plot and the solid black lines denoting isolines of the critical success index (CSI) with the values within the plot area on the right. The perfect forecast lies at the upper right corner of the plot. The different thresholds are visualised with different symbols, the results of the different models have different colours. The cross on each symbol depicts the 95th quantile of the sampling uncertainty derived from bootstrapping (with N=1000 random draws) giving an indication about the confidence in the result.

3.1.3 Post-processed products

Verification of extended range forecasts

In addition to our verification work on daily output of long range forecasts (see section 2.1.1), we performed similar skill analyses of post-processed extended range forecasts against ground observations of our national measurement network.

Figure 5 shows the continuous ranked probability skill score (CRPSS) for a sample of 16 initialization dates (Jan-April) and 10 observation sites using the ER hindcasts of the past 20 years (model cycle 40r1). Skill varied significantly, but overall there seems to be potential to predict site-specific weekly means up to week 3. The application of more sophisticated post-processing techniques and doing regional averages rather than site-specific predictions offer potential for improvements.

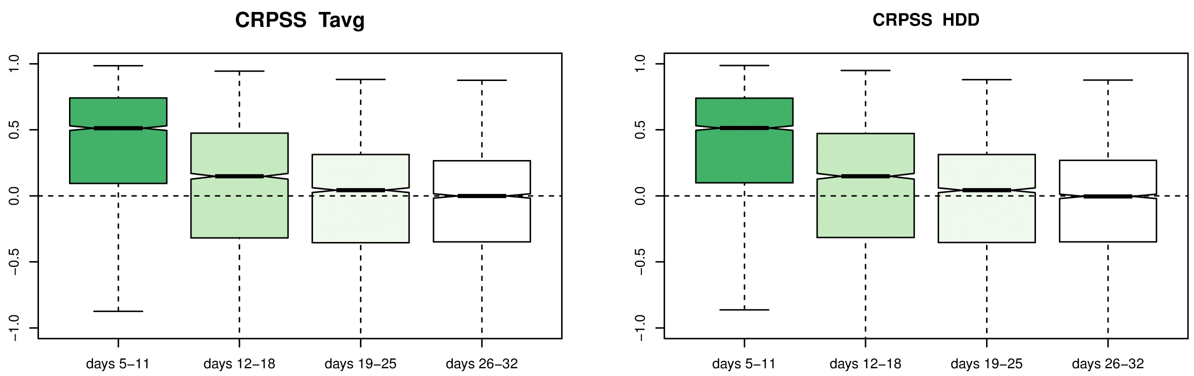


Figure 5: Continuous ranked probability skill score of weekly mean temperature forecasts and heating degree days (HDD) forecasts for 10 sites in Switzerland.

e.g. Kalman filtered products, calibrated EPS probabilities, etc

e.g. Kalman filtered products, calibrated EPS probabilities, etc.

3.1.4 End products delivered to users

3.2 Subjective verification

3.2.1 Subjective scores (including evaluation of confidence indices when available)

In Locarno a subjective very simple point forecast verification has been conducted for a long period of time (since 1997). Following daily station parameter are considered: precipitation amount, relative sunshine duration and temperature deviation from the climatology. The forecast method and verification are quite consistent in time. For this reason the improvement of the forecast is essentially due to the improvements of the available numerical models, particularly in the medium range (i.e. ECMWF). Figure 3 shows that the score of 2012 was comparable with the mean values of the previous 5 years for all forecast lead times. Interesting is the improvement in the forecast day 3 / 7 compared to the mean values of the period 2002 – 2006. For the forecast day 8 / 10 the results apparently don't show an improvement of the accuracy of the forecast for Locarno in the last 10 years. We think that this is more due to the behaviour of the forecaster in the forecasting range 8 / 10 days, than in a lack of model improvements. The absolute values of the verifications are very specific to the verification method.

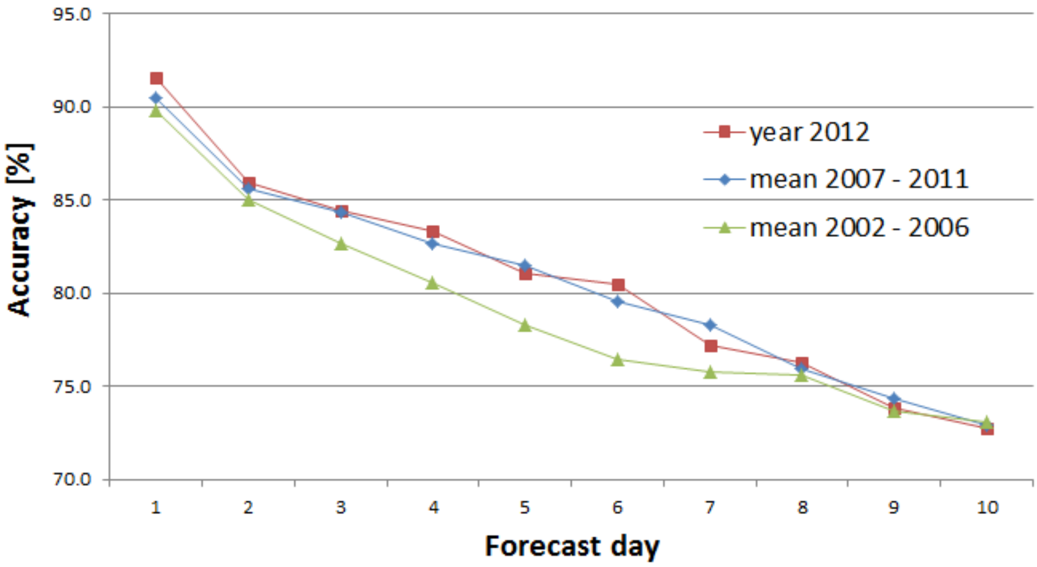


Figure 6 Forecast verification for Locarno. Results for year 2012 compared with the mean values of the previous 5 years and for the period 2002 – 2006, for all forecast lead time

3.2.2 Synoptic studies

Including evaluation of the behaviour of the model

4. References to relevant publications

Mahlstein, I., M. A. Liniger, and C. Appenzeller: Estimating daily climatologies for climate indices derived from climate model data and observations. Submitted to Journal of Geophysical Research – Atmospheres.

Roebber, P.J., 2009: Visualizing multiple measures of forecast quality. Wea. Forecasting, 24, 601-608.