Progress report on 3DVAR testing at CHMI

Alena Trojáková, December 2013

Contents

1 Introduction

This report summarizes studies performed at CHMI during the first half of 2013. The upper-air analysis scheme was of our main interest with aim to replace the operational digital filter (DF) blending scheme (Brožkova et al 2001) by 3DVAR based technique which uses observation directly.

The analysis scheme was tested in a simplified experimental framework. The aim was to quickly check performance of the ALADIN 3DVAR and to prepare a baseline setup for further development.

The experimental setup is described in section 2. Section 3 is dedicated to background errors diagnostics. Impact of selected background errors is presented in section 4. The background and observation error diagnostics and their tuning is examined in section 5. Performance of the baseline setup is presented in section 6 and the conclusions are given in the last section.

2 Experimental setup

The local ALADIN/CE model release 36t1ope as operational in July 2011 was used. It comprised horizontal resolution of 4.7 km and 87 vertical levels, linear truncation E269x215, mean orography, 3h coupling interval and time step 360 s.

The ALADIN 3DVAR analysis scheme was used with ensemble based background errors in the **simplified experimental framework**. It consisted of an experiment without assimilation cycling, the surface analysis and initialization were suppressed. The conventional data SYNOP (ϕ) and TEMP ($T, q, wind$) were assimilated. Dynamical adaptation was used as reference.

The two days forecasts were performed for the two weeks **period of February 1-14, 2013**. The objective scores against observation were evaluated by verification package VERAL. The BIAS, STDE and RMSE scores were computed from differences between the forecasts and observations (SYNOP and TEMP). The same scores were computed with respect to ARPEGE and/or ECMWF analyses and the significance tests described in Fisher (2002) of the RMSE differences were performed.

3 Background error statistics

Background error statistics (further referred also as **B**) are essential 3DVAR component. There are several typical characteristics which are usually examined, e.g. standard deviations which correspond to the expected amplitude of background errors, correlations (or length-scales) that determines how local observation is spatially filtered and propagated to the neighborhood, and cross-covariances between the different variables (divergence, vorticity, temperature, surface pressure and humidity) which usually reflect physical couplings between different variables, e.g. geostrophic balance.

The background errors were produced for three months period of February - May 2011 based on NMC, NMC lagged and ensemble method. Their typical diagnostics were briefly checked. The first four members (mem_i) from Assimilation Ensemble of global model ARPEGE (AEARP) were downscaled to ALADIN/CE resolution. The background errors were sampled from 240 +6H forecast differences ($mem_1 - mem_2$, $mem_3 - mem_4$) valid at 00, 06, 12 and 18 UTC and 960 differences valid in all (00, 06, 12 and 18UTC) analysis times, they will be referred as ENS_00, ENS_06, ENS_12, ENS_18 and ENS_all.

The ensemble based standard deviations were mostly the smallest, while Berre et al (2006) and Fischer et al (2005) presented the ensemble based standard deviations intermediate between NMC lagged and NMC. The same feature was apparent on the horizontal variance spectra, see Fig 1. One possible explanation for this can be that AEARP uses 4D-Var (instead of 3D-Var) since 2010 and the derived standard deviations were smaller for Arpege ensemble 4D-Var than for Arpege ensemble 3D-Var by a factor around 1.5 (L. Berre personal communication). Regarding cross-covariances except generally smaller values for the ensemble based **B** no particular discrepancy was found, for an illustration see Fig 2.

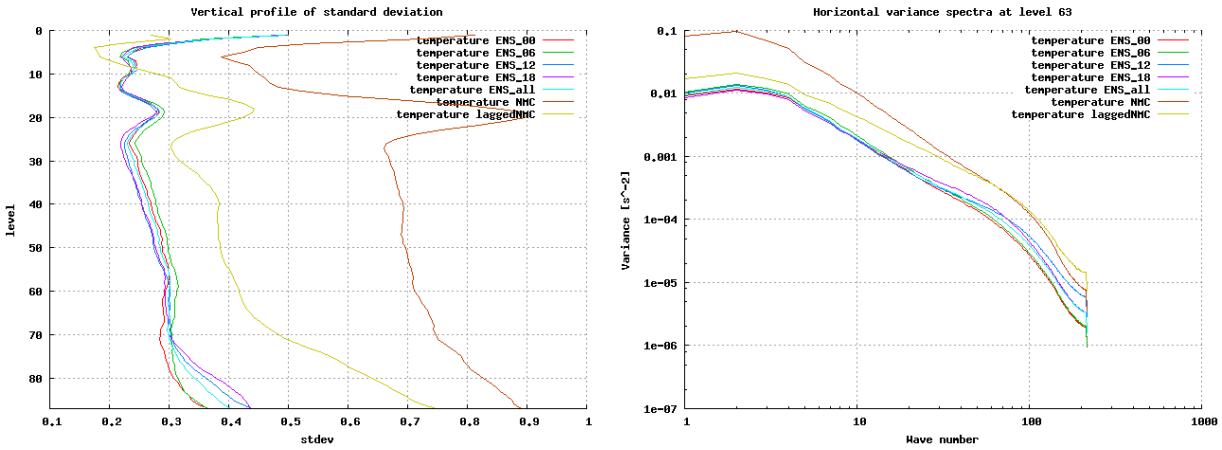

Fig. 1: Vertical profile of the standard deviation for temperature [K] and variance spectra of temperature at level 63 (around 844 hPa) for the NMC, the NMC lagged and the ensemble based methods.

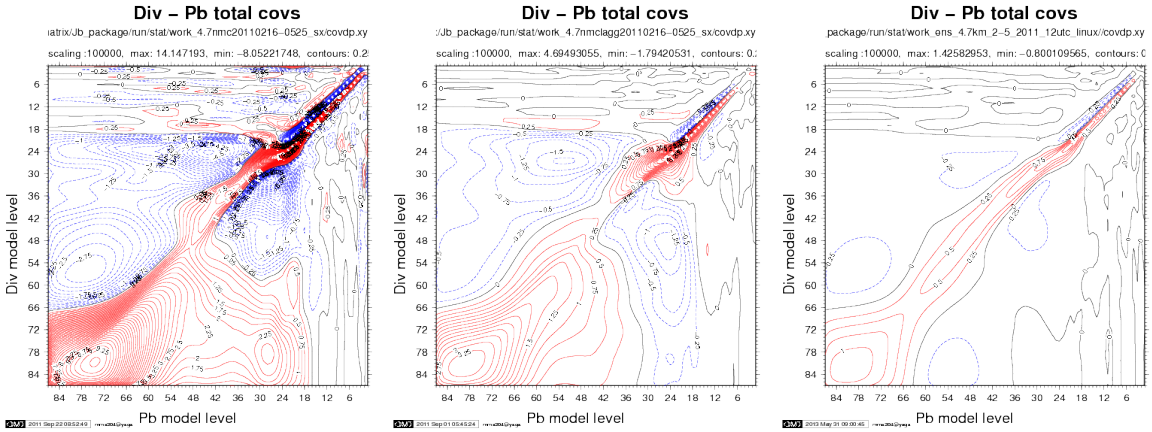

Fig. 2: Divergence/vorticity coupling - mean vertical cross-covariance between divergence and vorticity-balanced ϕ [10^{-5} J/kgs^{-1}] for the NMC (left), the NMC lagged (middle) and the ENS_12 (right).

The length-scales of ensemble based **B** were mostly the largest above 100 hPa, see Fig 3. Disregarding the changes in AEARP, the NMC method based on differences between 36H and 12H forecasts was expected to provide overestimated error correlations, thus smaller ensemble based covariances values seemed reasonable. It was hard to make any judgment regarding an intercomparison of NMC lagged and ensemble based **B**.

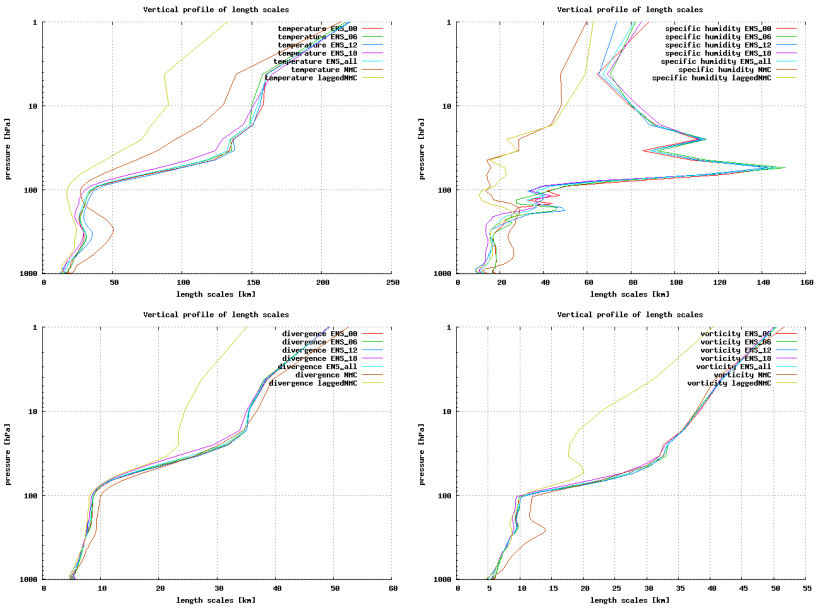

Fig. 3: Vertical profile of the length-scales.

The ensemble based background errors are believed to better represent analysis step and the short range forecast range (Berre et al 2006). A dependency of the **B** sampling on analysis time was of our interest. The vertical profile of the standard deviations, displayed on Fig 4, illustrated this dependency. The differences for 12 and 18UTC mainly in lower levels may be related to atmospheric stratification. Similarly broader and larger auto-correlation and cross covariances were found for 12 and 18UTC, see Fig 5.

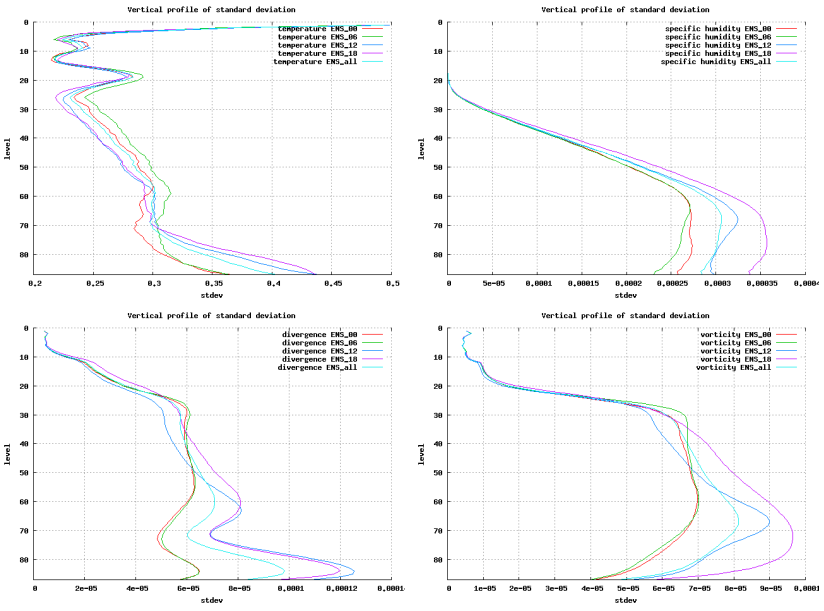

Fig. 4: Vertical profile of the standard deviation for temperature $[K]$, specific humidity $[kg/kg]$. divergence $[s^{-1}]$ and vorticity $[s^{-1}]$.

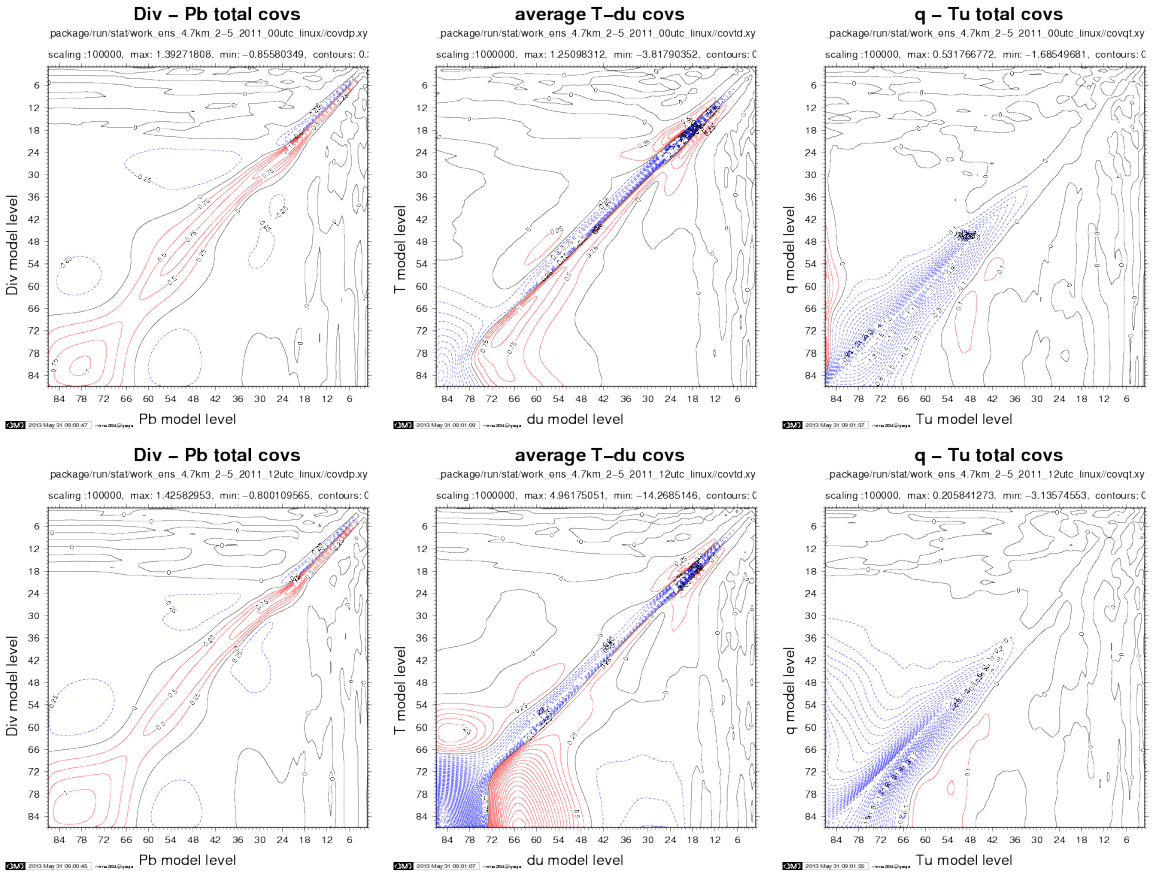

Fig. 5: Mean vertical cross-covariance between divergence and vorticity-balanced $[10^{-5}J/kgs{-}1]$ (left), temperature and unbalanced divergence $[10^{-6}K/s]$ (middle) and humidity and unbalanced divergence $[kgkg^{-1}s^{-1}]$ (right) for 00UTC (top) and 12UTC (bottom)

Detailed analysis of various background errors estimations methods was beyond the scope of this report. The typical diagnostics were presented to briefly check available background errors. Further diagnostic figures can be found in appendix A.

4 Impact of selected background errors

Following experiments with ENS_00 and ENS_12 were performed to examine impact of various background errors. The forecasts starting from 00 and 12 UTC were run for the evaluation period.

- **Y84** - 3DVAR ENS_12
- **Y86** - 3DVAR ENS_00

4.1 Scores against observations

Better performance of ENS_00 for 00UTC forecasts while ENS_12 for 12UTC ones was expected, but only very small differences between their usage were found in comparison with respect to default reference - dynamical adaptation. The results are illustrated on following Fig. 6-Fig. 8 for the humidity scores. Other variables had similar behavior (not shown).

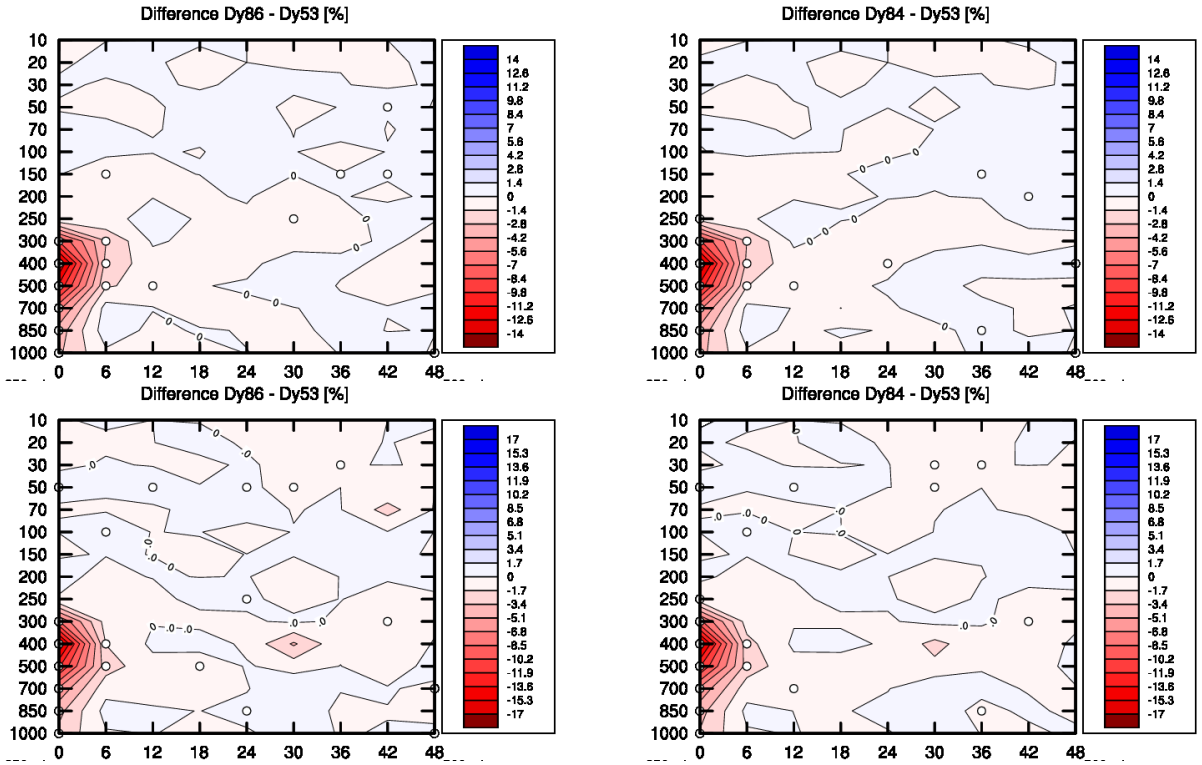

Fig. 6: RMSE differences of RH for 1-14 Feb 2013 of 00 UTC (top) and 12 UTC (bottom) runs. Red areas denote positive impact of the usage of ENS_00 (left) and ENS_12 (right) with respect to dynamical adaptation. The white circles points that RMSE difference is better/worse with significance 95 % two-side confidence interval.

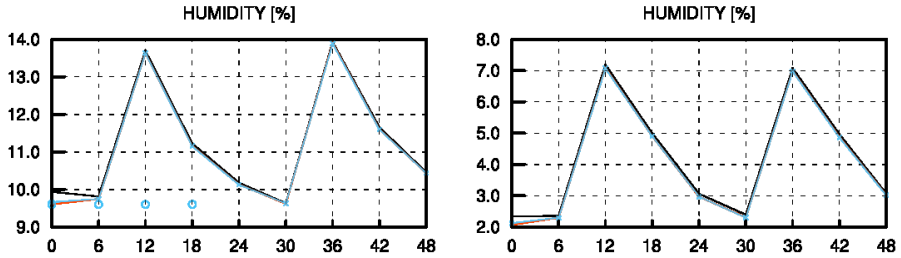

Fig. 7: RMSE (left) and BIAS for $RH2m$ (right) for 1-14 Feb 2013 of 00 UTC runs. The experiment **Y86** in blue, **Y84** in red and reference **Y53** in black.

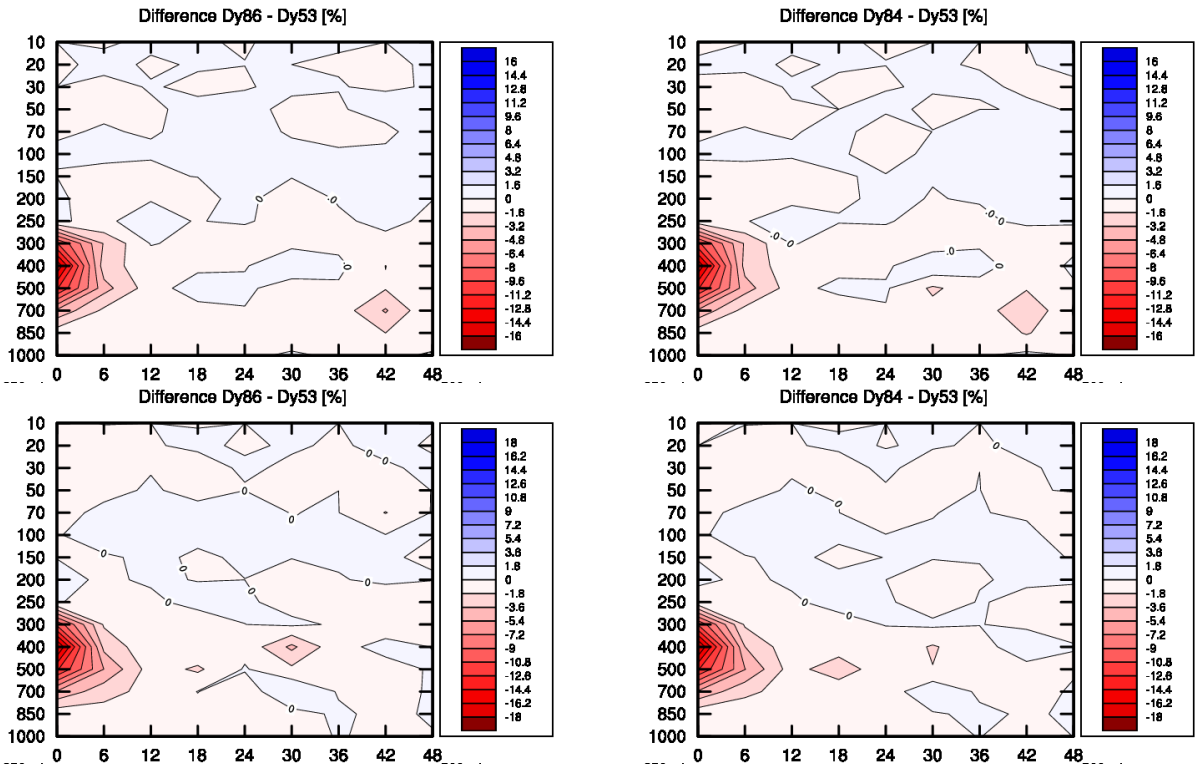

Fig. 8: as previous figure, but BIAS differences of RH. Blue/red areas denote positive/negative differences of the BIAS scores.

Direct inter-comparison of the two experiments **3DVAR ENS_00 vs 3DVAR ENS_12** against observation showed quantitative differences, but more or less neutral scores.

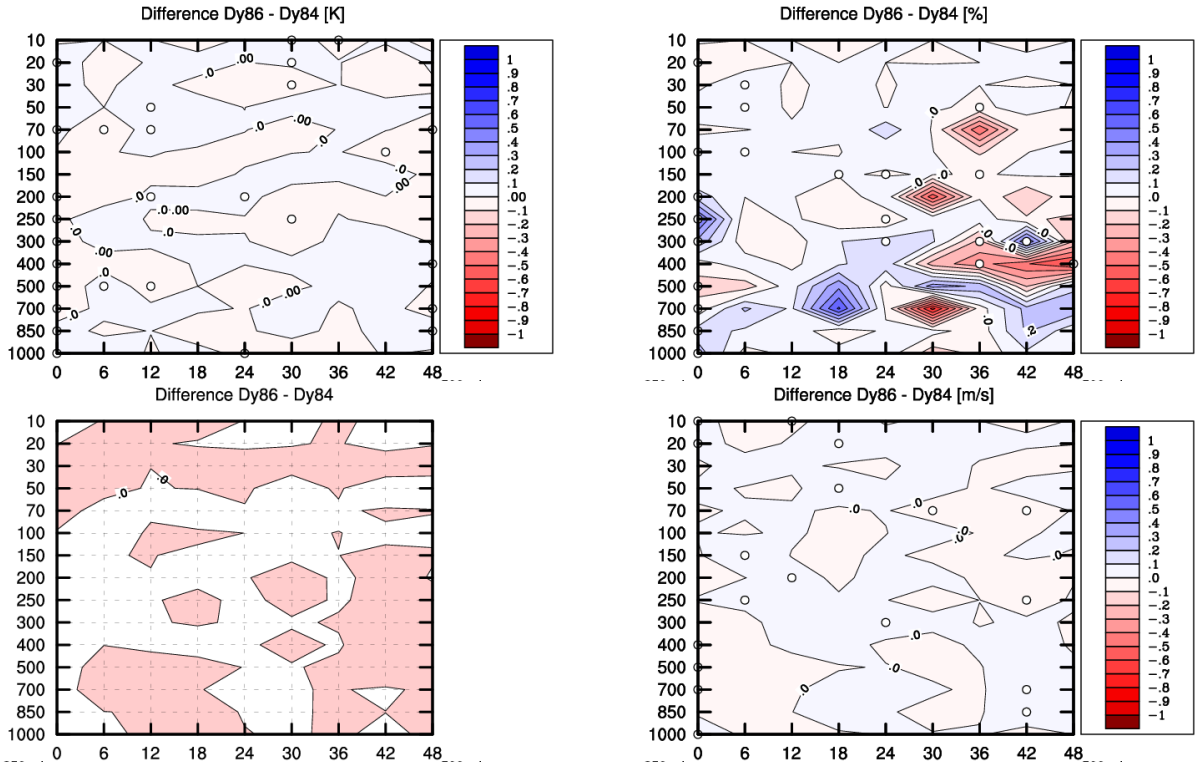

Fig. 9: RMSE differences of T (upper-left), RH (upper-right), ϕ (bottom-left) and wind speed (bottom-right) for 1-14 Feb 2013 of 00 UTC runs. Red areas denote better performance of ENS_00 over ENS_12. The white circles points that RMSE difference is better/worse with significance 95 % two-side confidence interval.

At analysis time for **both 00 UTC and 12 UTC runs** the ENS_00 performed a little bit better

for relative humidity near 500 hPa, for temperature and wind speed between 850-200 hPa. While ENS_12 had smaller RMSE for geopotential up to 200 hPa, for humidity below 850 hPa and between 300-200 hPa, for temperature below 850 hPa and for wind speed at 1000 hPa at analysis time.

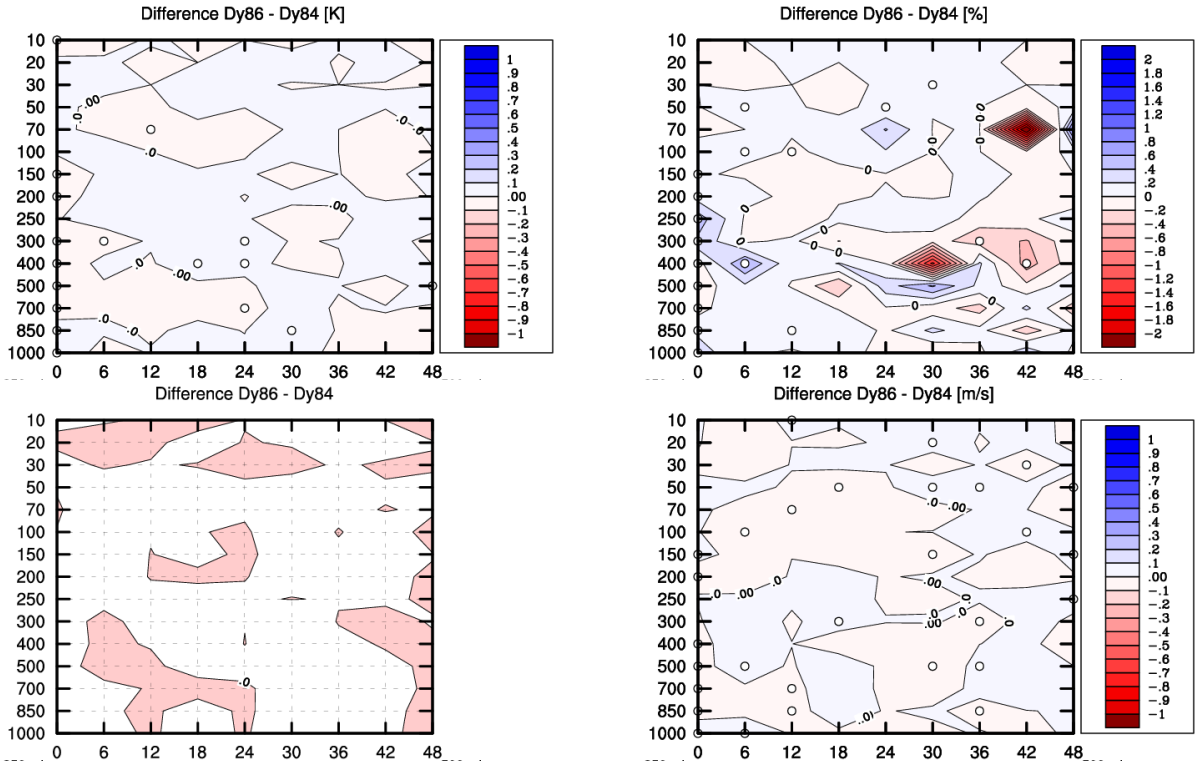

Fig. 10: RMSE differences of T (upper-left), RH (upper-right), ϕ (bottom-left) and wind speed (bottom-right) for 1-14 Feb 2013 of 12 UTC runs. Red areas denote positive impact of the usage of ENS 12UTC B matrix. The white circles points that RMSE difference is better/worse with significance 95 % two-side confidence interval.

The correlation between the impact on analysis and the background standard deviations was noticed, e.g. the higher the background stde of temperature the bigger positive impact, see next Figure. This might hamper a fair experimental evaluation.

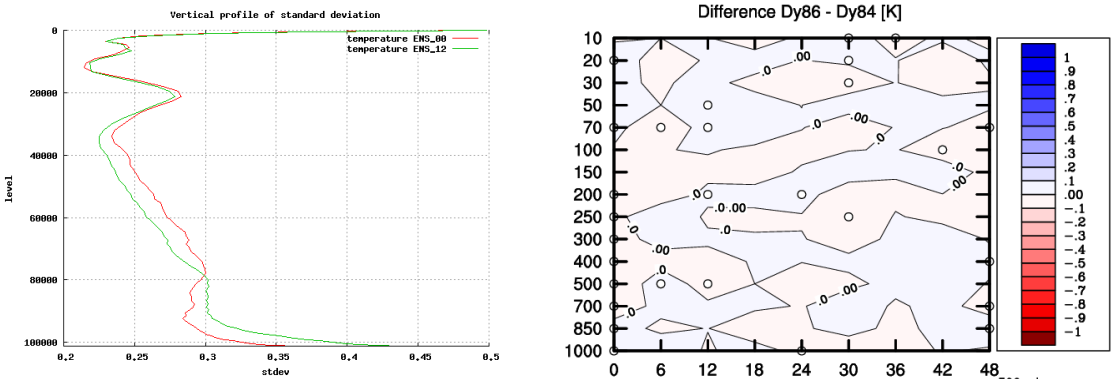

Fig. 11: Vertical profile of temperature background error standard deviation (left), the RMSE differences of T for 3DVAR ENS_00 with respect to 3DVAR ENS_12 for 00UTC runs (right).

4.2 Scores against ARPEGE analyses

The scores against ARPEGE analyses were mostly neutral. Only very small signal was found in relative humidity, for **both 00 UTC and 12 UTC run** ENS_00 performed slightly better at 400 hPa while ENS_12 at 700-850 hPa.

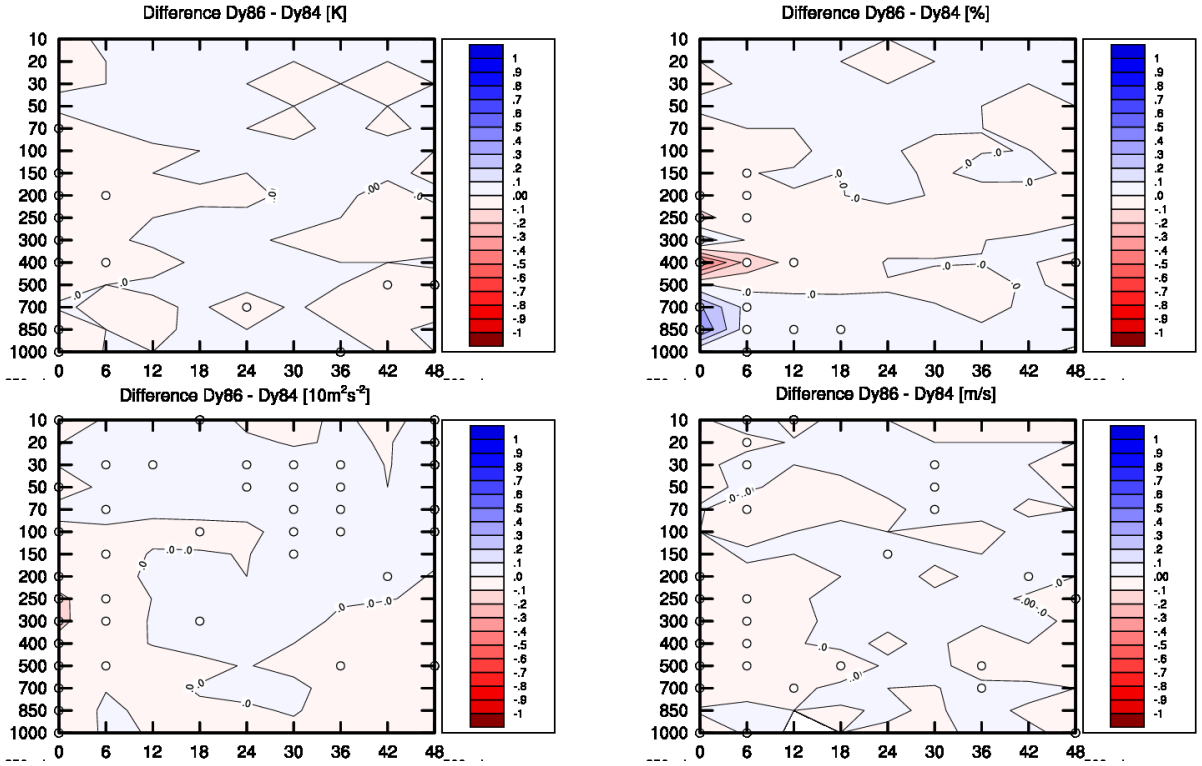

Fig. 12: RMSE differences of T (upper-left), RH (upper-right), ϕ (bottom-left) and wind speed (bottom-right) for 1-14 Feb 2013 of **00 UTC runs**. Red areas denote better performance of ENS_00 over ENS_12.

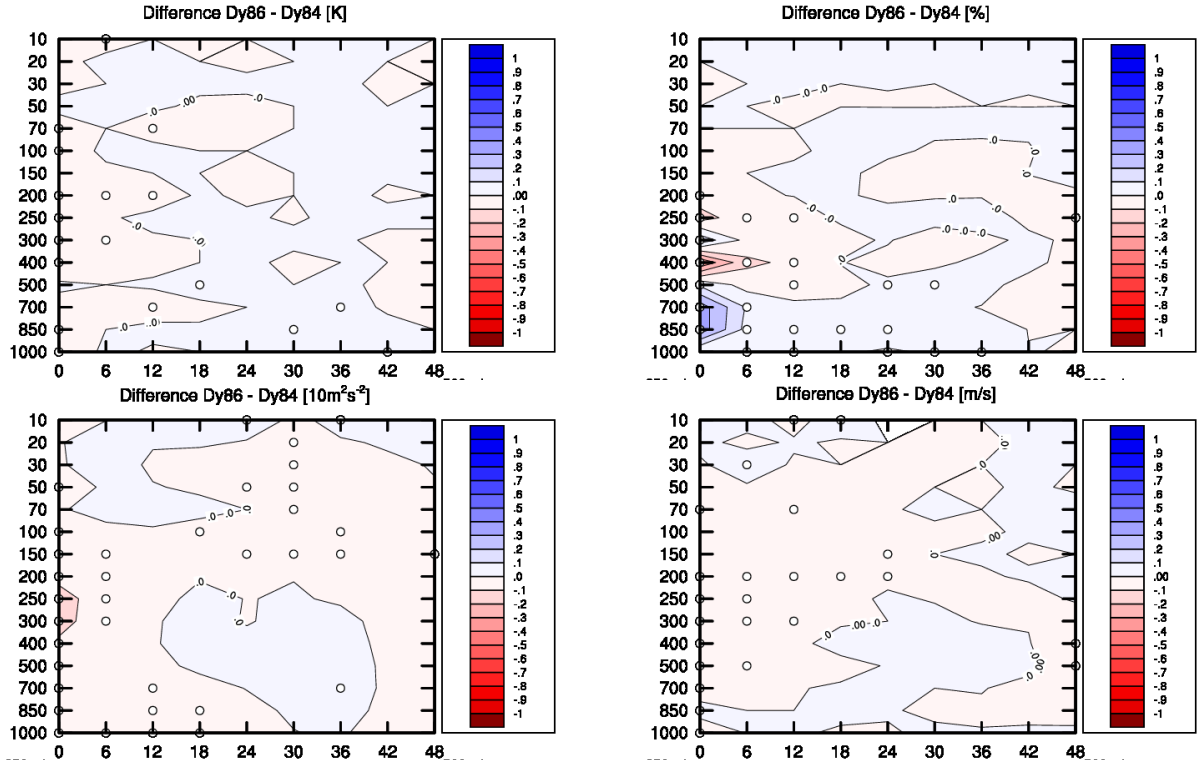

Fig. 13: RMSE differences of T (upper-left), RH (upper-right), ϕ (bottom-left) and wind speed (bottom-right) for 1-14 Feb 2013 of **12 UTC runs**. Red areas denote better performance of ENS_00 over ENS_12.

4.3 Scores against ECMWF analyses

The scores against ECMWF analyses were qualitatively the same as ARPEGE ones.

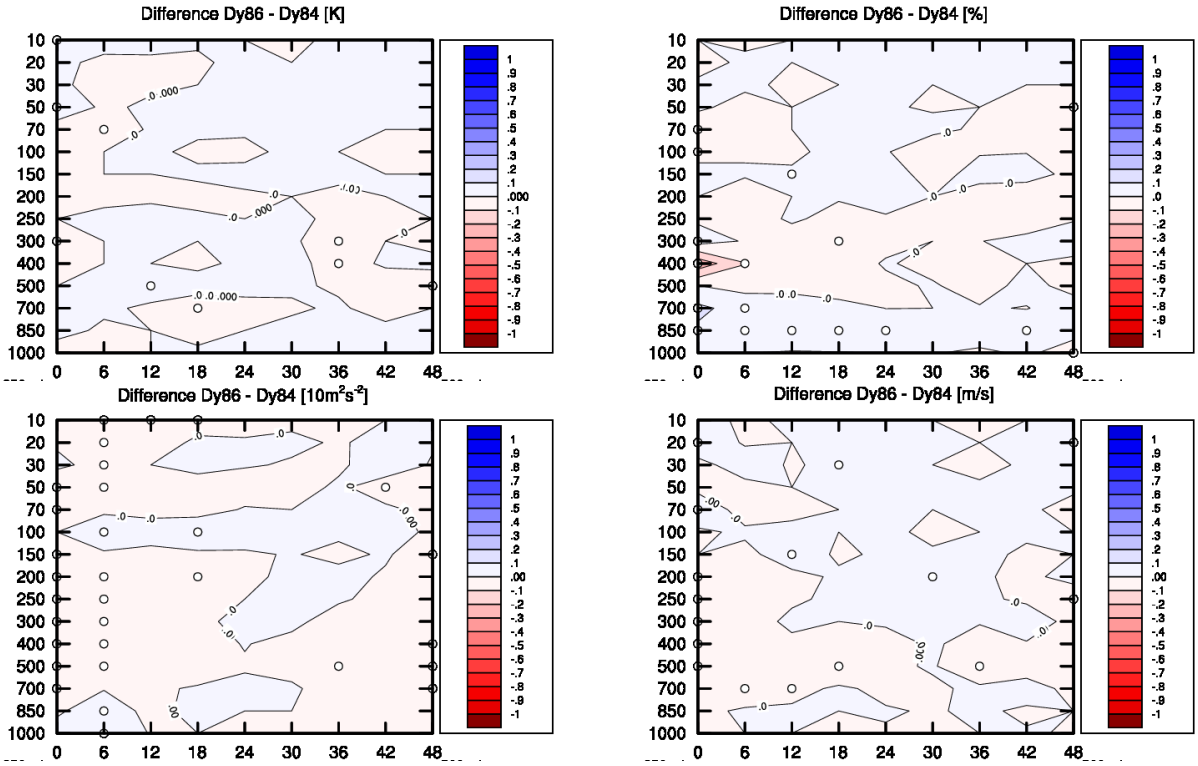

Fig. 14: RMSE differences of T (upper-left), RH (upper-right), ϕ (bottom-left) and wind speed (bottom-right) for 1-14 Feb 2013 of **00 UTC runs**. Red areas denote better performance of ENS_00 over ENS_12.

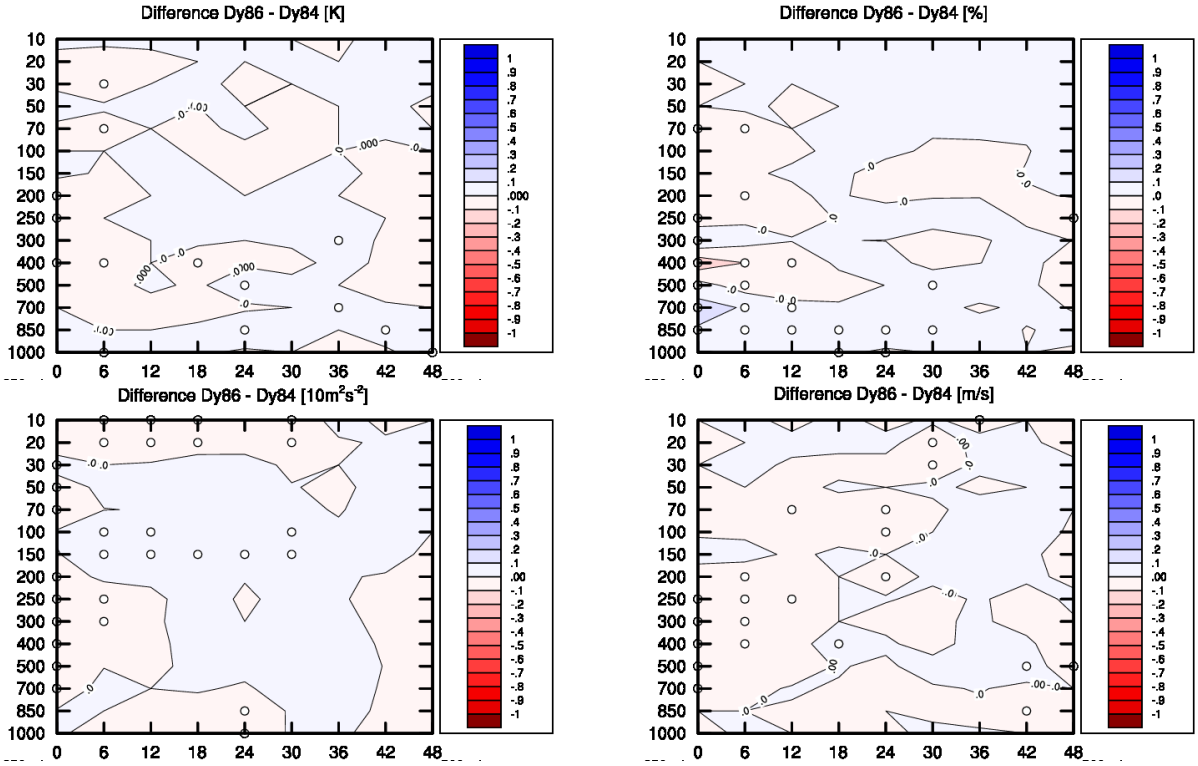

Fig. 15: RMSE differences of T (upper-left), RH (upper-right), ϕ (bottom-left) and wind speed (bottom-right) for 1-14 Feb 2013 of **12 UTC runs**. Red areas denote better performance of ENS_00 over ENS_12.

Summary: Based on performed comparisons it was not possible to make any conclusion about better performance of ENS_00 or ENS_12 for the 3DVAR baseline setup. Considering quantitative differences of various ensemble matrices the experimental evaluation as presented in this section was not found beneficial.

5 Error diagnostics and tuning

Background and observations contribute to analysis proportionally to their errors. The errors are not perfectly known and their determination is an important task for data assimilation research and development.

Desroziers et al (2005) proposed a posteriori diagnostics that are supposed to represent the real standard deviations of observation and background errors in the given data assimilation system. The diagnosed values divided by the predefined ones define the ratio $r = \frac{\sigma_{diagnosed}}{\sigma_{predefined}}$ to be used for tuning of analysis.

The package for the computation of the diagnostics was kindly provided by Bölöni (2010). In ALADIN 3DVAR the tuning was done by multiplying REDNMC and SIGMAO_COEF namelist parameters via the mean ratios. It was found that SIGMAO_COEF implementation in BATOR (CY36T1) is not applied to SYNOP geopotential, T2m, RH2m and to TEMP geopotential, humidity, T2m and RH2m. This was corrected and the modified BATOR was used in following experiments.

5.1 3DVAR ENS_all

The error diagnostics of 3DVAR ENS_all were studied for the testing period. The decision on used background errors was ad hoc. The diagnostics showed that the background errors were underestimated, except for kinetic energy, while the observation errors are overestimated, see Table 1.

Exp	ENS_all		
Var	cases	r_o	r_b
q	10321	0.61941	1.21313
T	17424	0.84735	1.53390
Ek	17655	0.75862	0.78490
Mean	45400	0.76589	1.21538

Table 1: The ratios of diagnosed/predefined standard deviations for observations r_o and background r_b

The diagnostic and the new experiment with retuned REDNMC and SIGMAO_COEF (iteration) were performed three times to obtain mean tuning ratio close to one. The diagnostics results were summarized in Table 2.

Exp	Y88			Y89			Y90		
Var	cases	r_o	r_b	cases	r_o	r_b	cases	r_o	r_b
q	10327	0.67346	1.11380	10327	0.67063	0.94799	10327	0.64998	0.85763
T	17428	1.00326	1.62716	17429	1.06875	1.54322	17430	1.07456	1.45898
Ek	17657	0.88722	0.78246	17660	0.94684	0.70839	17660	0.95286	0.65623
Mean	45412	0.89190	1.23946	45416	0.94298	1.14605	45417	0.94459	1.07313

Table 2: The ratios of diagnosed/predefined standard deviations for observations r_o and background r_b

The list of the error tuning experiments follows:

- **Y81** - 3DVAR **ENS_all**

- **Y88** - 3DVAR **ENS_all** REDNMC=1.2 SIGMAO_COEF=0.8 (iteration 1)
- **Y89** - 3DVAR **ENS_all** REDNMC=1.5 SIGMAO_COEF=0.7 (iteration 2)
- **Y90** - 3DVAR **ENS_all** REDNMC=1.7 SIGMAO_COEF=0.67 (iteration 3)

5.2 Scores against observation

The impact on the forecast was evaluated for all three iterations. Better fit to observations was obtained mostly for analysis.

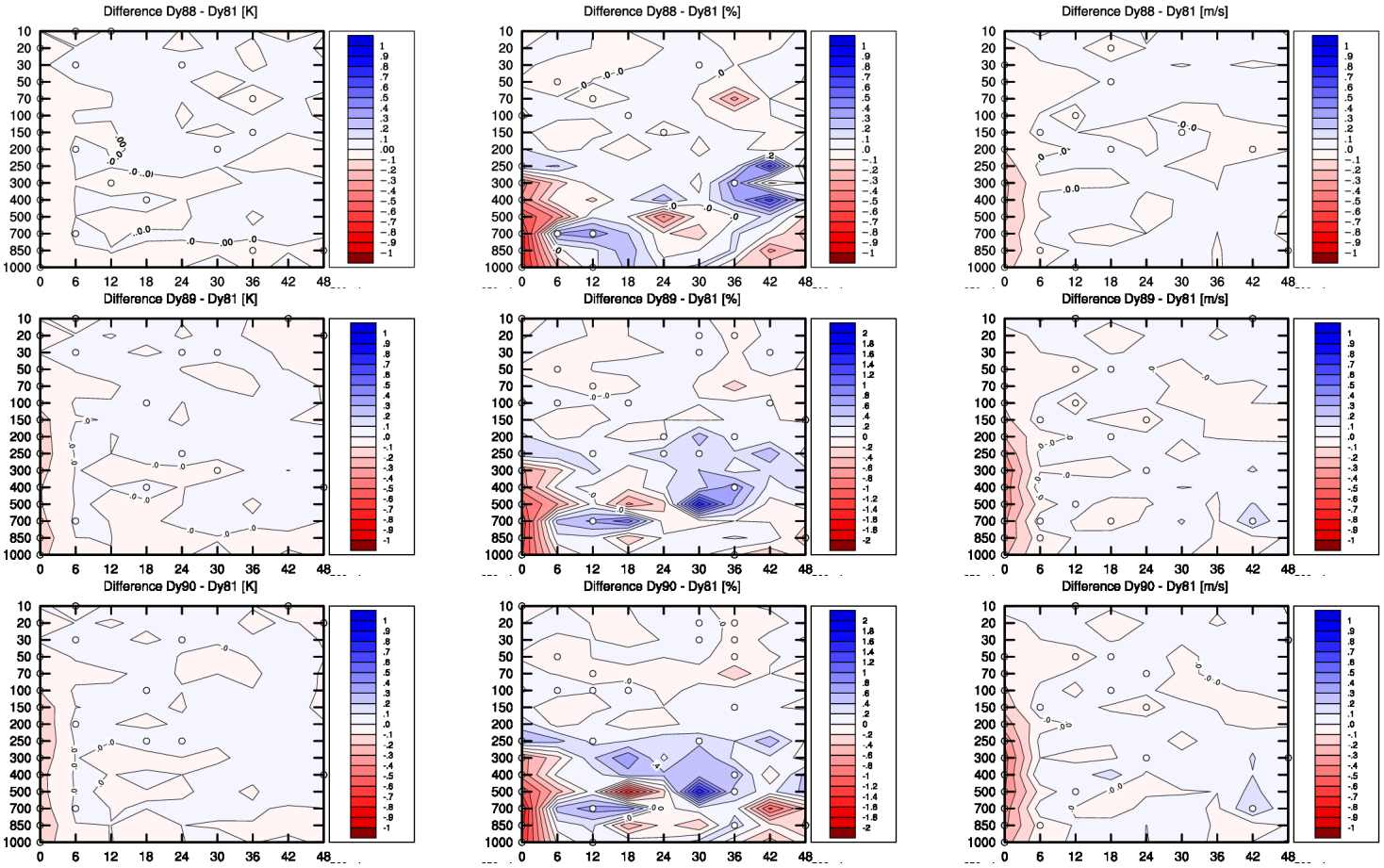

Fig. 16: RMSE differences of T (left), RH (middle), wind speed (right) for 1-14 Feb 2013 of 00 UTC runs. Red areas denote positive impact of the usage of the tuning iterations.

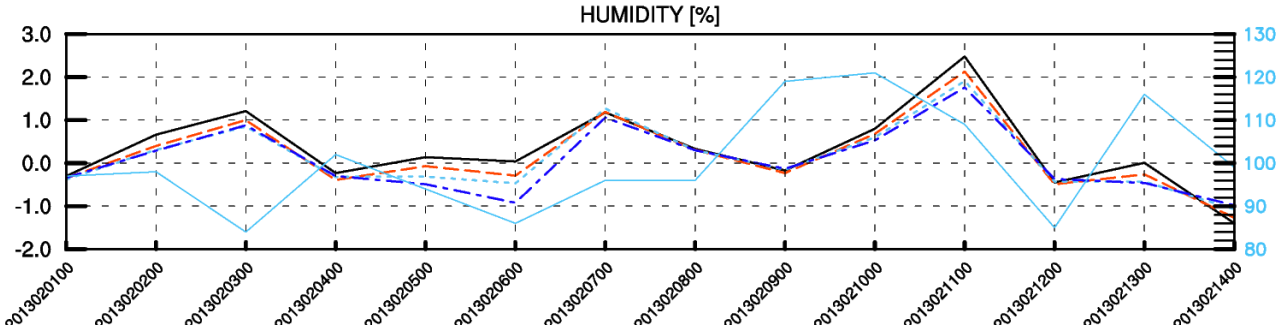

Fig. 17: The time evolution of *RH700* BIAS at analysis time **Y81** in black, **Y88** in red, **Y89** in light dashed blue and **Y90** in dark blue color. The number of TEMP observation used for verification in solid blue.

Summary: The diagnostics showed that the ensemble based background errors are underestimated, except for kinetic energy, while the observation errors are overestimated. There were quantitative differences for each variable and the use of mean value is thus questionable. The tuning of errors proved potential to improve analysis mostly.

6 Baseline setup

The experiment **Y90** - 3DVAR ENS_all with REDNMC=1.7 and SIGMAO_COEF=0.67 was considered as the baseline setup and the results were evaluated with respect to the dynamical adaptation.

6.1 Scores against observation

The comparison against observation showed clear positive impact for analysis and also for longer forecast ranges of some parameters, e.g. *T850* and *RH500* up +18H.

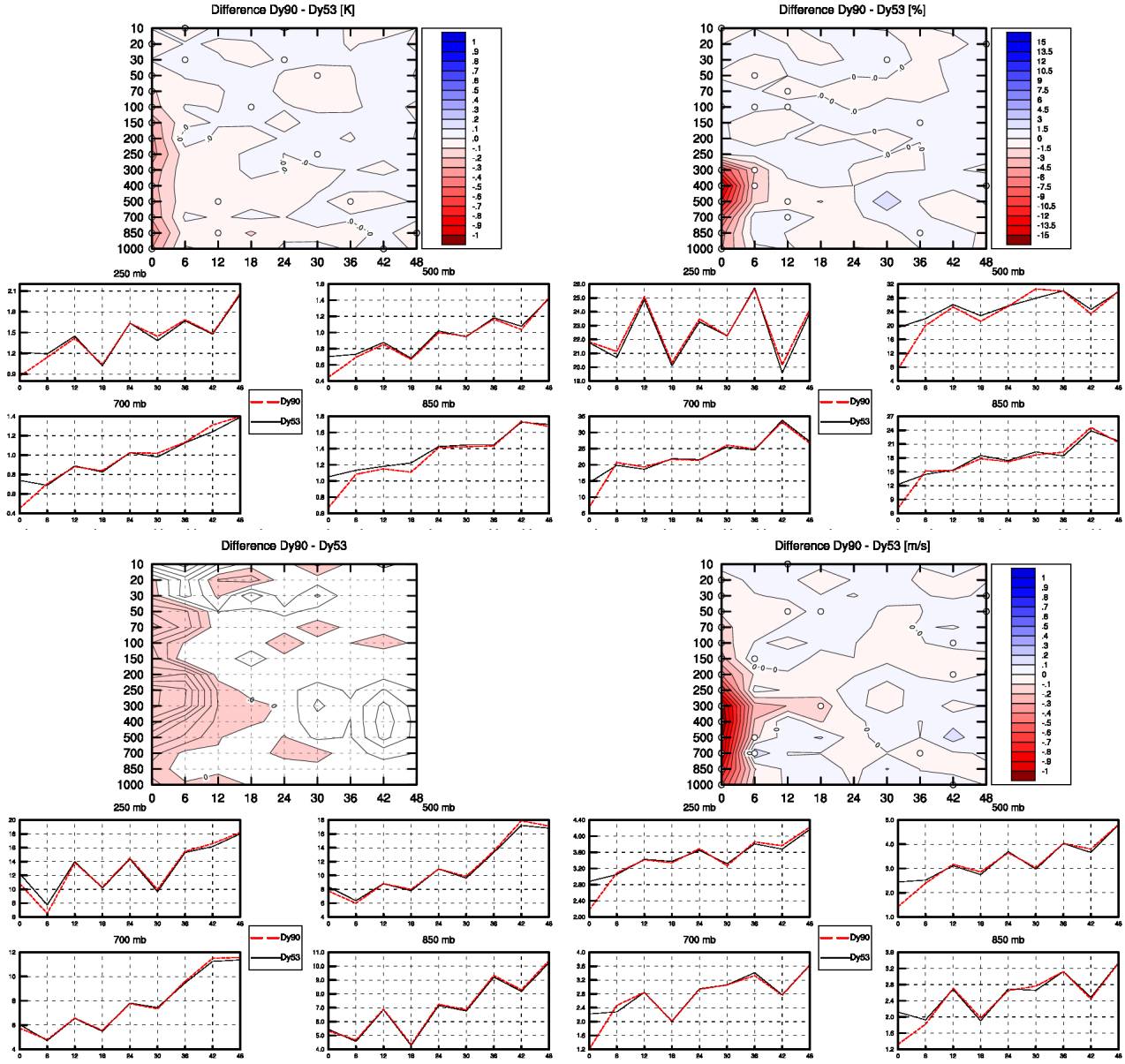

Fig. 22: RMSE differences of T (upper-left), RH (upper-right), ϕ (bottom-left) and wind speed (bottom-right) for 1-14 Feb 2013 of 00 UTC runs. Red areas denote positive impact of the usage of the **Y90** vs dynamical adaptation. The white circles points that RMSE difference is better/worse with significance 95 % two-side confidence interval.

The BIAS scores were mostly improved, the most for *RH500* by more than 10 %, for other variables in a lesser extent. Qualitatively the 3DVAR implied slight cooling below 700 hPa and upward of 200 hPa and warming around 500 hPa, for humidity prevails intensive drying between 700 hPa and 300 hPa.

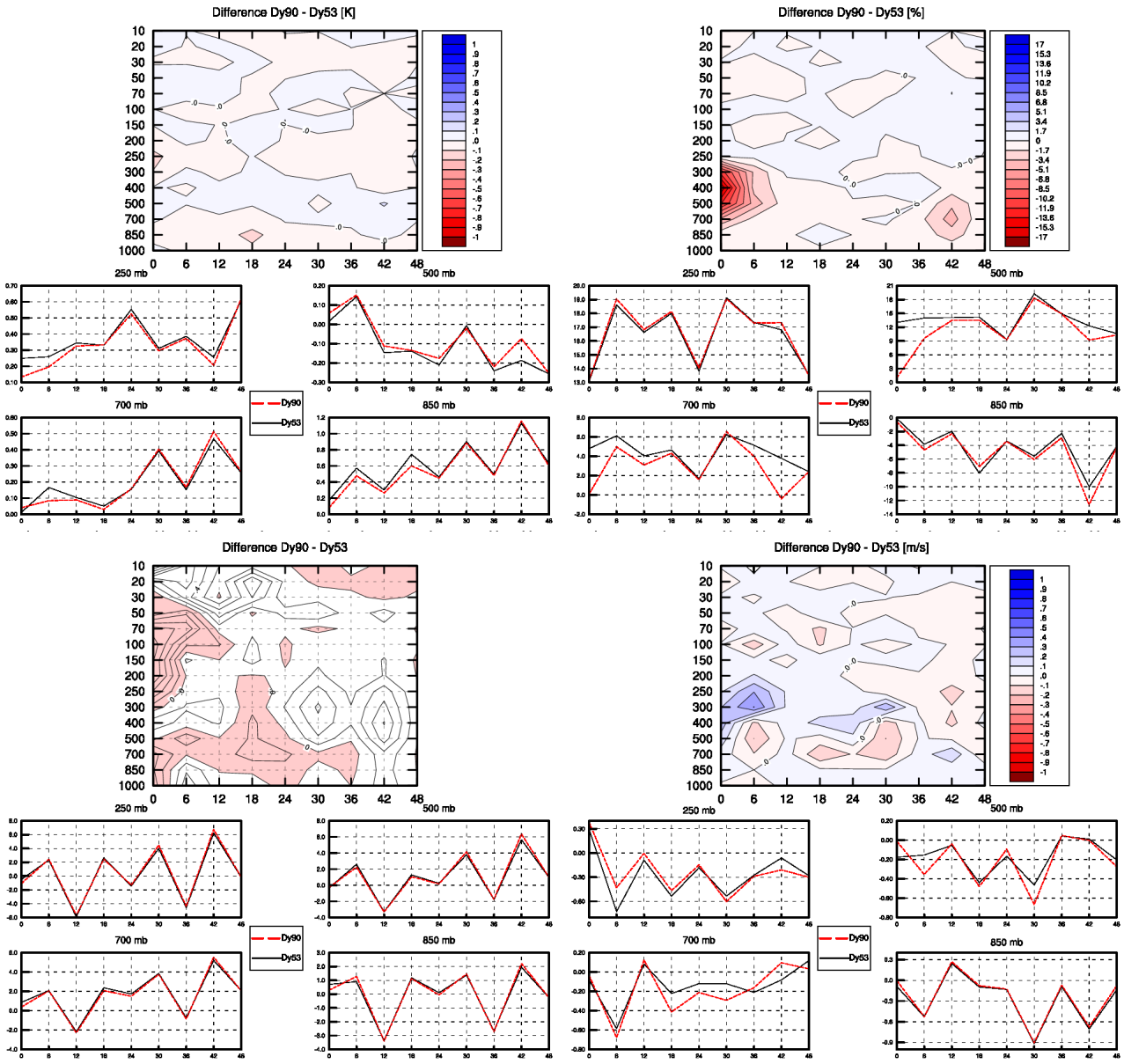

Fig. 23: BIAS differences of T (upper-left), RH (upper-right), ϕ (bottom-left) and wind speed (bottom-right) for 1-14 Feb 2013 of 00 UTC runs. Blue/red areas denote positive/negative differences of the BIAS scores.

The screen level parameters showed a small improvement for analysis time, except cloudiness.

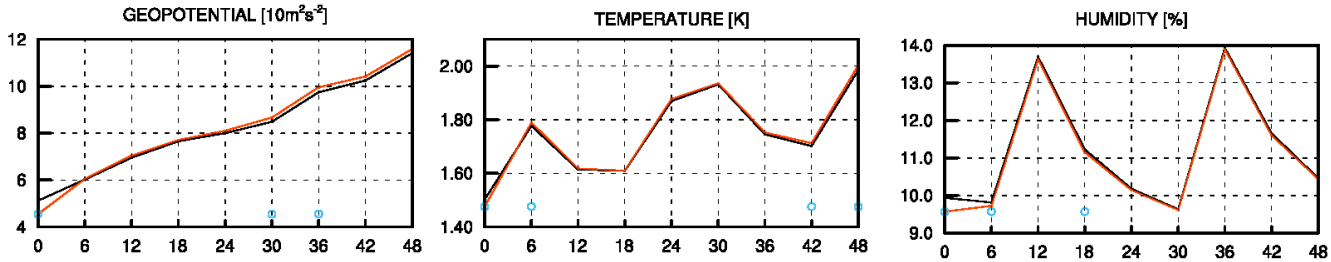

Fig. 24: RMSE of $MSLP$ (left), $T2m$ (middle) and $RH2m$ (right) for 1-14 Feb 2013 of 00 UTC runs. The experiment **Y90** red and **Y53** in black. The circles points that RMSE difference is better/worse with significance 95 % two-side confidence interval.

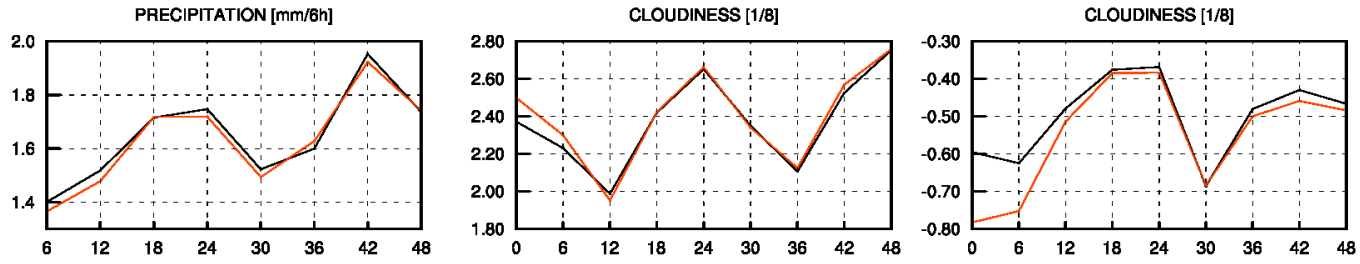

Fig. 25: RMSE of accumulated 6h precipitation (left), cloudiness (middle) and BIAS for cloudiness (right) for 1-14 Feb 2013 of 00 UTC runs. The experiment **Y90** in red and **Y53** in black.

6.2 Scores against ARPEGE analyses

The RMSE scores against ARPEGE analyses showed overall degradation for analysis, +6H and many parameters at +12H forecasts. The dynamical adaptation as reference posed very hard limit for comparison. The distance of the 3DVAR experiment increased with respect to ARPEGE.

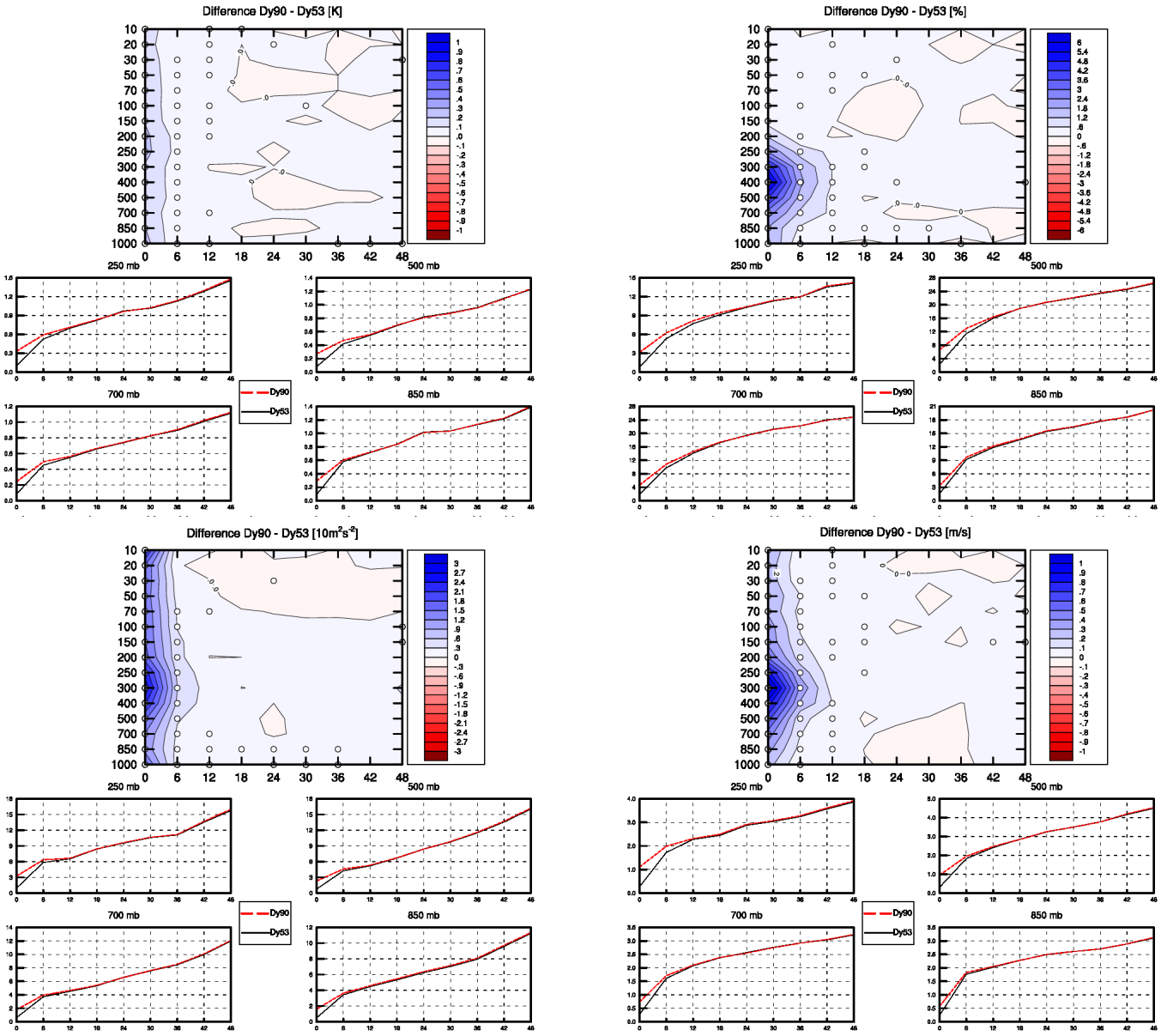

Fig. 1: RMSE differences of T (upper-left), RH (upper-right), ϕ (bottom-left) and wind speed (bottom-right) for 1-14 February 2013 of 00 UTC runs. Red areas denote positive impact of the **Y90** vs dynamical adaptation. The white circles points that RMSE difference is better/worse with significance 95 % two-side confidence interval.

The BIAS scores against ARPEGE analyses of the 3DVAR baseline experiment **Y90** showed slight warming between 700 - 400 hPa and cooling at 300-100 hPa, for humidity there was clear drying up to 250 hPa and moistening above up to 100 hPa. The wind speed BIAS was slightly larger, while for the geopotential slightly smaller values were found at analysis, except level 400 hPa.

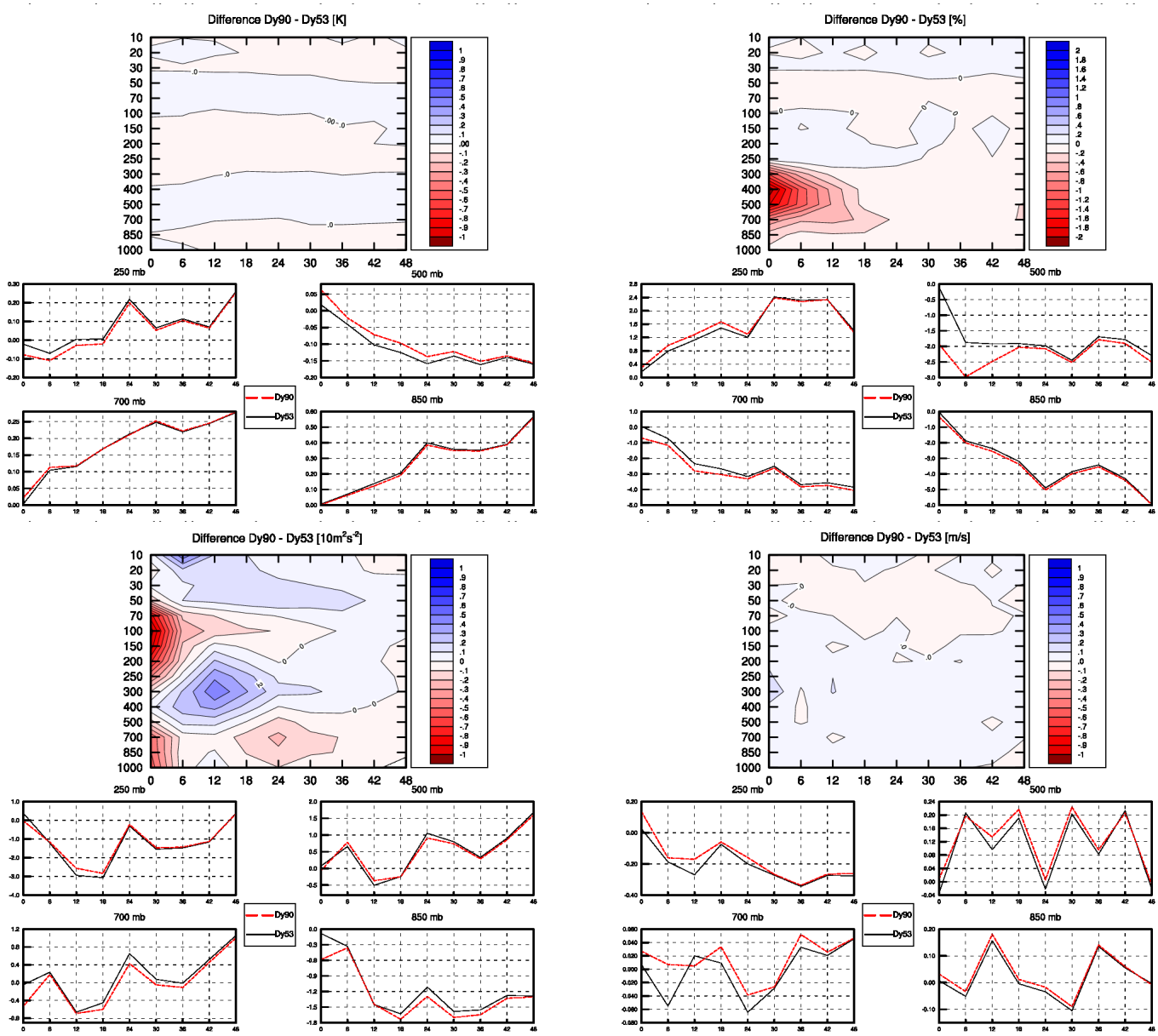

Fig. 2: BIAS differences of T (upper-left), RH (upper-right), ϕ (bottom-left) and wind speed (bottom-right) for 1-14 February 2013 of 00 UTC runs. Blue/red areas denote positive/negative differences of the BIAS scores.

6.3 Scores against ECMWF analyses

The RMSE scores against ECMWF analyses showed smaller differences than in the comparison against ARPEGE.

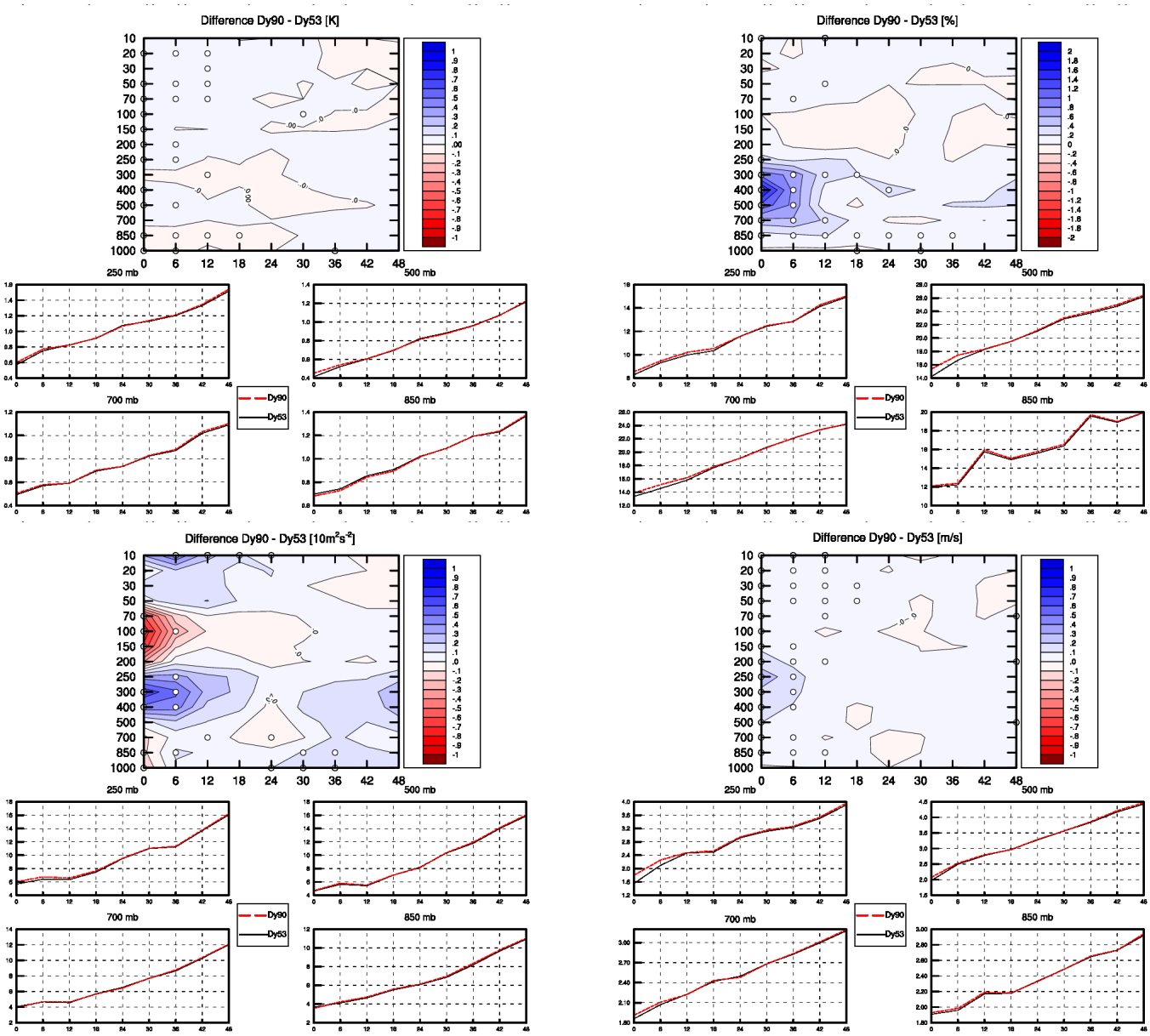

Fig. 1: RMSE differences of T (upper-left), RH (upper-right), ϕ (bottom-left) and wind speed (bottom-right) for 1-14 February 2013 of 00 UTC runs. Red areas denote positive impact of the **Y90** vs dynamical adaptation. The white circles points that RMSE difference is better/worse with significance 95 % two-side confidence interval.

The BIAS scores against ECMWF analyses were qualitatively the same as against ARPEGE.

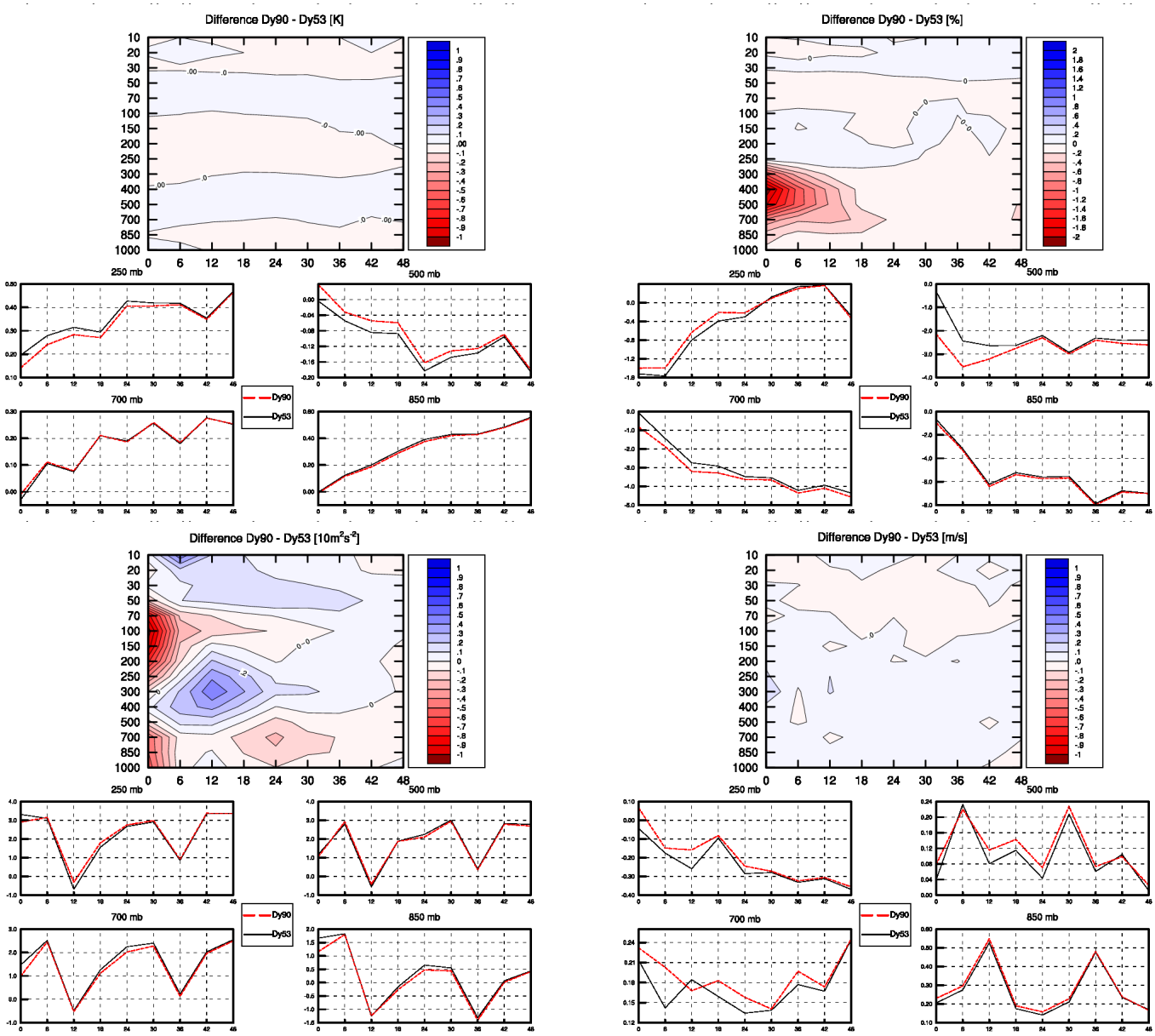

Fig. 2: BIAS differences of T (upper-left), RH (upper-right), ϕ (bottom-left) and wind speed (bottom-right) for 1-14 February 2013 of 00 UTC runs. Blue/red areas denote positive/negative differences of the BIAS scores.

Summary: The RMSE scores against analyses were mostly degraded, but the scores against observations showed improvement (better fit to assimilated observations). This discrepancy should be better understood. The BIAS scores (against observations and both analyses) showed the consistent signal: drying up to 250 hPa, slight cooling around 850 hPa and 200 hPa and warming between 700-400 hPa.

7 Conclusions

The studies dedicated to the 3DVAR baseline setup were presented. Performed experiments did not bring an indication for optimal background error sampling. The bug was identified in the SIGMAO_COEF implementation in BATOR and with the correction the tuning of the errors showed potential to improve analysis mostly. The encouraging baseline setup results were obtained and will be used in further tests with more observation types.

Acknowledgment
The author wishes to thank Antonín Bučánek who kindly provided all background errors used in this study.

References

Berre, L., S.E., Stefanescu, and M. Belo Pereira, 2006: The representation of the analysis effect in three error simulation techniques. *Tellus*, **58A**, 196-209.
Brožkova, R., and Coauthors, 2001: DFI blending: An alternative tool for preparation of the initial conditions for LAM. *Research activities in atmospheric and oceanic modelling, WMO CAS/JSC WGNE Rep. 31*, 1.71.8 Bölöni, G., 2010: A posteriori diagnosis (tuning) of B and R variances, *RC LACE Forum*
available online http://www.rclace.eu/forum/viewtopic.php?f=30&t=248
Desroziers, G.,L. Berre, B. Chapnik, and P. Poli, 2005: Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteor. Soc.*, **131**, 33853396.
Fischer, C., T. Montmerle, L. Berre, L. Auger, S. Stefanescu, 2005: An overview of the variational assimilation in the ALADIN/France numerical weather-prediction system. *Quart. J. Roy. Meteor. Soc.*, **131**, 34773492.