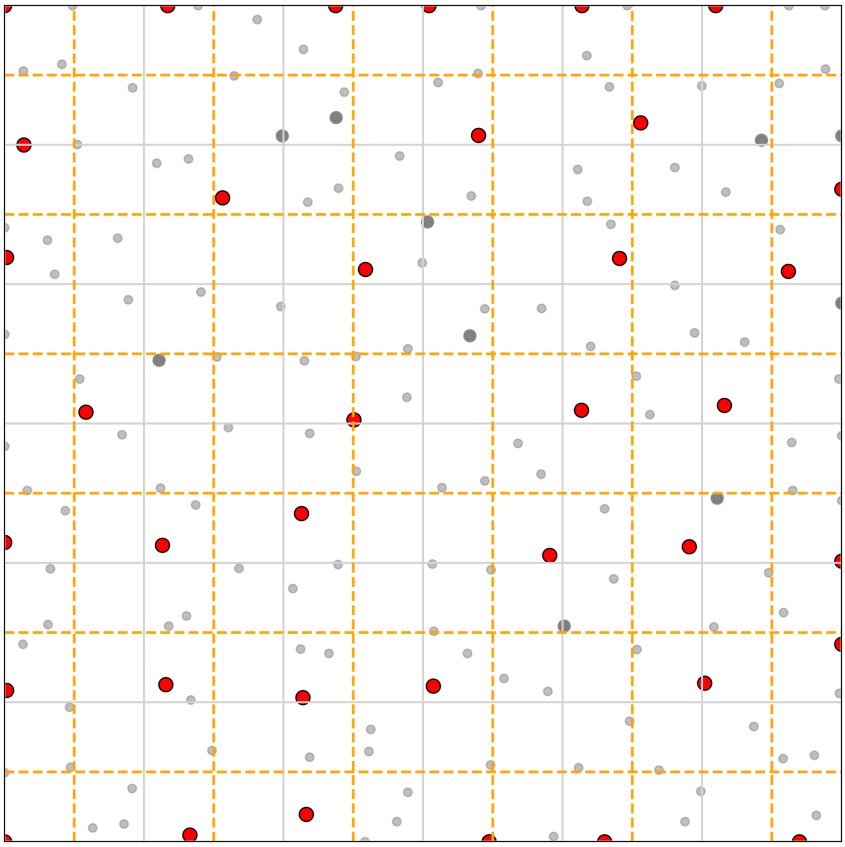
Regional Cooperation for
Limited Area Modeling in Central Europe

LACE
nwp central europe

Thinning of radar reflectivity observations in ALARO



Author:	Laura Gréta Magyar (HungaroMet)
Host:	Antonín Bučánek (CHMI)
Purpose:	Report from RC LACE stay
Place:	CHMI, Czech Republic, Prague
Date(s):	3. 11. – 28. 11. 2025

Contents

1 Introduction

The assimilation of high-density radar reflectivity observations offers great potential for improving convective-scale numerical weather prediction, but it also poses significant challenges for data assimilation systems. In this article, we present work carried out during an ACCORD/LACE stay at the Czech Hydrometeorological Institute (CHMI), focusing on optimizing the thinning distance for radar reflectivity assimilation in ALARO cy46. Using an existing method based on observation-error correlations, we first derived a theoretically optimal thinning distance and then conducted sensitivity experiments to assess the impact of different thinning choices on verification scores. Proper tuning of the thinning distance is particularly important in the 3D-Var implementation of the ACCORD CSC, where we expect that two observations do not have correlated errors, and therefore the observation-error covariance matrix R is modelled as diagonal. An appropriate thinning strategy helps ensure that this assumption is respected, prevents the introduction of spurious correlations, and supports a stable, efficient assimilation that translates into measurable forecast improvements.

1.1 Thinning procedure for reflectivity in ALARO

In ALARO cycle 46, radar reflectivity data are assimilated using the so-called "1D+3D-Var" approach, as described by Wattrelot et al. (2014) [10]. In this framework, radar reflectivity is not assimilated directly; instead, pseudo-observed relative humidity columns are derived from the observed reflectivity columns and are used in the 3D-Var system. Figure 1 illustrates the preprocessing steps applied prior to the 3D-Var assimilation in our configuration.

During the assimilation cycle, radar observations can be filtered at two stages. The first opportunity arises immediately after reading the reflectivity data from the HDF5 files, during their conversion from polar to Cartesian coordinates within the BATOR module. At this step, the effective sampling distance can be adjusted using the HODIM%Sample namelist parameter.

A second filtering step is applied in the screening module following the Bayesian inversion. At this stage, a horizontal thinning of the relative humidity pseudo-observations is performed by defining a square grid using the RMIND_RADAR namelist parameter. Only one column per grid cell is retained, with preference given to columns located closer to the radar and those containing a larger number of elevation angles. This process is then repeated using an arbitrarily chosen grid size (RFIND_RADAR) and an optional horizontal shift (XYSHIFT_THIBOX) relative to the initial grid. Figure 2 visualizes this two-stage thinning process by illustrating the two thinning grids and their relative displacement.

Figure 1 highlights a third option for reducing the density of radar observations. This procedure, known as superobbing [3], merges closely spaced observation points into a single representative super-observation whose value best reflects the characteristics of the combined measurements. The HOOF tool (developed by Peter Smerkol [9]), used through the BATOR module to homogenize OPERA HDF5 ODIM files for assimilation in ALARO, includes the capability to perform superobbing. Comparing this method with the “traditional” filtering approach described above would be highly informative, but such an analysis lies beyond the scope of the present work.

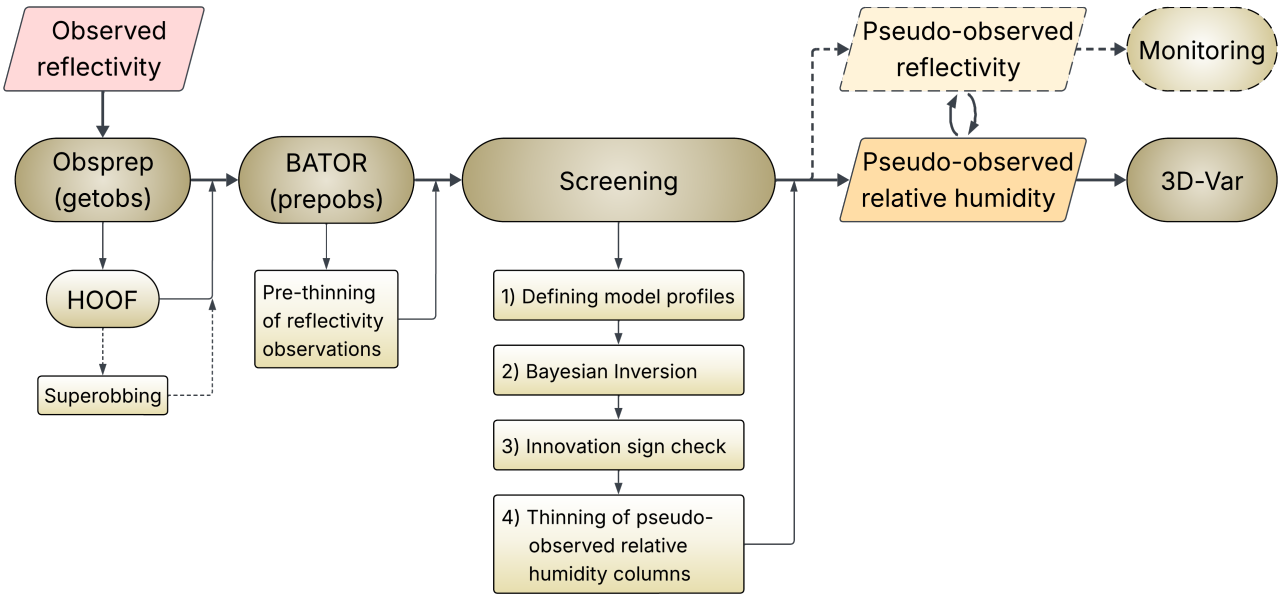


Figure 1: Flowchart of preprocessing steps for radar reflectivity observations using Bayesian inversion prior to 3D-Var assimilation

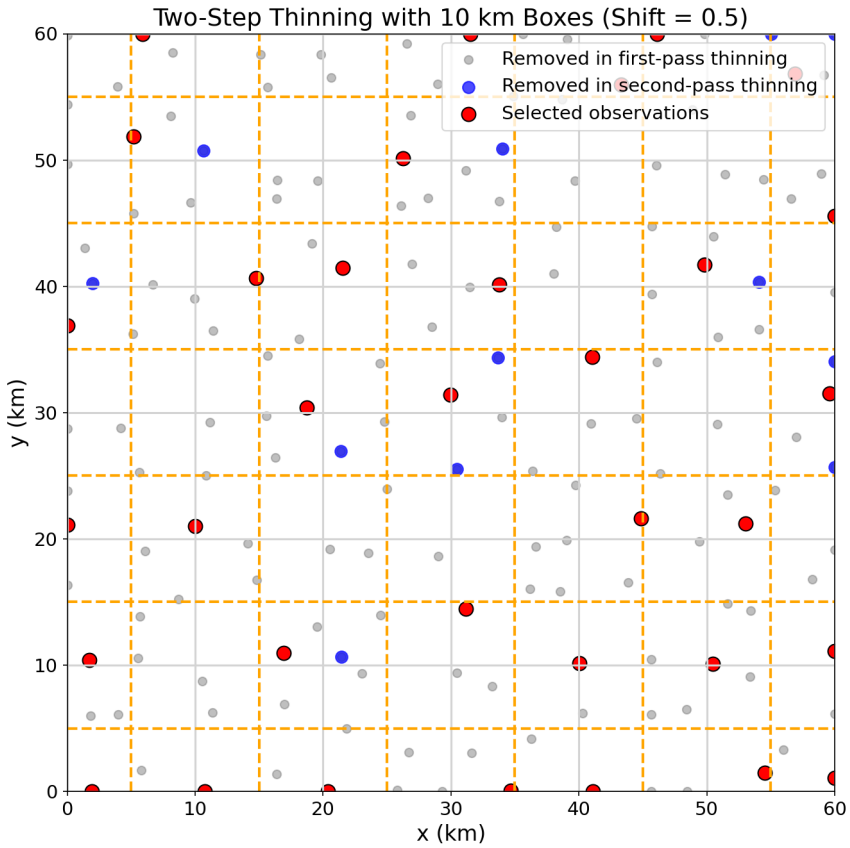


Figure 2: Illustration of the two-stage horizontal thinning process applied to relative humidity pseudo-observations (note: the figure is based on synthetic data and does not use real observations)

2 Methodology

2.1 Observation-error correlations

To select a suitable thinning distance, we adopted the criterion of Liu and Rabier (2003) [5], according to which observational sampling becomes counterproductive when error correlations between adjacent observations rise above about 0.2, leading to reduced analysis and forecast quality.

To compute the observation-error correlations, we used the ObsTool software (written in R) developed by Patrik Benáček at CHMI, which implements the method described by Desroziers et al. (2005) [2]. The tool uses first-guess and analysis departures to diagnose the error correlations of the observation-error covariance matrix R in observation space. For each observation type, pairs of first-guess and analysis departures are selected within a defined time window and subsequently grouped according to their spatial separation using predefined distance bins. This allows the observation-error correlations to be estimated as a function of distance. The results obtained from running ObsTool for our experimental period are reported in Section 2.3.

2.2 Operational model setup

The model configuration used in this study, along with the operational namelist settings relevant to thinning, is outlined below.

- **Model:** ALARO NH-v1B
- **Local CHMI cycle:** 46t1as_op1
- **Domain (ALARO/CZ):** Δx $\approx$ 2.3 km; 1069x853 grid points; 87 vertical levels
- **Coupling:** 3h space consistent coupling from ARPEGE; synchronous
- **B-matrix:** static EDA
- **Assimilation & initialization:** 3h assimilation cycle; BlendVar (digital filter spectral blending + 3D-Var for upper-air) + OI (for surface), incremental DFI in short cut-off production (no DFI in long cut-off 3h cycle)
- **Assimilated observations:** SYNOP + AWS(soil), AMDAR/MRAR/EHS, AMV/HR, PROFILER, TEMP, ASCAT, SEVIRI, OPERA reflectivity

---

- **Thinning-related namelist settings:**
  - HODIM%Sample = 5000 (m)
  - RMIND_RADAR = 16275 (m)
  - RFIND_RADAR = 16275 (m)
  - XYSHIFT_THIBOX(13) = 0.5

The value RMIND_RADAR = 16275 was chosen using the $7\Delta x$ criterion, where Δx corresponds to the 2325 m horizontal grid spacing of the operational ALARO/CZ model. This choice reflects the principle that assimilating scales smaller than the model's effective resolution provides no meaningful benefit. The concept follows Skamarock (2004) [8], who identified the effective resolution of the WRF model to be approximately $7\Delta x$.

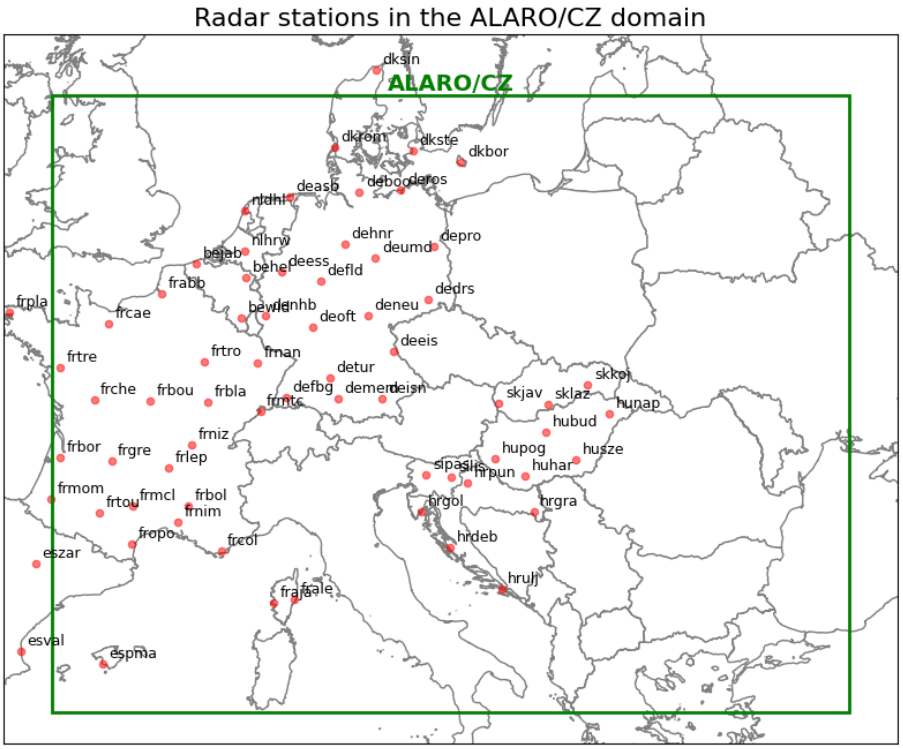


Figure 3: Radar stations used in the experiments. The stations are indicated by their ODIM code [4], and the approximate ALARO/CZ domain is marked by the green rectangle.

Operational assimilation of OPERA radar reflectivity data has been implemented in the ALARO/CZ model since May 2025. All radars included in the experiments discussed in the following sections are shown in Figure 3.

2.3 Experimental setup

In the following sections, a series of experiments is presented to identify the most appropriate thinning distance for radar reflectivity assimilation. For this purpose, an experimental period from 28 August to 7 September 2025 was selected, during which considerable precipitation was observed over Czechia.

In August, the operational configuration was reverted from the Abel–Boutle raindrop size distribution [1] to the previously used Marshall–Palmer formulation [6] within the radar reflectivity diagnostics and observation operator, and therefore a double-suite was running during our selected period. Because this change has a noticeable impact on the precipitation fields, it was essential to reproduce the double-suite configuration to ensure that our experiments remained comparable with the operational reference. To this end, we prepared a dedicated test experiment with namelist settings identical to those of the operational model. The test confirmed that both the double-suite and the operational run were successfully replicated, validating the experimental setup. All experiments described in the following are therefore based on this replicated configuration, ensuring full consistency with the operational run, which is thus taken as the reference throughout.

We set up a series of experiments with different thinning configurations for the selected period, with their settings summarised in Table 1 alongside the operational setup. Experiments

Table 1: Thinning-related namelist settings used in all experiments and in the operational configuration

	Oper.	a52	a53	a54	a55	a56
HODIM%Sample	5000	5000	2500	5000	5000	15000
RMIND_RADAR	16275	1000	1000	10000	20000	16275
RFIND_RADAR	16275	1000	1000	10000	20000	16275
XYSHIFT_THIBOX(13)	0.5	0	0	0.5	0.5	0.5

a52 and a53 were purposefully configured to obtain the CCMA ODB datasets required to run ObsTool, and therefore, only the 3-hour assimilation cycle was performed in these two cases. To obtain the most reliable estimates of the observation-error correlations, we needed to "turn off" the filtering. This was achieved by substantially reducing the thinning distance; consequently, RMIND_RADAR and RFIND_RADAR were set to 1000 m, and the horizontal shift of the second grid was disabled (XYSHIFT_THIBOX(13) = 0). In a52, the sampling distance in BATOR (HODIM%Sample) was kept at 5000 m, whereas in a53 the aim was to reduce thinning even further to allow as many observations as possible to pass through. We therefore set the BATOR sampling distance to 1000 m, which corresponds to the resolution of the OPERA radar data. However, the experiment exceeded the available memory despite using the maximum number of compute nodes allowed by the system. In practice, all 12 nodes were fully utilised, each equipped with 384 GB of memory, meaning that more than 4.5 TB of total memory was still insufficient to process the resulting number of observations. Consequently, the sampling distance had to be increased to 2500 m to keep the experiment within memory limits. Because the computational cost increased substantially, only four assimilation cycles could be completed, covering the first 12 hours of the experimental period. The upper panels of Figures 4 and 5 display the observation-error correlations of R and the corresponding number of assimilated observations as a function of distance for experiment a52, while the results obtained from the shorter a53 period are shown in the lower panels. A detailed analysis of these plots and the associated results is provided in Section 3.1.

Building on these findings, we conducted a sensitivity study consisting of three additional experiments, a54, a55, and a56, each employing different thinning configurations. These experiments were designed in light of the ObsTool diagnostics, which provided a guideline for a theoretically optimal thinning distance. In contrast to a52 and a53, both the assimilation and the production were run for these experiments, and verification was carried out to assess the impact of the different thinning setups on forecast quality. In this configuration, the system produces a +48-hour forecast at each 00 and 12 UTC. The corresponding results are presented in Section 3.2.

3 Results

3.1 Optimal thinning based on observation-error correlations

What we observed from the ObsTool diagnostics is that extending the time period – and thus increasing the total number of assimilated observations – leads to smoother and generally lower diagnosed correlations. This behaviour became evident when we began running experiment a52: as more assimilation cycles were completed, the diagnosed correlation curves progressively smoothed out and their magnitudes decreased. Additionally, the increased observational density in the higher-resolution setup (a53) also shaped the diagnostics. Figures 4 and 5 show that a53 configuration produces smoother and more internally consistent diagnostics; however, it also causes the observation-error correlations to decrease more rapidly with distance. In a53 setup, the correlations fall below 0.2 at approximately 7–10 km, whereas in the more robust a52 experiment this range shifts to about 10–15 km. Although a longer time period is desirable for improved statistical reliability, these results nonetheless provide a useful indication of a theoretically optimal thinning distance.

3.2 Sensitivity studies

3.2.1 Effects of thinning on observation density and assimilation metrics

To illustrate how different thinning configurations affect the spatial density of observations, we begin by introducing Figure 6, which compares experiments a54, a55, and a56 with the operational reference (see Table 1 for the thinning configurations used in each experiment). The plots show the western part of Czechia and provide a direct visual comparison of how each thinning setting modifies the density and spatial distribution of assimilated observations.

As Figure 6 illustrates, even when a specific thinning distance is prescribed during screening, it is not possible to prevent observations from falling closer than the target spacing. This behaviour follows directly from the characteristics of the thinning procedure described in Section 1.1. Figure 7 further highlights this effect by showing histograms of the minimal intervals between observations for experiments a52, a54, a55, and a56. In the cases of a54 and a55, the largest counts in the distributions cluster around the respective target thinning distances, yet more than half of the observation pairs still appear below these. In contrast, the distribution for experiment a56 exhibits a slightly different pattern, with observation separations shifted more towards higher values.

As a complementary quantitative perspective, Table 2 summarises the number of observations retained after screening, together with the subset of pseudo-observed relative humidity that was actually assimilated, for all experiments. This comparison provides a clearer view of how the different thinning configurations translate into effective observation usage within the assimilation cycle. Table 2 also reports the number of Lanczos iterations reached during the minimisations. Since this iteration count is limited to 50 in ALARO/CZ, it is informative to examine whether any of the experiments approach or reach this threshold. In addition, the table lists the final values of the cost function and the running times of both the screening and the minimisation processes, providing a more complete picture of the computational behaviour across the different thinning configurations.

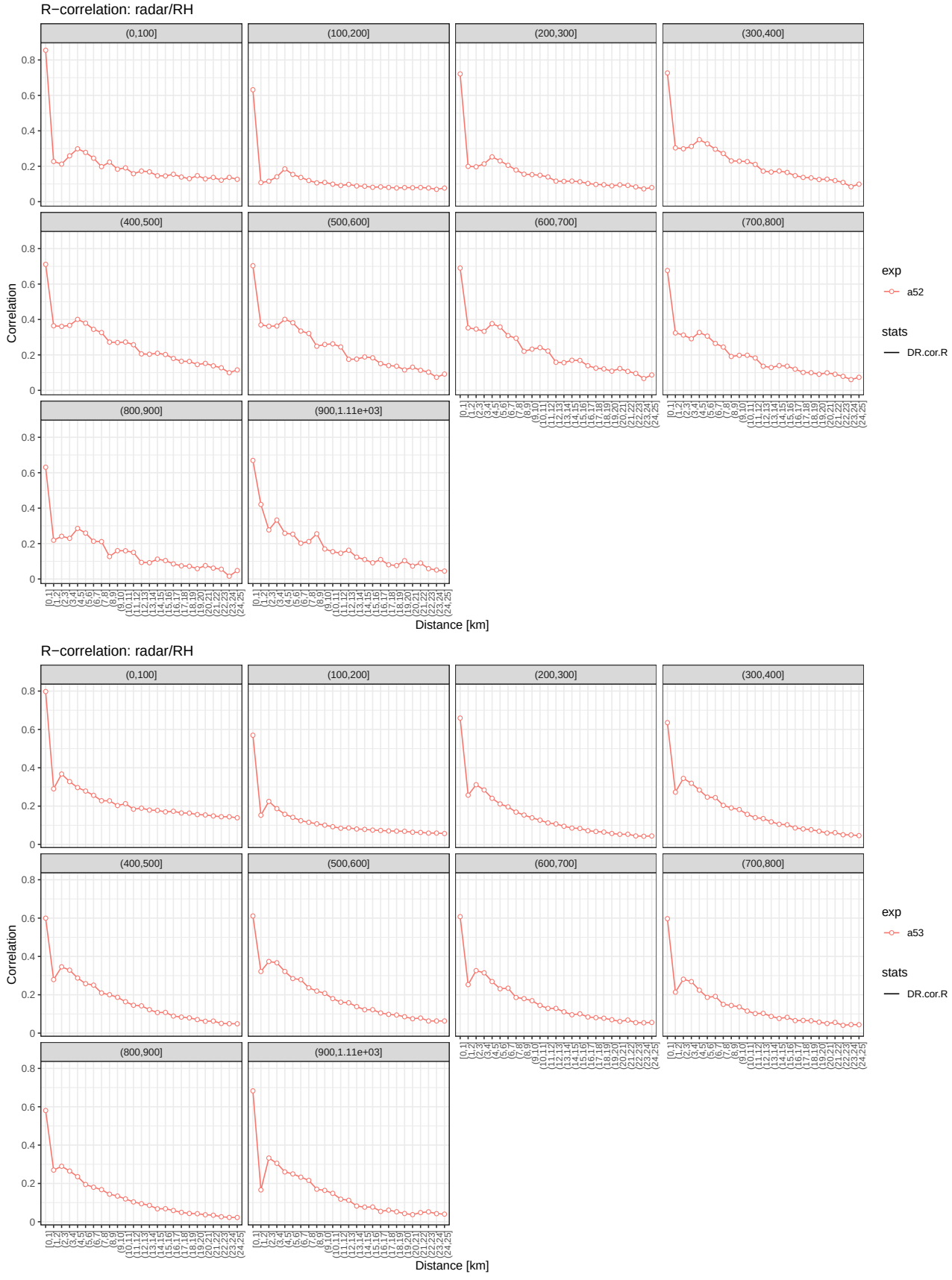


Figure 4: Diagnosed observation-error correlations of R as a function of distance for predefined vertical intervals in the a52 (upper) and a53 (lower) experiments

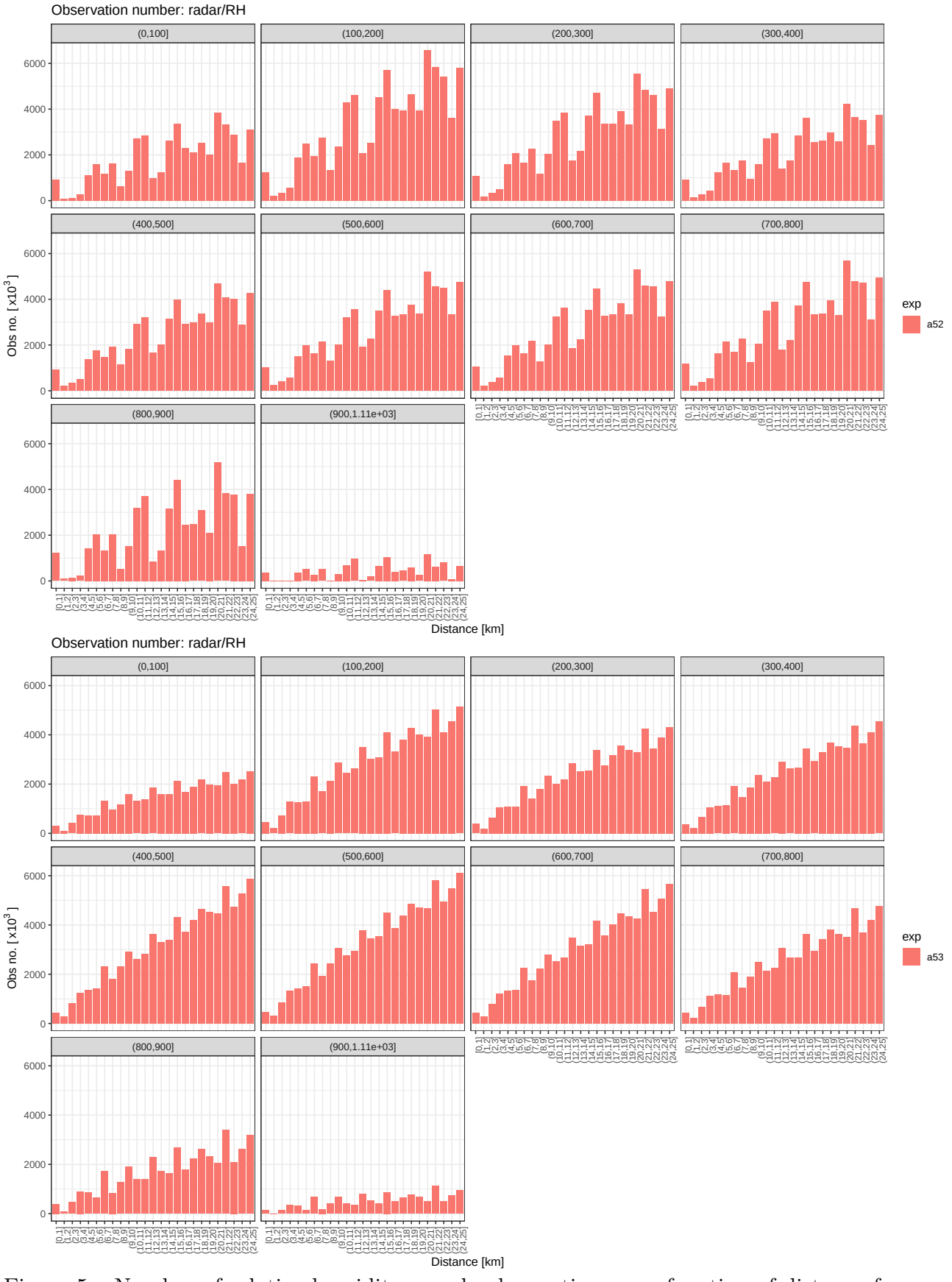


Figure 5: Number of relative humidity pseudo-observations as a function of distance for predefined vertical intervals in the a52 (upper) and a53 (lower) experiments used in the ObsTool diagnostics. (Note that experiment a52 covers a 10-day period (80 assimilation cycles), whereas experiment a53 spanned only 4 assimilation cycles (12 hours).)

Table 2: Summary of active observations, minimisation behaviour, and computational metrics for the different thinning configurations. RFL indicates reflectivity observations; H indicates pseudo-observed relative humidity; JO/n indicates that the cost-function value has been normalised by the number of assimilated observations.

	Data Count Screening		Data Count Minimization	No. of iterations performed	Runtime (sec)		Cost function (JO/n)
	H	RFL	H		Screening	Minimization	
a52	8682	382573	8649	36	143.6	129.9	0.04
a53	677570	1987678	631250	50	241.7	2362.7	0.02
a54	2816	335112	989	30	139.4	120.2	0.05
a55	2863	335112	504	30	140.4	121.3	0.06
a56	290	37082	232	30	101.5	108.5	0.05

From the table, it is evident that experiment a53, which has by far the highest number of assimilated observations, reaches the maximum allowed number of Lanczos iterations. In parallel, both the screening and minimisation runtimes increase substantially compared to the other experiments, reflecting the greater computational demand associated with processing such a large observation set. The table also highlights that increasing the sampling interval in BATOR (experiment a56) has a more pronounced impact on observation selection during the screening step, which in turn results in somewhat more favourable computational costs.

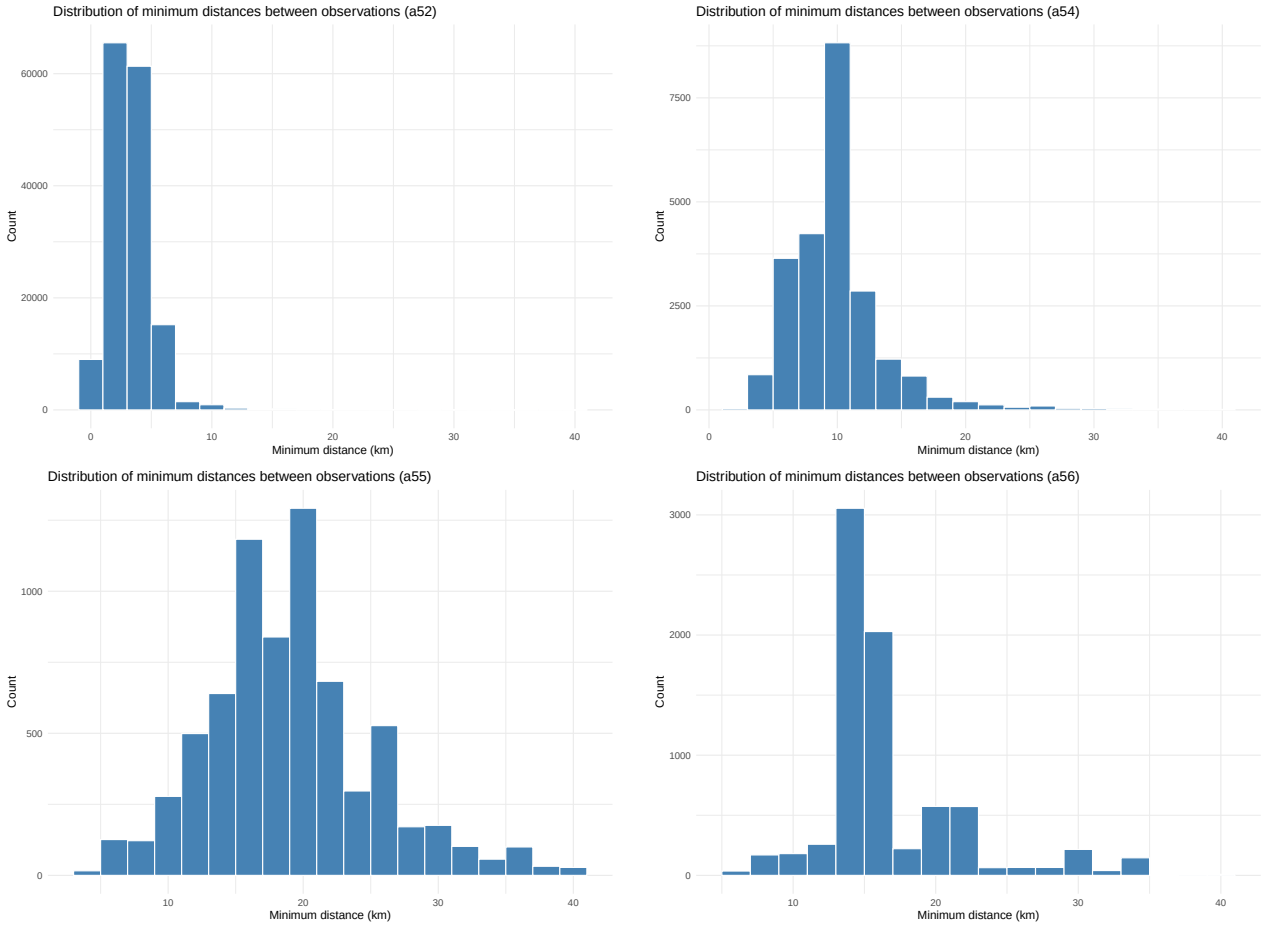


Figure 7: Histograms of minimal distances between observations for experiments a52, a54, a55, and a56. Here experiment a52 serves as a reference case in which the filtering was "turned off".

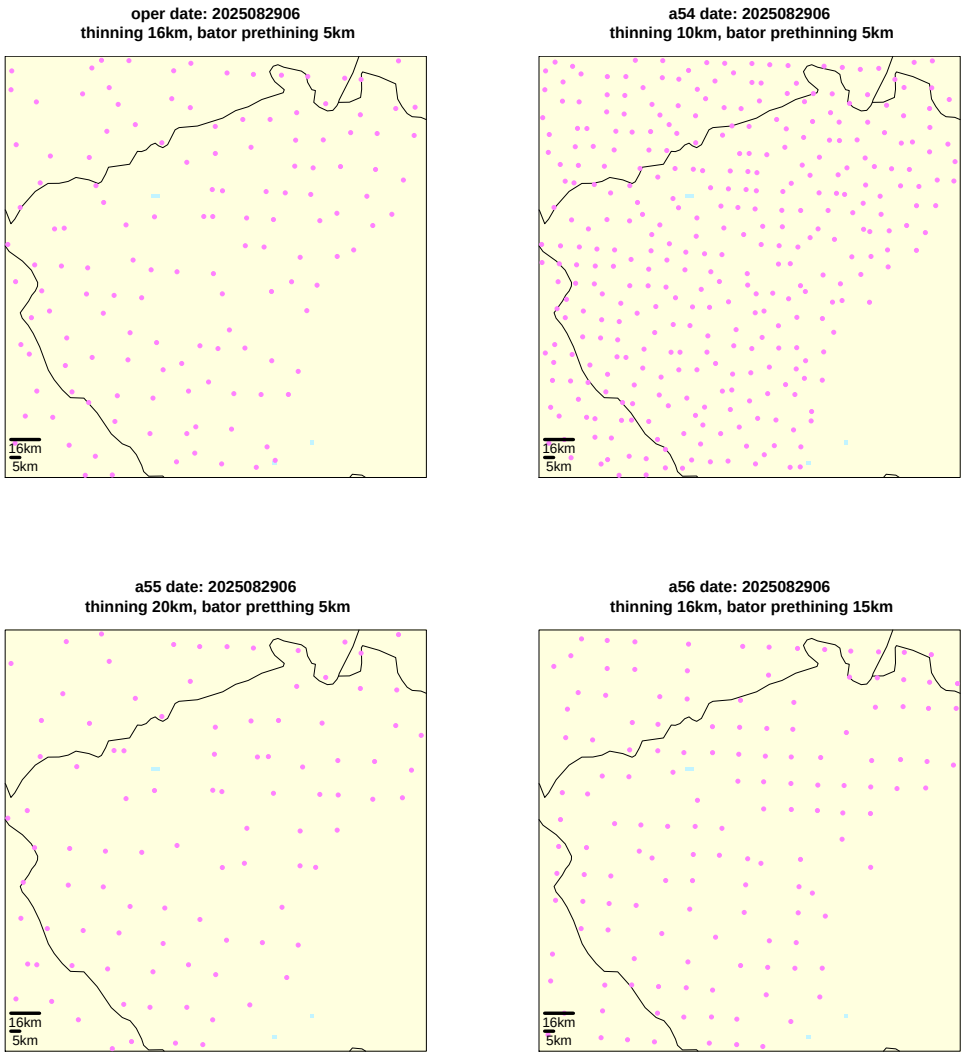


Figure 6: Observation density over western Czechia for the operational reference and thinning experiments a54, a55, and a56

3.2.2 Impact of thinning distance configurations on forecast and verification scores

As we mentioned in Section 2.3, we conducted a verification for experiments a54, a55, and a56 to evaluate how the different thinning configurations affected the resulting forecasts. The verification was performed at 6-hour intervals by computing departures of the forecasted meteorological parameters from their corresponding SYNOP observations at the observation locations, and these differences were then used to derive statistical scores.

The results indicate that, of all meteorological quantities examined, only the precipitation scores showed noticeable deviations from the operational reference, suggesting that changes in thinning distance influenced mainly the precipitation forecast, while other fields remained essentially unchanged. Figure 8 illustrates the evolution of the mean error (bias) and the root mean square error (RMSE) as a function of forecast lead time for all experiments, shown for precipitation in the forecasts initiated at 00 UTC and 12 UTC, respectively.

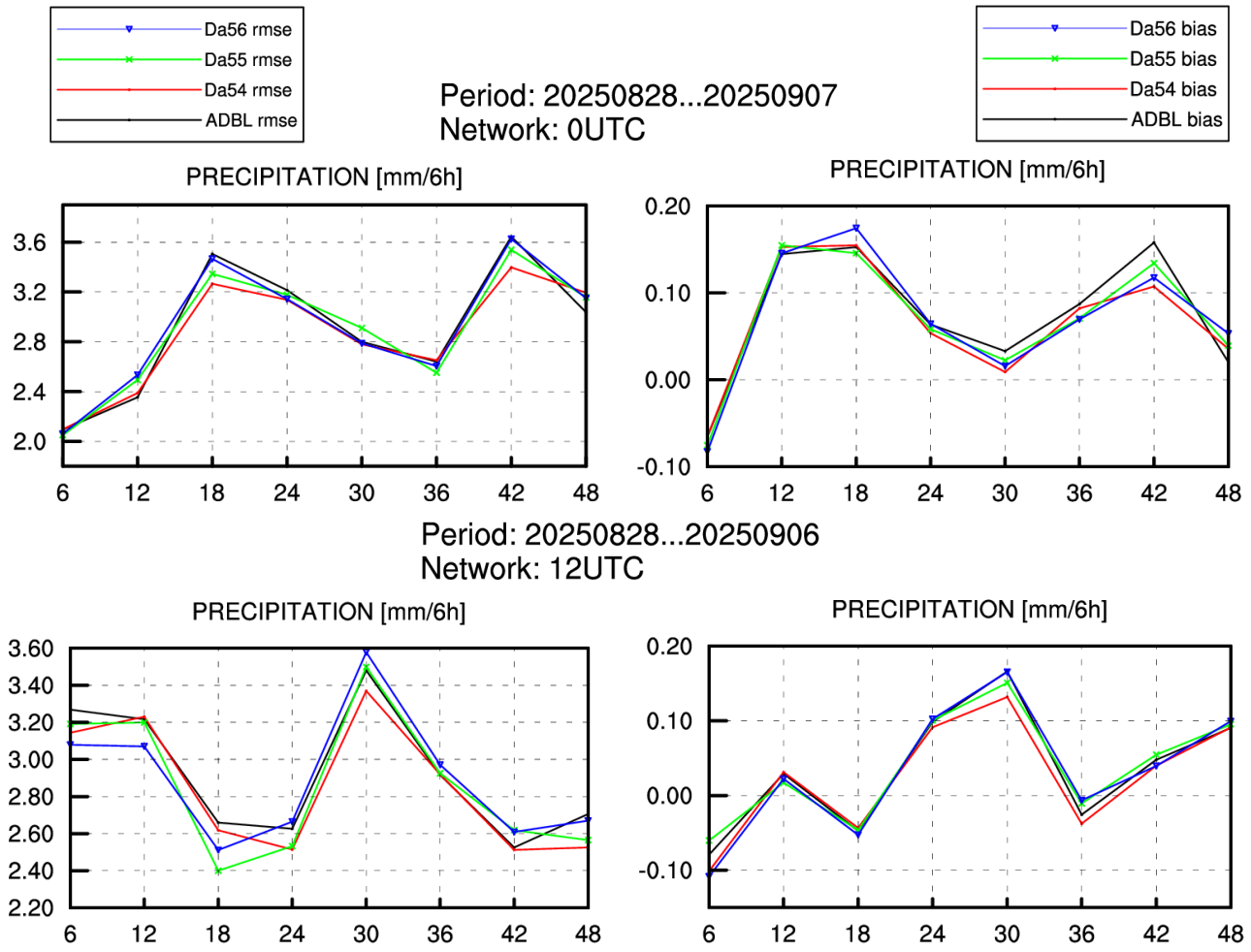


Figure 8: Evolution of RMSE (left) and bias (right) as a function of forecast lead time for all experiments for precipitation in the 00 UTC (upper) and 12 UTC (lower) production cycles, with ADBL denoting the operational reference.

Based on these plots, the precipitation verification shows either a neutral or slightly positive impact across all three experiments. Only in a few cases does experiment a56—which is the configuration where the sampling distance was increased in BATOR—exhibit a minor degradation relative to the operational reference. Among the tested configurations, a54—corresponding to the lowest thinning distance of 10 km—appears to perform the best overall, as its RMSE remains consistently below, or at worst comparable to, that of the operational reference throughout the forecast range.

To visualise the results, Figure 9 presents the total precipitation fields for all experiments alongside the operational run, complemented by the observational reference derived from the combined radar reflectivity and rain gauge product over Czechia. The spatial structures appear to be remarkably similar in all configurations, with only minor variations in precipitation intensity, indicating that the different thinning distances exert minimal influence on the overall forecast performance.

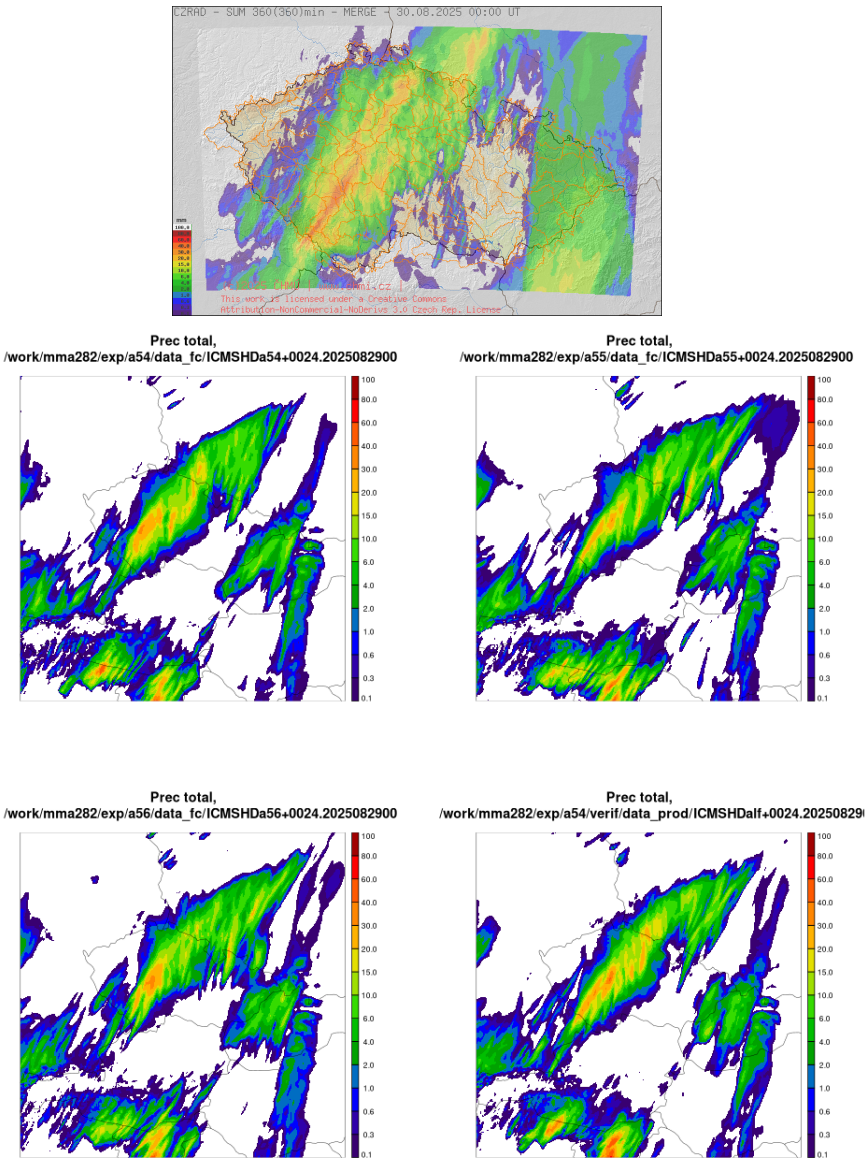


Figure 9: 24-hour forecast of total precipitation from the 00 UTC 29 August 2025 runs for experiments a54 (upper left), a55 (upper right), and a56 (lower left) with varying thinning distances, compared with the operational configuration (lower right) and the combined radar reflectivity and rain gauge precipitation estimate (top). A 6-h accumulation period is applied to all fields.

To further assess the impact of the different thinning distances, we computed observation–minus–guess (OMG) statistics for active pseudo-observed relative humidity, pseudo-observed reflectivity, and observed reflectivity. These results are shown in Figure 10. The first-guess departures are also plotted separately for dry (flgdyn = 0) and wet (flgdyn = 8) observations to evaluate their individual contributions to the overall statistics. The rightmost curves in each plot clearly illustrate the influence of the thinning distances on the number of active observations: experiment a54 separates noticeably from the others, reflecting the expected increase in sample size associated with its reduced thinning distance. This warrants some caution, as a larger sample may shift the resulting statistics (something we could already notice when discussing the observation-error correlation diagnostics in Section 3.1). What also immediately stands out is that dry observations are far more numerous than wet ones and therefore seem to drive the overall statistics. In addition, experiment a56 appears to dry the atmosphere the most among all configurations, shifting the mean relative-humidity departures further into negative values.

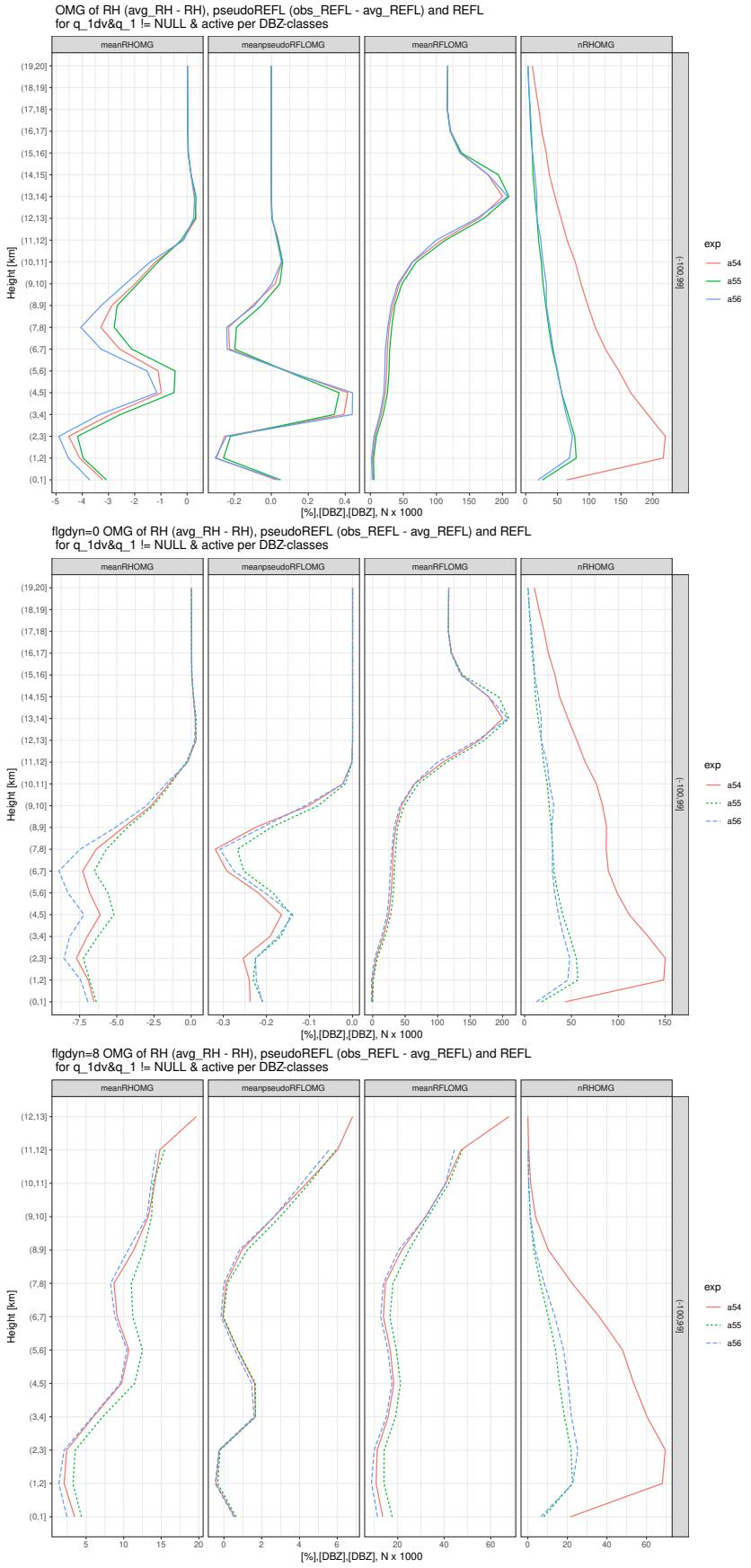


Figure 10: Observation–minus–guess statistics for relative humidity (meanRHOMG), pseudo-observed reflectivity (meanpseudoRFLOMG), and observed reflectivity (meanRFLOMG) for experiments a54, a55, and a56. The upper, middle, and lower panels show results for all active, dry (flgdyn = 0), and wet (flgdyn = 8) observations, respectively. The rightmost panels display the number of active relative humidity pseudo-observations.

As a spatial verification metric, Fraction Skill Scores (FSS) [7] were computed over the full evaluation period to compare the precipitation fields across the different experiments. The FSS is computed here against a combined product of radar reflectivity and rain-gauge measurements, which is available only over the area of the Czech Republic. Figures 11 and 12 show the fraction of useful forecasts—an FSS value must exceed a predefined threshold to be considered useful—across all forecast ranges for several precipitation thresholds, using a 1-hour accumulation period, for the three experiments relative to the operational reference.

Both figures indicate that all experiments, as well as the operational run, achieve very similar scores. Unfortunately, it is not possible to assess the FSS at higher precipitation thresholds in the present case, as the percentage of useful forecasts drops sharply from 4mm/1h onward, as illustrated in Figure 12. For light precipitation, however, more than 70% of the forecasts are useful with reasonably good spatial accuracy across the area. This percentage decreases progressively with increasing thresholds, and the displacement error becomes more pronounced as well. The experiments differ most clearly at the 1mm/1h and 2mm/1h thresholds. At 2mm/1h, experiments a54 and a55 show a very slight improvement, while around the 1mm/1h threshold, experiment a56 performs the worst among all configurations.

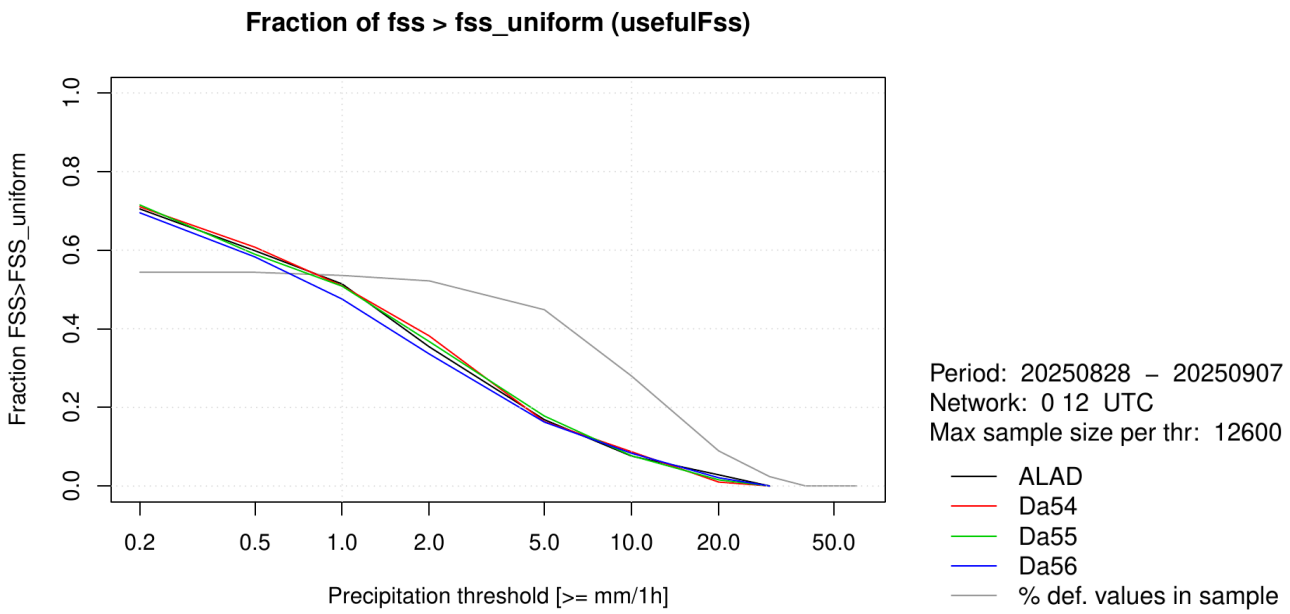


Figure 11: Fraction of useful forecasts (FSS > FSS_uniform) for all experiments compared to the operational reference (ALAD). The grey line indicates the percentage of observations that meet the criterion for the FSS to be considered useful.

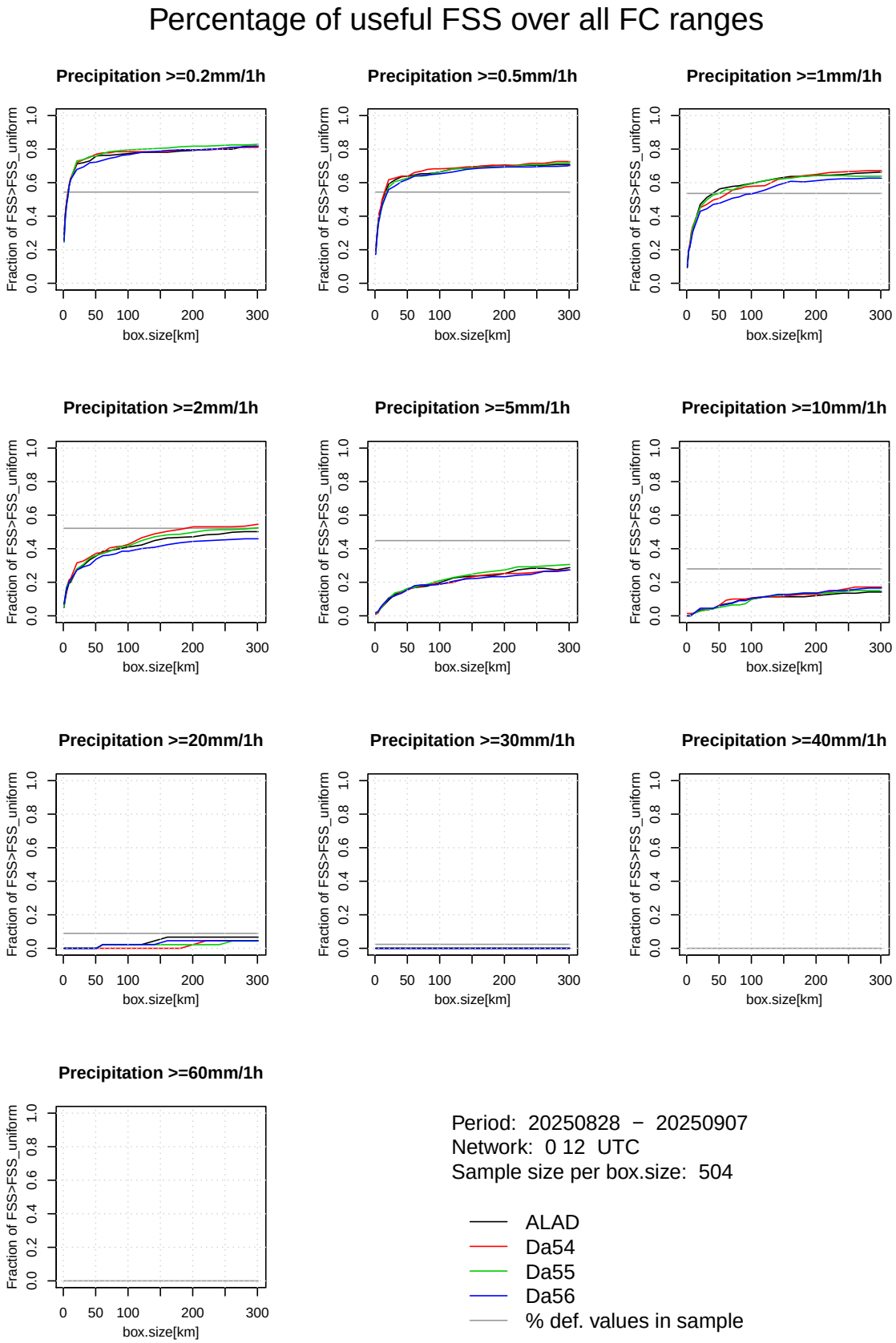


Figure 12: Percentage of useful forecasts as a function of box size across all forecast ranges for several precipitation thresholds (1-hour accumulation) for the three experiments relative to the operational run (ALAD). The grey line shows the percentage of observations that meet the criterion for the FSS to be considered useful.

4 Conclusion

This study examined the impact of different thinning-distance configurations on the assimilation of radar reflectivity in ALARO cy46, with the goal of identifying a setup that balances observational information content, error correlations, and computational efficiency. Using observation–error correlation diagnostics based on the Desroziers method [2], we estimated that a suitable thinning distance lies roughly between 7 and 15 km. At the same time, the diagnostics clearly showed that both the length of the evaluation period and the number of assimilated observations strongly influence the diagnosed correlations: longer periods and larger sample sizes produce smoother and generally lower correlation curves.

In light of these findings, three sensitivity experiments (a54, a55, a56) were carried out to assess how different thinning configurations affect the assimilation cycle and subsequent forecasts. The verification results indicated that thinning distance does have a measurable influence on assimilation behaviour and forecast quality, although the overall impact remained modest. Adjusting the pre-thinning in BATOR proved to be an effective way to address memory and computational-performance issues, as seen in experiment a56. However, increasing the sampling distance in BATOR—and thus filtering out many observations before later quality-control steps—may be compromising: across all verification metrics, a56 consistently performed somewhat worse than both the operational reference and the other thinning configurations. The other two experiments behaved similarly to the operational run, with a54 (10 km thinning in screening) showing the most consistently neutral or slightly positive impact.

Both the ObsTool diagnostics and the verification results suggest that a longer experimental period would be beneficial for obtaining more precise and reliable conclusions. Extending the evaluation period would help stabilise the diagnosed correlations and provide a more accurate assessment of the optimal thinning configuration.

In addition to refining the thinning distance itself, future work should also examine the potential of superobbing as an alternative means of reducing observation density, to assess whether it provides a more efficient or more physically consistent approach than the current thinning strategy applied to radar reflectivity data.

5 Acknowledgements

This work was supported by the ACCORD consortium, whose financial contribution is gratefully acknowledged. Special thanks are also extended to the Czech Hydrometeorological Institute for their warm hospitality and for providing an excellent working environment during the research stay.

References

[1] Abel, S. J., and I. A. Boutle (2012), An improved representation of the raindrop size distribution for single-moment microphysics schemes, *Quarterly Journal of the Royal Meteorological Society*, *138*(669), 2151–2162, doi:https://doi.org/10.1002/qj.1949.

[2] Desroziers, G., L. Berre, B. Chapnik, and P. Poli (2005), Diagnosis of observation, background and analysis-error statistics in observation space, *Quarterly Journal of the Royal Meteorological Society*, *131*, 3385–3396, doi:https://doi.org/10.1256/qj.05.108.

[3] Duncan, D., N. Bormann, A. Geer, and P. Weston (2023), Superobbing and Finer Thinning for All-sky Humidity Sounder Assimilation, doi:https://doi.org/10.21957/5c3b9c8d9f.

[4] EUMETNET OPERA Programme (n.d.), OPERA Radar Data Centre (ODC) Database.

[5] Liu, Z.-Q., and F. Rabier (2003), The potential of high-density observations for numerical weather prediction: A study with simulated observations, *Quarterly Journal of the Royal Meteorological Society*, *129*, 3013–3035, doi:https://doi.org/10.1256/qj.02.170.

[6] Marshall, J. S., and W. M. K. Palmer (1948), The distribution of raindrops with size, *Journal of Atmospheric Sciences*, *5*(4), 165 – 166, doi:https://doi.org/10.1175/1520-0469(1948)005<0165:TDORWS>2.0.CO;2.

[7] Mittermaier, M., N. Roberts, and S. A. Thompson (2013), A long-term assessment of precipitation forecast skill using the fractions skill score, *Meteorological Applications*, *20*(2), 176–186, doi:https://doi.org/10.1002/met.296.

[8] Skamarock, W. C. (2004), Evaluating Mesoscale NWP Models Using Kinetic Energy Spectra, *Monthly Weather Review*, *132*(12), 3019 – 3032, doi:https://doi.org/10.1175/MWR2830.1.

[9] Smerkol, P. (2022), *Documentation for the HOOF2 preprocessing tool*, ARSO, technical documentation for the HOOF2 radar preprocessing tool.

[10] Wattrelot, E., O. Caumont, and J.-F. Mahfouf (2014), Operational Implementation of the 1D+3D-Var Assimilation Method of Radar Reflectivity Data in the AROME Model, *Monthly Weather Review*, *142*, 1852–1873, doi:https://doi.org/10.1175/MWR-D-13-00230.1.