Short Term Scientific Mission Report

EUMETNET - Weather Forecasting Cooperation E-WFC

Prepared by: Martin Petrovič

For electronic submission to: e_wfc@eumetnet.eu

Report's summary

Title of STSM	Radar radial wind data assimilation
Name of candidate	Mgr. Martin Petrovič
Institute of candidate	Comenius University, Faculty of mathematics, physics and informatics & Slovak Hydrometeorological Institute, Slovakia
Host institute	Czech Hydrometeorological Institute, Prague
Contact point in host institute	Mgr. Alena Trojáková
Date and length of STSM	26.4.-16.5.2026, three weeks

1. Abstract and key words

This STSM investigated the assimilation of Doppler radar radial wind (DOW) observations from the EUMETNET OPERA network into the ALADIN high-resolution NWP model (Cycle 48). Six experiments for a test period in March 2026 – a passive reference (r01) and five active runs (r02–r06) – were designed as controlled pairs to isolate the individual effects of dealiasing, radar network density, observation rejection limit, and spatial filtering method on both innovation statistics and forecast quality. In the three-country baseline (r02), whose radars all have a Nyquist velocity (NI) above 30 m/s, folded velocities still passed the background check and left the assimilated innovations the least Gaussian of all experiments, as measured by the Block Bootstrap Anderson-Darling (BBAD) test. Enabling dealiasing alone (r04) reduced the raw innovation standard deviation by 14.8 percent, removed the aliasing bands from the innovation field and lowered the data loss required to reach a Gaussian innovation distribution by 2.3 percentage points at 40 of the 43 radars. Expanding to the full 77-station, 11-country network (r03) with dealiasing enabled for sites with a NI of less than 30 m/s showed that such radars can be added without degrading the data quality or the forecast. A tighter rejection limit (r05) made the assimilated data more Gaussian mainly by truncation, while superobservations (r06) degraded the fit to near-surface observations. None of the configurations produced a clearly positive forecast impact against the operational suite over the two-week period, and a positive mean innovation of +0.21 to +0.28 m/s in all experiments points to the observation and background error statistics as the next step. Taken together, the results provide a systematic, quantitative basis for tuning multi-country radar wind assimilation in ALADIN NWP system and demonstrate the value of the BBAD quality dashboard.

Key words:
Doppler radar, radial wind, data assimilation, ALADIN, observation error, dealiasing, quality control, BBAD, OPERA network, nowcasting

2. Introduction (purpose / scope)

The assimilation of Doppler radar radial wind observations into high-resolution numerical weather prediction (NWP) models, such as ALADIN NWP system (Termonia et al, 2018) with its canonical model configurations ALARO and AROME, is a critical priority for enhancing the accuracy of short-range forecasts, particularly for convective events and local wind fields. Doppler radar radial wind (DOW) observations provide high-resolution measurements of atmospheric dynamics at scales (< 5 km) critical for convection forecasting and local wind field

analysis. However, integrating these observations into NWP models across a heterogeneous, multi-country European network presents challenges:

- Aliasing: When real wind speed exceeds Nyquist velocity, the radar measures opposite direction or wrong velocity, creating observation quality issues. A more detailed explanation is in Sireci (2005) and Brown and Wood (2007).
- Data heterogeneity: Radar network has different configurations, metadata standards, processing levels across countries

Scientific objectives of this STSM are aimed at:

- Quantify forecast impact of DOW in ALADIN system operated at Czech Hydrometeorological Institute (ALADIN/CHMI)
- Isolate effects of dealiasing, network density, observations error specification and filtering methods
- Validate framework for multi-country quality assessment and assimilation
- Provide scientific basis for implementation of DOW DA in ALADIN/CHMI

3. Data and method

Primary radar observations came from the EUMETNET OPERA, which provides the quality-controlled volume data at the NIMBUS production line in HDF5 format (EUMETNET, 2024). The homogenization of the raw radar data was performed using HOOF2.2.py, which is the Homogenization of OPERA Files tool (Smerkol et al., 2023).

The study included 43 stations in a three-country (France, Germany, Slovakia) configuration (r02, r04, r05, r06), and 77 stations across 11 (Belgium, Czech Republic, Denmark, France, Germany, Poland, Romania, Slovakia, Slovenia, Spain, United Kingdom) countries in the full-network configuration (r01,r03). The radar observations were available at 15-minute intervals, but only the data valid at 00 minutes, containing measurements from the previous 10–15 minutes, were used and assimilated in a 3-hour cycle. The data included extensive quality preprocessing including data homogenization, dealiasing using the torus mapping method and spatial filtering.

To evaluate the impact of the modified radar processing strategies, forecast performance was assessed using a dual approach. First, direct inter-comparisons were made between the experimental runs and the operational ALADIN (ALAD) reference to quantify the spatial and structural impact of the changes. Second, objective verification was performed against standard observation networks to evaluate forecast accuracy at various lead times. Upper-air forecast verification (vertical profiles of wind and temperature) was conducted using the TEMP radiosonde network, while ground-level verification relied on SYNOP surface observations.

Model configuration:

This study used the ALARO configuration of the ALADIN system's canonical model (Termonia et al., 2018). Specifically, it is the operational model configuration used by the Czech Hydrometeorological Institute (Brožková et al., 2019). This configuration couples non-hydrostatic (NH) dynamics and a set of ALARO-1vB physical parameterizations that are suited to modelling atmospheric motions from planetary to meso-gamma scales. The model cycle 48t3mas_op1 is used, with a horizontal resolution of 2.3 kilometers and 87 vertical levels. The model runs every three hours and provides forecasts up to 54 hours ahead, from 00 and 12 UTC,

for evaluation purposes. The integration domain is the ALADIN/CHMI operational domain (Figure 1).

Data assimilation includes surface analysis based on optimal interpolation (OI) and BlendVar analysis for upper fields. This consists of digital filter spectral blending (Brožková et al., 2001), followed by 3DVAR analysis (Fischer et al., 2005), which assimilates the following observations: SYNOP, TEMP, AMDAR, Mode-S, SEVIRI, WP, HR-AMV, ASCAT and radar reflectivity.

Preprocessing workflow:

1. HOOF2.2 Homogenization:
   - Format conversion (HDF5 ODIM)
   - Metadata harmonization (Nyquist velocity, gain/offset)

2. HOOF2.2 Dealiasing:
   - Torus mapping algorithm (Smerkol et al. 2025) using model wind projection
   - Quality flagging (QI – quality indicator)

3. Superobservation (SO) generation – HOOF2.2:
   - Spatial binning: configurable range/angle factors
   - multiple nearby observation points are combined into a single representative point (Ridal et al. 2023)

4. BATOR preprocessing and spatial filtering (Montmerle and Faccani, 2009):
   - Wind cleaner (elevation-level median filter)
   - Spatial median filter (per-pixel neighbourhood analysis)
   - Background check (rejection limit)
   - Thinning: Sample=5000, Resolution=1000 m

The output was an ODB (Observation Data Base) file prepared for assimilation into the ALADIN system. After this, the ODB file continues with the Screening (quality control) process, producing ECMA files containing the differences between the model background (also known as the 'guess') and the observations. These differences are known as OMG (observations minus guess), fg_depar (first guess departure) or innovations; all three names denote the same quantity, and the term innovation is mainly used in this report. Up to this point the process is passive assimilation: the observations are not yet used to compute the new analysis from which the forecast is integrated.

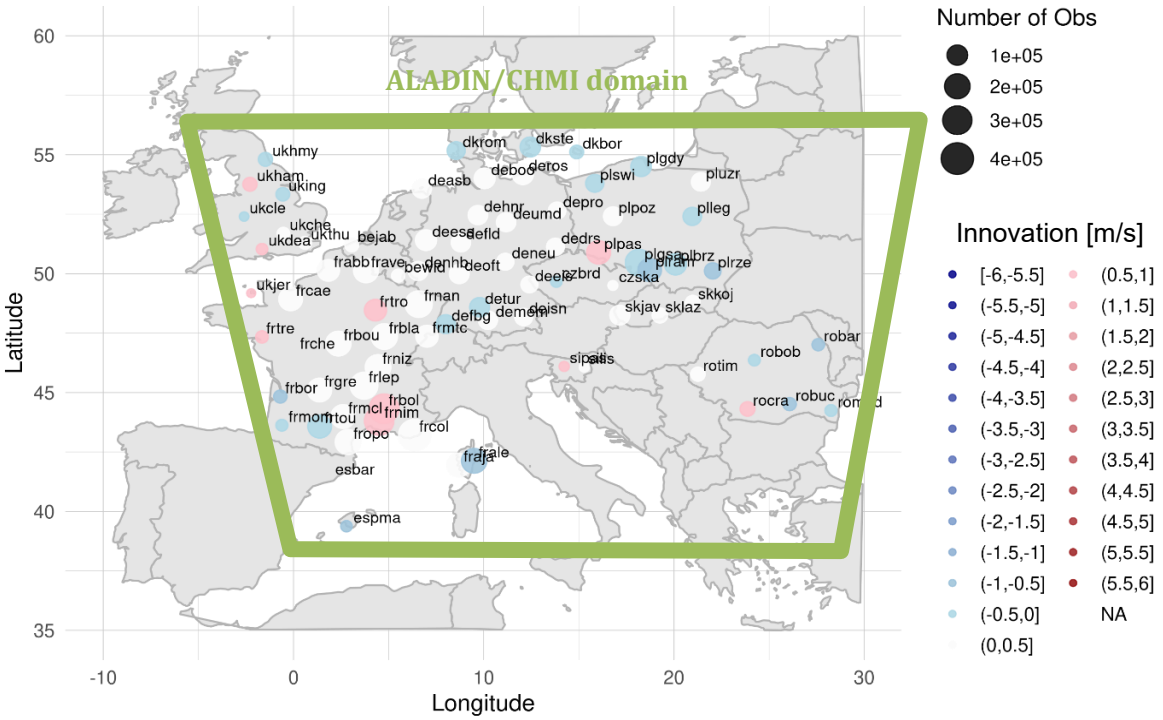


Figure 1 *Model and radar network domain of used radars, showing mean innovation value and number of observations of each radar from experiment r01.*

Quality assessment methodology

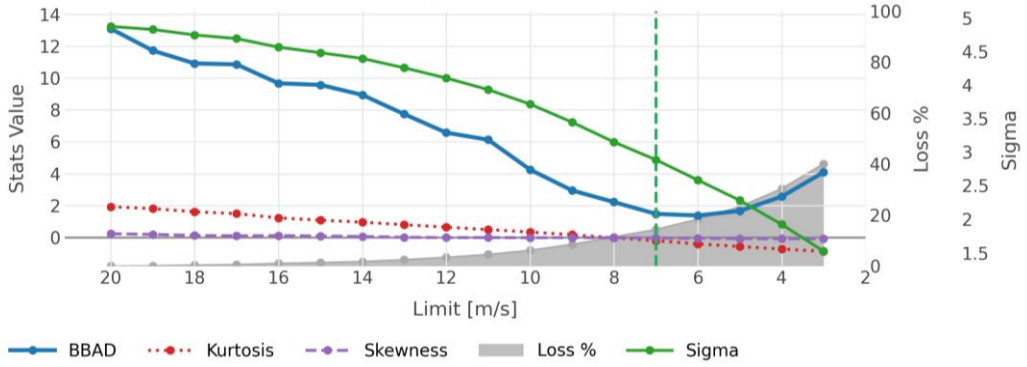
The basis was an initial assessment of statistical properties: data volume, error distribution, innovations and OMA (Observations Minus Analysis) departures, including innovation histograms. The inverse histogram visualization method (Anderson and Järvinen, 1999) was used to identify the cutoff value (rejection limit) for data exclusion. In variational data assimilation, the innovation vector is assumed to be Gaussian distributed. To objectively evaluate this assumption for radar radial wind observations, we implemented a robust statistical framework based on the Anderson–Darling test (AD).

The methodological contribution involved application of the Block Bootstrap Anderson-Darling (BBAD) normality test, adapted for large radar datasets (Vávra, 2015). Standard AD tests are highly sensitive to sample size, causing rejection of the Gaussian hypothesis even for negligible deviations when dealing with hundreds of thousands of observations. The BBAD approach addresses this by resampling 1,000 observations in contiguous 10-point blocks to preserve spatial and temporal correlation structure, repeating this process 100 times to calculate a median AD statistic.

The BBAD threshold classification was established as follows: values below 1.50 pass the test indicating good Gaussian fit; values between 1.50 and 2.50 warn; values above 2.50 fail, indicating non-Gaussian behaviour and likely clutter contamination. The multicriterial scoring system evaluated each radar, elevation angle, and distance range using four metrics. The primary metric was BBAD for normality testing. Secondary metrics included skewness to measure asymmetry and systematic bias, and kurtosis to measure tail weight and extreme values. A tertiary metric was the R-squared coefficient measuring goodness of fit to a Gaussian distribution.

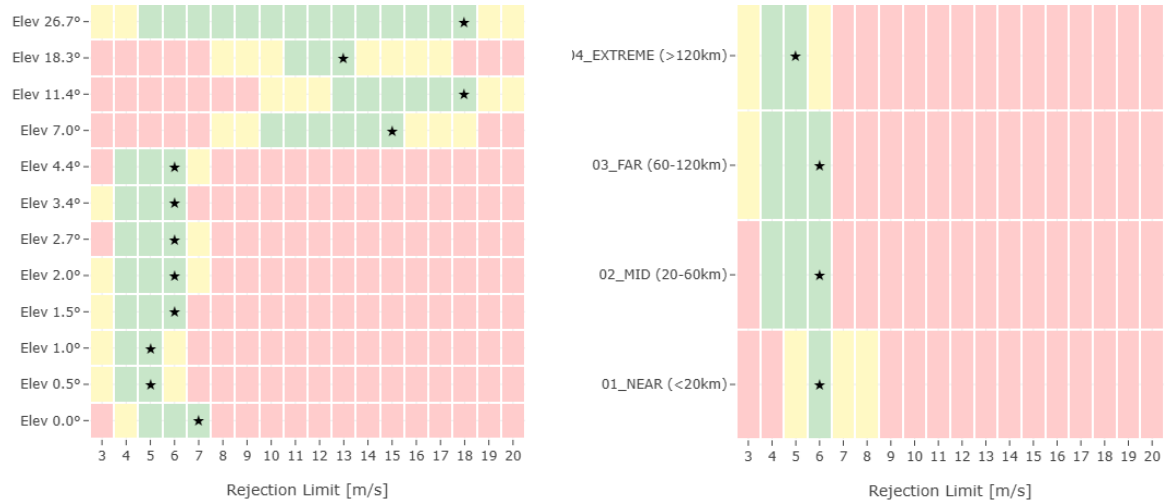
The automated recommendation system created interactive heatmaps showing BBAD score across different rejection limits and spatial bins. This allowed identification of the "sweet spot" rejection value where the maximum number of spatial bins passed quality criteria while minimizing data loss. Furthermore, the quality dashboard was run on global mode (evaluated per radar), on elevations mode (evaluated per radar per elevation) or analogously for distance mode. The dashboard identified problematic radars across the entire network, detected clutter at low elevation angles and beam broadening at high angles and long distances, and quantified degradation with range from the radar. Figure 2 illustrates what can be read from the dashboard in each of the three modes; further examples, together with the results, are presented in the next section.

a) Global mode – one radar, all elevations and distances together



b) Elevation mode – status matrix per elevation angle

c) Distance mode – status matrix per distance segment



Quality Assessment Methodology

The **OVERALL** status is determined by the worst-performing metric among R2, Skewness, Kurtosis, and AD (Bootstrap).

Metric	PASS (Green)	WARN (Yellow)	FAIL (Red)	Description
Block Bootstrap AD (BBAD)	< 1.50	1.50 - 2.50	> 2.50	Normality test. Uses BBAD to normalize results for large datasets.
R-Squared (R2)	> 0.950	0.900 - 0.950	< 0.900	Goodness of fit for Gaussian distribution.
Skewness	< 0.500	0.500 - 1.000	> 1.000	Absolute value of distribution asymmetry.
Kurtosis	< 1.500	1.500 - 3.000	> 3.000	Absolute value of distribution sharpness/peakedness.

Figure 2 *Illustration of how the BBAD quality dashboard is read, for radar BEJAB (active observations, experiment r01): (a) global mode, (b) elevation mode, (c) distance mode.*

In Figure 2, the global mode (a), the statistics of one radar are followed as a function of the rejection limit: the curves of BBAD, kurtosis and skewness are judged against the pass, warn and fail thresholds, the recommended limit is marked by the green dashed line, and the price of a tighter limit is read from the standard deviation (sigma) and the percentage of rejected data. In the elevation mode (b) and the distance mode (c), every row is one elevation angle or one distance segment, and every cell is one rejection limit. The color of the cell is the overall status, given by the worst of the four metrics, and the black star marks the optimal limit of that row. Comparing the rows shows where the data are Gaussian already with a loose limit and where a tight limit is needed.

The dashboard is evaluated per radar, so observations from different radars are never pooled into a single test; the experiments are compared through the median and interquartile range of the per-radar values. Two operating points are used. At the common rejection window of 20 m/s the aliased departures are still present, so the BBAD value measures how strongly the innovations depart from a Gaussian; it is used here as a relative measure between experiments, and the PASS, WARN and FAIL categories, which are defined for screened data, are not applied at this point. At the rejection limit recommended by the dashboard almost every radar reaches a Gaussian distribution by construction, so the informative quantities there are the recommended limit itself, the fraction of data it discards and the standard deviation of the remaining core of the distribution. In addition, a Gaussian is fitted by non-linear least squares to the histogram of the innovations of each experiment, with its area constrained to the number of observations; its standard deviation (σ_{core}) describes the width of the core of the distribution and is less sensitive to individual outliers than the sample standard deviation.

Experimental design

Six experiments (r01–r06) were designed around two anchor points, the passive reference experiment r01 and the active pivot experiment r02 (Figure 3). The passive experiment r01 does not use the observations to compute the analysis and provides reference statistics for the whole radar network. The active experiments r02–r06 assimilate the observations. Every active experiment differs from its partner in exactly one setting.

- **r01 – passive reference:** all 77 radars from 11 countries, 6 weeks, dealiasing disabled, superobservations disabled and BATOR filtering enabled. It served as reference statistics.
- **r02 – active pivot:** three-country configuration of France, Germany and Slovakia with 43 radar stations, 2 weeks (5 to 19 March), dealiasing disabled, superobservations disabled and BATOR filtering enabled. The rejection limit was left at its default; for this purpose, we added the new variable RZFACT, adjustable via namelist, which adjusts the rejection limit (set to 1). This configuration covers the region with the highest data density in Europe and radars with Nyquist velocity (NI) higher than 30 m/s.
- **r03 – full network:** all available radars, 77 stations across 11 countries, the same 2-week period, dealiasing enabled, superobservations disabled and BATOR filtering enabled.
- **r04 – dealiasing test:** the same three-country configuration as r02, with dealiasing enabled.
- **r05 – rejection limit test:** the three-country configuration as r02, with a much tighter rejection limit, RZFACT = 0.16, to test whether aggressive rejection of large-innovation observations improves the results.
- **r06 – superobservations versus BATOR filters:** the three-country configuration as r02, with superobservation processing enabled and BATOR filtering disabled, providing a head-to-head comparison of superobservation versus native-resolution plus filtering approaches.

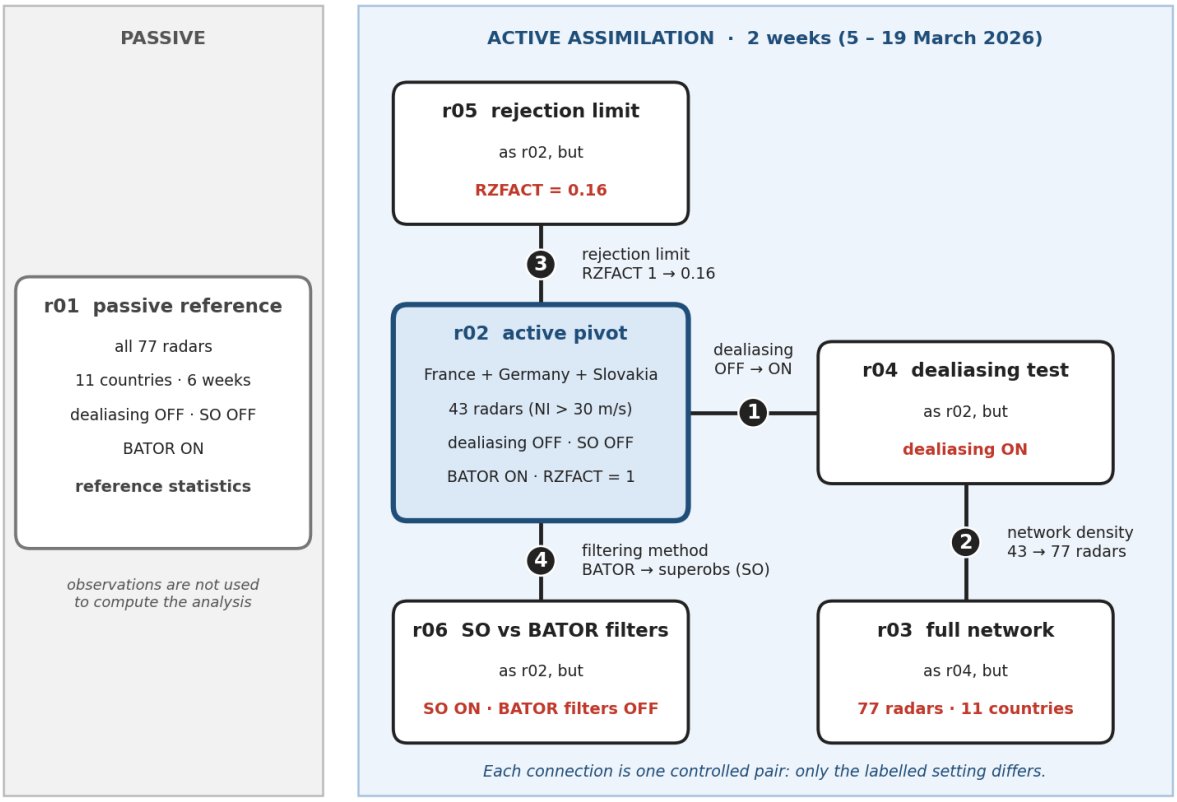


Figure 3 *Design of the experiments. Each connection between two experiments is a controlled pair in which only the labelled setting differs. SO – superobservations, BATOR – BATOR preprocessing and spatial filtering, RZFACT – factor adjusting the rejection limit, NI – Nyquist velocity.*

The design principle ensured that each experimental pair changed exactly one variable, enabling clean comparisons. The pairs numbered in Figure 3 isolate the following effects (also called "degrees of freedom"): (1) r02 versus r04 – dealiasing, (2) r03 versus r04 – network density, (3) r02 versus r05 – rejection limit and (4) r02 versus r06 – the choice of spatial filtering method.

4. Results

Test period 5th – 19th March 2026

From a synoptic perspective, this was a meteorologically dynamic pre-spring period. The change of extensive pressure systems over Europe brought a shifting of phases, ranging from deep atmospheric stability to dynamic frontal situations. At the beginning of this period, the centre of stable high-pressure system shifted from Central Europe eastward. Around March 9th, the centre of this anticyclone moved over Kazakhstan, though its ridge continued to influence the atmosphere over the Carpathian region. Concurrently, a deep trough of low pressure began to deepen from the Atlantic Ocean. By March 11th, this dynamic system manifested as a massive, continuous precipitation band stretching thousands of kilometres—from the Iberian Peninsula and Southern France, up through the Benelux countries and Western Germany, and extending all the way into Scandinavia (Figure 4). Supported by strong warm-air advection ahead of the front and active dynamics from the Atlantic trough, this highly organized boundary delivered widespread, intense precipitation across Western and Central Europe simultaneously.

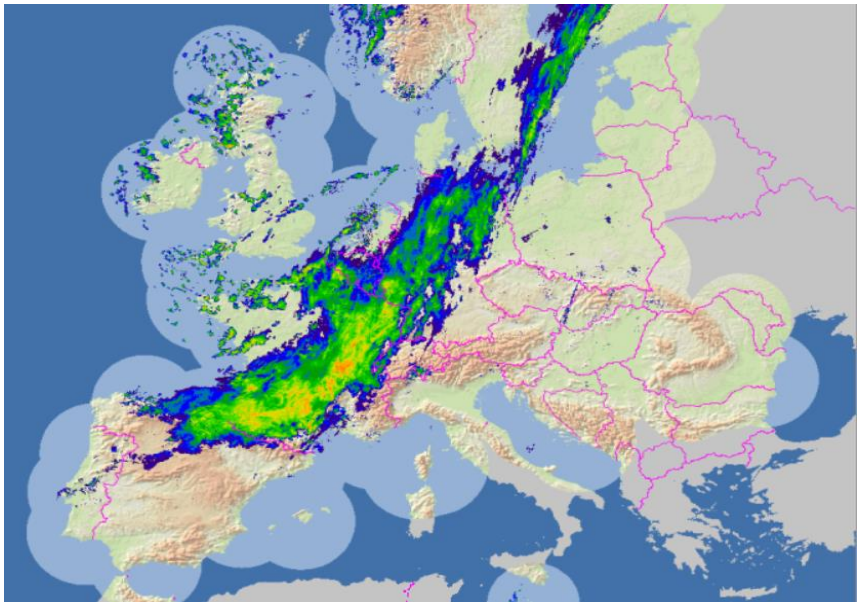


Figure 4 *Radar reflectivity composite of 11.03.2026 from EUMETNET OPERA.*

Experiment results

4.1. The passive reference **experiment r01** established the baseline quality of OPERA radar wind data across the full 11-country network over six weeks (5th March – 15th April). The mean innovation was +0.276 m/s across all radars in the network. A positive bias of comparable magnitude appeared in every experiment as well, ranging from +0.21 to +0.28 m/s (Table 1); its persistence regardless of network configuration, dealiasing status or spatial filtering method points to a systematic origin. At the common rejection window of 20 m/s the innovations are strongly non-Gaussian (median BBAD 15.9 for the assimilated observations, Table 2). The dashboard recommends a considerably tighter limit for each radar, 6 m/s in the median, at which the radars do reach a Gaussian distribution, at the price of discarding 7.8 % of the observations (Table 2). The r01 dataset therefore documents the raw quality of the network and serves as the reference for the per-radar quality assessment. Figure 5 shows the daily number of all and active observations per country; France, Germany and Poland contribute by far the most observations, which reflects mainly the size of their networks rather than their Nyquist velocity.

Exp	Dealias	Network	Mean innov.	N all (10^3)	STD all	σ_core all	N active (10^3)	Active [%]	STD act	σ_core act
r01	OFF	77 st / 11 cty	+0.276	22 800	4.835	2.484	1 907	8.4	4.643	2.613
r02	OFF	43 st / FR+DE+SK	+0.276	5 527	4.333	2.368	415	7.5	4.347	2.375
r03	ON	77 st / 11 cty	+0.245	4 952	3.881	2.328	392	7.9	4.078	2.346
r04	ON	43 st / FR+DE+SK	+0.256	4 382	3.692	2.294	330	7.5	3.962	2.258
r05	OFF*	43 st / FR+DE+SK	+0.270	5 527	4.313	2.337	381	6.9	2.649	2.170
r06	OFF	43 st / FR+DE+SK	+0.210	1 275	4.781	2.421	341	26.7	4.201	2.543

Table 1 *Summary statistics of DOW innovations for each experiment. N all is the number of monitored observations entering the statistics, N active the subset that passed the background check and active [%] their ratio; the statistics labelled "all" and "act" refer to these same two samples. σ_core is the standard deviation of a Gaussian fitted to the histogram of the innovations with its area constrained to the number of observations and describes the width of the core of the distribution. The observation counts are not directly comparable between experiments: r01 covers six weeks (5 March – 15 April) and 77 radars, whereas r02–r06 cover the two-week period (5–19 March), with 77 radars in r03 and the 43-radar three-country network in r02 and r04–r06. In r06 the superobservation processing replaces the native-resolution observations by spatially binned ones, which reduces N all by a factor of four and raises the active fraction correspondingly. * r05 has dealiasing disabled but a tight rejection limit (RZFACT = 0.16, ± 8 m/s), which lowers the active STD by truncation; it uses the same input data as r02.*

Exp	BBAD at 20 m/s median (IQR)	Improved vs r02	Rec. limit median	Data loss at rec. limit	σ at rec. limit
r01	15.9 (9.0–22.7)	–	6 m/s	7.8 %	2.23 m/s
r02	17.0 (9.3–22.6)	–	6 m/s	7.9 %	2.15 m/s
r03	8.9 (3.8–19.6)	34/43	6 m/s	6.7 %	2.13 m/s
r04	10.6 (4.7–21.8)	32/43	6 m/s	4.8 %	1.94 m/s
r05 *	3.8 (2.9–4.9)	42/43	6 m/s	3.2 %	2.12 m/s
r06	14.8 (11.1–19.1)	25/43	6 m/s	8.9 %	2.47 m/s

Table 2 *Quality of the assimilated observations from the dashboard in global mode, one value per radar, given as the median and interquartile range (IQR) over radars. BBAD at 20 m/s is evaluated at the common rejection window, where the aliased departures are still present, and is used as a relative measure between experiments; the PASS and FAIL categories are not applied at this point. "Improved vs r02" counts the radars shared with r02 at which this value is lower than in r02. The remaining columns refer to the rejection limit recommended by the dashboard for each radar, at which almost every radar reaches a Gaussian distribution: the limit itself, the fraction of observations it discards and the standard deviation of the remaining core of the distribution. * In r05 the observations are already truncated at ± 8 m/s by RZFACT = 0.16, so its value at the 20 m/s window is not comparable with the other experiments.*

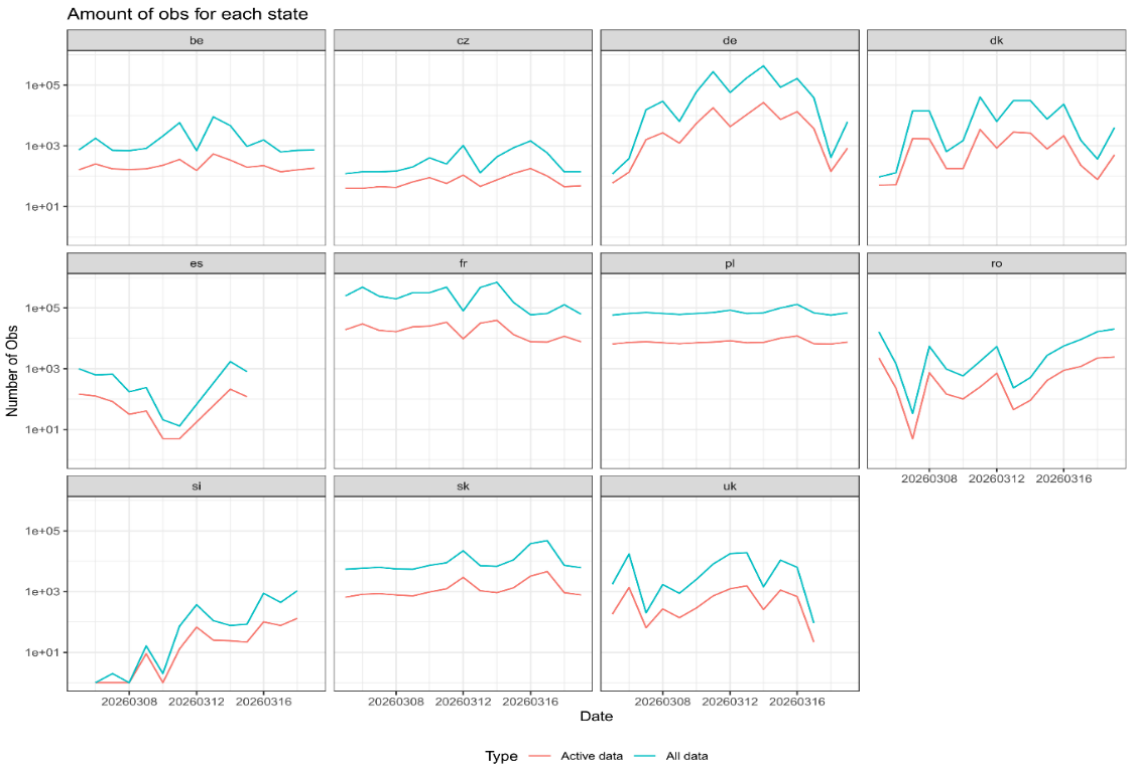


Figure 5 *Visualization of the total (blue) and active (red) Doppler wind observations from each country radars over a selected two-week period from experiment r01.*

4.2. Moving to active assimilation in **experiment r02** – the three-country baseline without dealiasing – the assimilated innovations are the least Gaussian of all experiments: a median BBAD of 17.0 at the common 20 m/s window (Table 2). Reaching a Gaussian distribution requires discarding 7.9 % of the observations, the largest loss of all experiments apart from r06.

The raw innovations reached ±70 m/s, which is physically impossible and confirms the presence of aliased observations. Most of them are rejected by the background check, but folded values that fall inside the ±20 m/s window pass it. This can be seen on histograms of innovations, typically German radars (Figure 6, left). The same pattern can be shown on the French radar with "heavy tails" around ±12 m/s (Figure 6, right). The forecast impact was negative against the operational suite, with wind RMSE degrading at multiple levels and OMA/OMG changes showing degradation concentrated in the surface and lower troposphere.

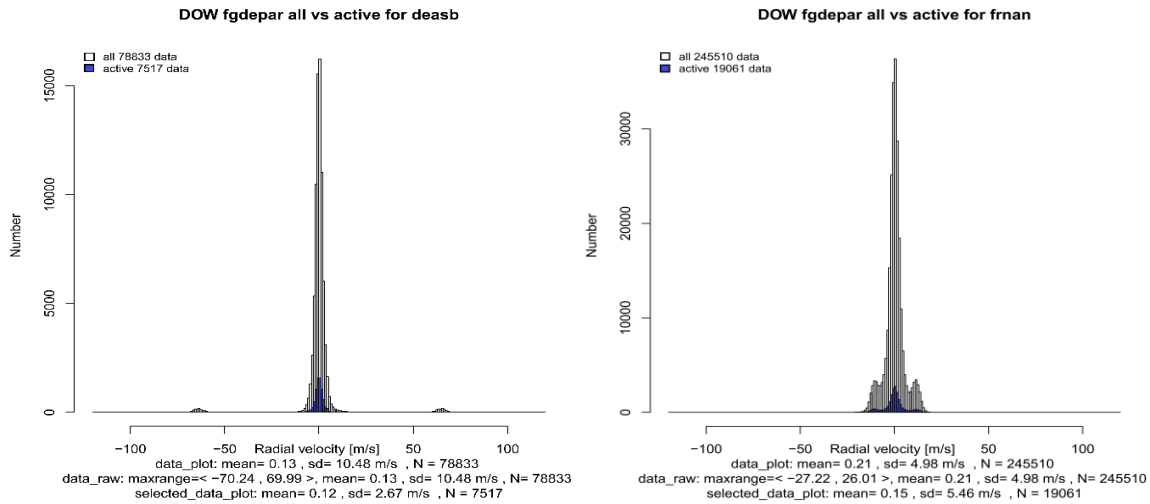


Figure 6 *Histogram of innovations for German radar DEASB and French radar FRNAN, from experiment r02 (gray – all data, blue – active data).*

4.3. The full-network **experiment r03**, which added dealiasing and expanded the station network to 77 stations across 11 countries, gave the most Gaussian assimilated innovations among the experiments that keep the full rejection window, with a median BBAD of 8.9 at 20 m/s (Table 2) and an innovation standard deviation of 3.881 m/s for all observations. This network-wide value is partly a composition effect, as the radars added outside France, Germany and Slovakia are cleaner than the three-country subset. On the 43 radars shared with r02 the BBAD improved at 34 of them, and the data loss needed to reach a Gaussian distribution decreased by 2.0 percentage points in the median. The forecast impact remained slightly negative but was measurably less degraded than r02. The histogram of innovations and the transformed histogram (Anderson and Järvinen, 1999), which linearises the expected Gaussian distribution into a straight line and makes deviations from normality visually apparent, illustrate this improvement (Figure 7).

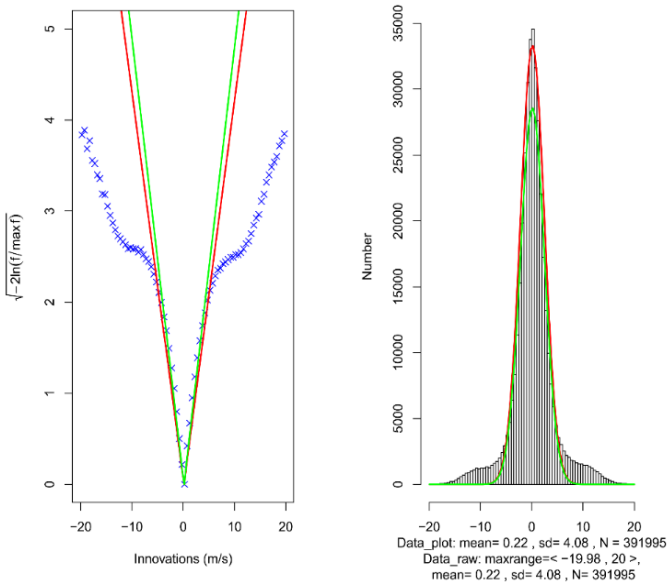


***Figure** 7 Histogram of innovations of active data (right) and transformed histogram (left) from experiment r03 for all radars. Red, or green, describes the normal distribution with offset, or without.*

4.4. Experiment r04, which applied dealiasing to the same three-country network as r02, achieved the tightest raw innovation distribution of all active experiments, with a standard deviation of 3.692 m/s – a reduction of 14.8 percent relative to r02's 4.333 m/s on the identical network. The active innovations also became more Gaussian: the median active-subset BBAD decreased from 17.0 to 10.6 and improved at 32 of the 43 radars (74 %, Table 2).

As example of results, we show the dashboard panel of German radar DEASB highlights how tracking statistical metrics (BBAD, Kurtosis, Skewness, Sigma, and Data Loss %) across distance ranges helps pinpoint the "sweet spot" where the maximum number of spatial bins satisfies Gaussian assumptions without suffering from severe data loss or contamination from aliased near-surface wind components (Figure 8).

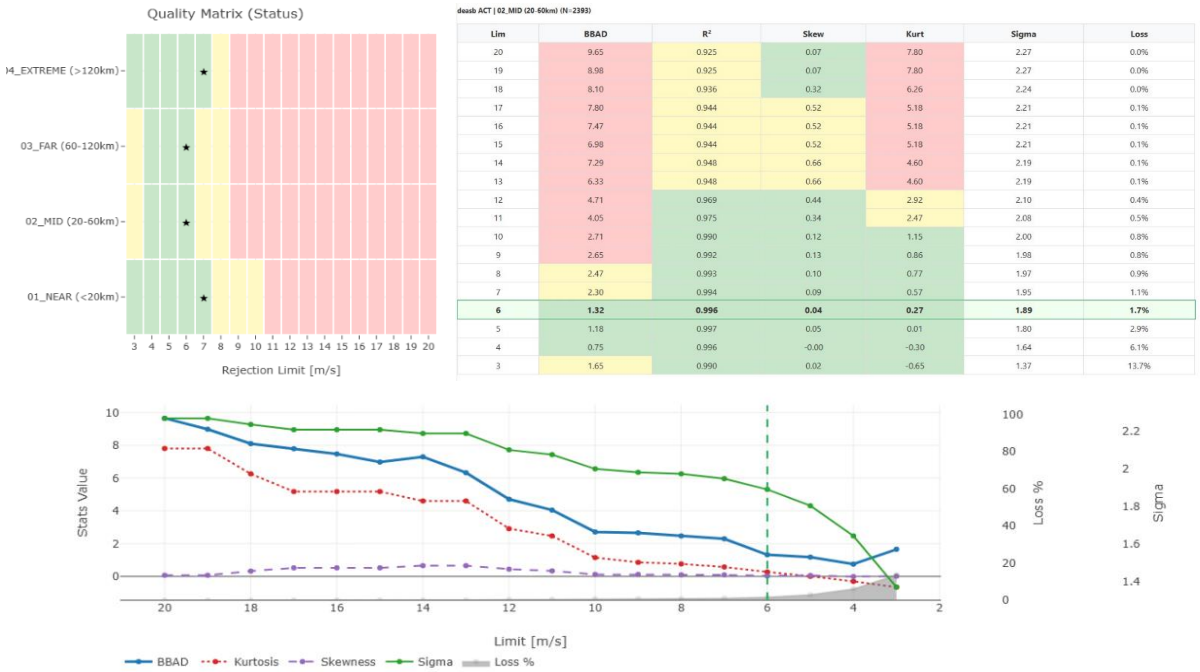


Figure 8 *Radar quality dashboard showing panels of distance segments of DEASB radar from experiment r04 (ACTIVE data). A more detailed description of the legend can be found in the Quality Assessment Methodology section.*

4.5. Experiment r05 used the same input data as r02 but with a much tighter rejection limit (RZFACT = 0.16), restricting the active window to approximately ±8 m/s. The value was chosen on the basis of the r01 quality assessment, and the tighter limit discarded 8.3 % of the active observations. As expected, this sharply reduced the active-subset standard deviation to 2.649 m/s by discarding all large-departure observations. The truncation itself removes most of the non-Gaussian tails, which is why the BBAD of r05 cannot be compared with the other experiments at the 20 m/s window. What remains comparable is the behaviour inside the accepted window: the dashboard still recommends a median limit of 6 m/s, and the standard deviation of the core of the distribution, 2.12 m/s, is essentially the same as in r02 (2.15 m/s) and larger than in r04 (1.94 m/s). Discarding the large departures therefore removes the tails, but it does not improve the core of the distribution in the way dealiasing does.

The resulting verification of forecast scores (Figure 9) for wind speed, wind direction, and geopotential root-mean-square error (RMSE) shows only small differences relative to the operational suite ALAD, i.e. an approximately neutral forecast impact.

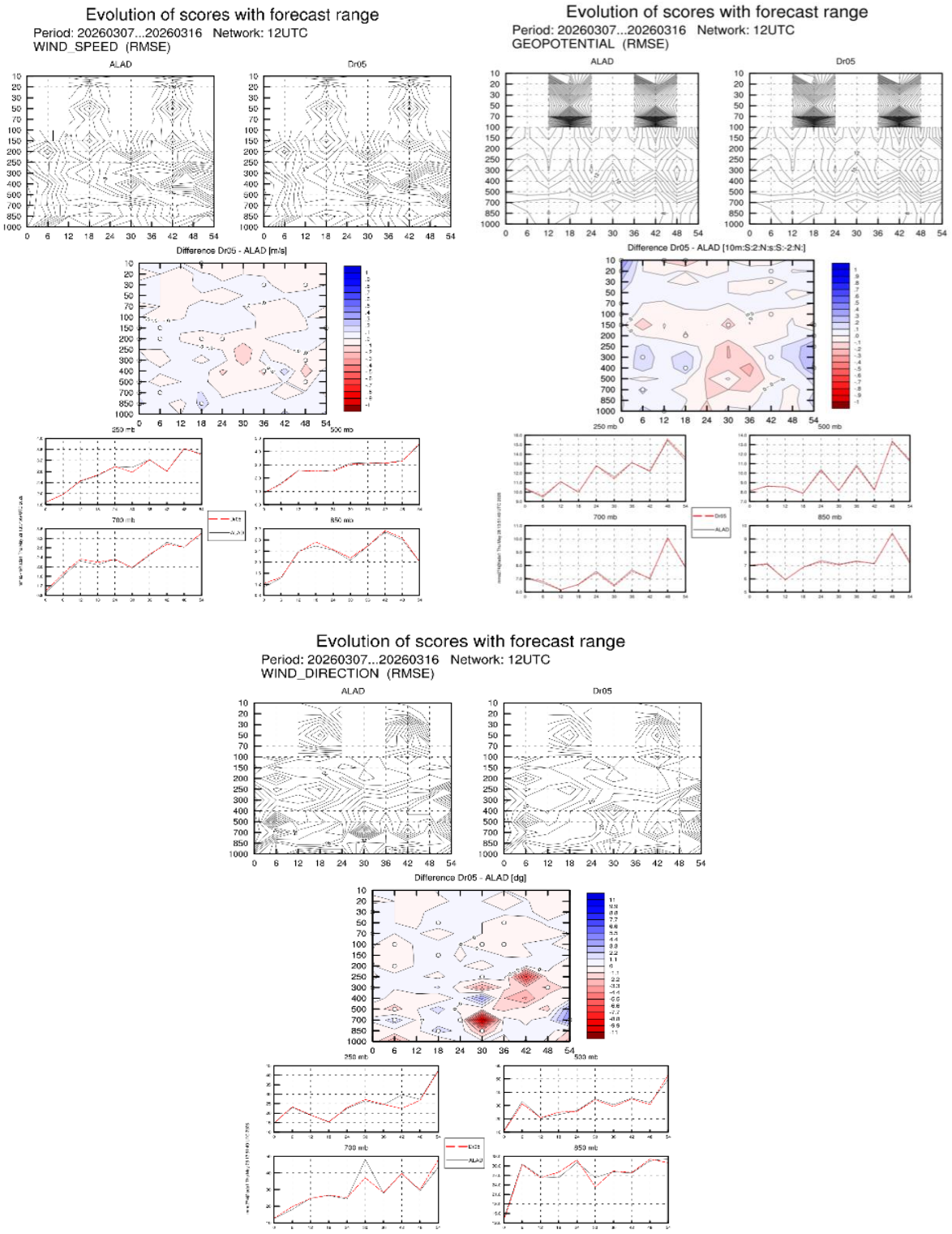


Figure 9 *Verification of forecast scores (exp r05) for the period 20260307–17 (12 UTC network). Three variables are presented: wind speed RMSE in meters per second (top right), wind direction RMSE in degrees (bottom), and geopotential RMSE in metres (top right).*

4.6. Experiment r06, which tested superobservation processing without BATOR filtering, showed the widest raw innovation spread of all experiments (standard deviation 4.781 m/s), consistent with the absence of spatial filtering. Relative to r02 the median BBAD at the 20 m/s window decreased only from 17.0 to 14.8 (Table 2), and the price of reaching a Gaussian distribution remained the highest of all experiments: a median data loss of 8.9 % and the widest core of the distribution, 2.47 m/s. The superobbing algorithm discards bins with a large internal spread, which removes part of the folded velocities, but bins containing a mixture of folded and correct values can still produce biased superobservations; superobservations therefore compensate for aliasing only partially, while the spatial averaging widens the distribution.

However, evolution of forecast verification across seven atmospheric variables (temperature, geopotential, wind components, and humidity) reveals a pronounced degradation signal compared to the operational reference suite (Figure 10).

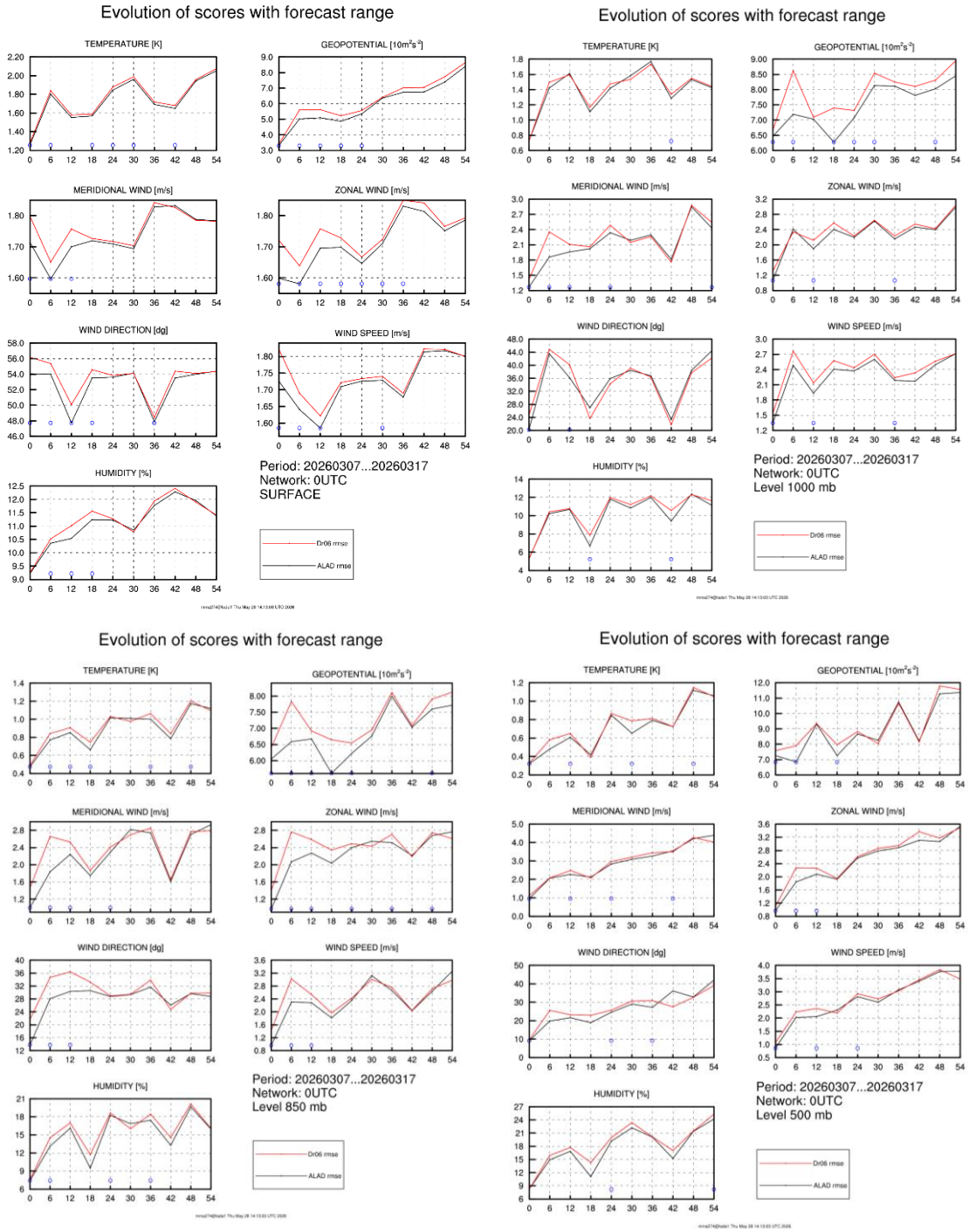


Figure 10 *Verification of forecast scores (exp r06 - Dr06) for the period 20260307–17 (00 UTC). Four vertical levels (Surface, 1000 hPa, 850 hPa, 500 hPa) are presented with 7 variables (Temperature, Geopotential, Meridional wind, Zonal wind, Wind speed, Wind direction and Humidity) showing RMSE for each of it compared to operational suite ALAD.*

4.7. Comparisons – Effect of each Degree of Freedom

The experimental design was constructed such that each pair of experiments isolates exactly one degree of freedom, enabling clean attribution of statistical and forecast changes to specific preprocessing or configuration choices. Four such comparisons are presented here.

4.7.1. Dealiasing Effect (r02 vs r04)

The direct comparison of r02 and r04, identical in network and all other settings but differing only in whether the torus mapping dealiasing algorithm is applied, quantifies the isolated impact of dealiasing. The raw innovation standard deviation decreased from 4.333 to 3.692 m/s (−14.8 percent), the aliasing bands disappeared from the innovation field (Figure 11), the BBAD improved at 32 of the 43 radars, and the data loss required to reach a Gaussian distribution decreased by 2.3 percentage points in the median, at 40 of the 43 radars. Dealiasing also reduced the number of monitored observations by 21 % (Table 1), with a similar relative reduction at every radar and an unchanged active fraction (7.5 % in both). The loss is therefore not a removal of folded observations but of data that the dealiasing step cannot process.

The physical impact of this correction is explicitly visualized in the innovation (fg_depar) scatter plots across distance ranges (Figure 11). In r02 configuration (no_dealias), severe velocity aliasing manifests as distinct, unphysical parallel stripes of extreme departures. For the German network (left panel), these aliasing artifacts are heavily concentrated as high-magnitude departures near 60 to 70 m/s. In the Slovak network (right panel), tighter aliasing bands can be seen extending up to 25 m/s. Activating the torus mapping algorithm (dealias) mostly removes non-Gaussian outlier structures, compressing the departures back around 0 m/s.

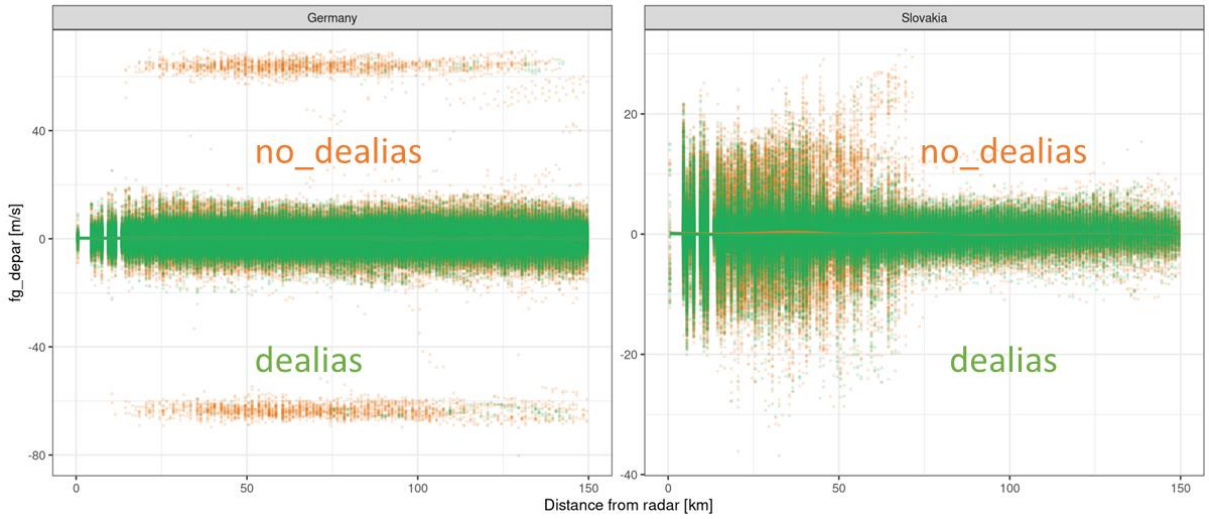


Figure 11 *Visualization of scatter plot of innovations across distance for German radars (left) and Slovak radars (right) for experiment r02 (orange color – without dealiasing) and experiment r04 (green – with dealiasing algorithm).*

As shown in the relative change charts (Figure 12), the introduction of dealiasing yields a profound reduction in the root-mean-square (RMS) impacts by observation type. The relative RMS departures for RADAR DOW drop by approximately -7.5 % for the OMA and -8.5 % for the OMG. This indicates a significantly cleaner fit of both the first guess and final analysis to the active radar observations. Other wind-sensitive observation types, such as aircraft data (AMDAR), show minor adjustments.

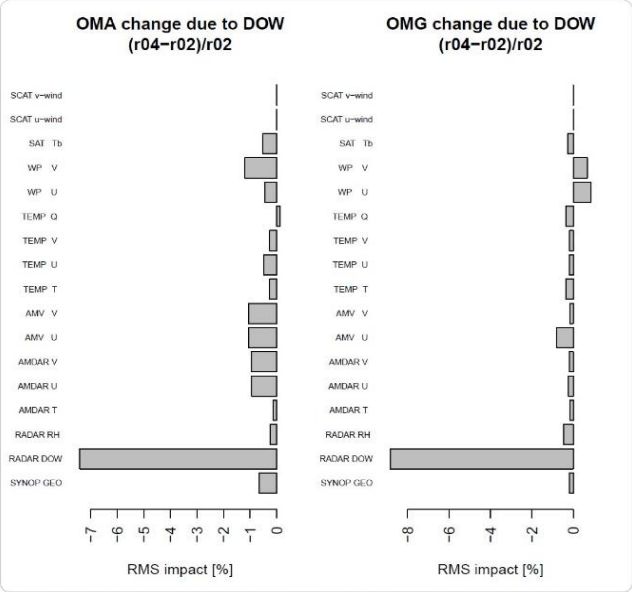


Figure 12 *Relative change in the RMS of OMA (left) and OMG (right) departures due to the application of dealiasing algorithm, across various observation types between experiments r04 and r02.*

4.7.2. Network Density Effect and Dealiased Radars with NI < 30 m/s (r03 vs r04)

Both r03 and r04 have dealiasing enabled, so their comparison isolates the effect of expanding from the three-country network (43 stations, FR+DE+SK) to the full network (77 stations, 11 countries). The additional radars in r03 include several with Nyquist velocities below 30 m/s from Belgium, Czech Republic, Denmark, Poland, Romania, Slovenia, Spain and the United Kingdom. These are radars that would contribute mostly aliased observations without dealiasing; with dealiasing enabled in both experiments, the comparison specifically quantifies their value once properly corrected.

The relative change charts (Figure 13) plot the impact of this network across distinct observation types. Dropping from the full network to three-country configuration results in a 3.2 % increase in the OMA RMS error for RADAR DOW, while the OMG RMS for RADAR DOW is 2.7 % lower. Because the two experiments assimilate different sets of radars, these two bars partly reflect the change of the radar sample itself and cannot be interpreted directly as a better or worse analysis. The observation types that are identical in both experiments change only by a few percent. On the 43 radars shared by both experiments the active-subset BBAD is nearly the same, while the radars added in r03 have a lower median active BBAD (7.8) than the shared radars (11.0).

Long-range verification scores (not shown), tracking the evolution of root-mean-square errors (RMSE) up to 54 hours ahead for wind direction, wind speed, geopotential, and temperature across multiple vertical levels (250, 500, 700, and 850 hPa), were similar for both experiments. The main result of this comparison is therefore that the additional radars, several of them with NI below 30 m/s, can be added once dealiased without degrading the forecast; a forecast benefit of the larger network was not demonstrated over this period.

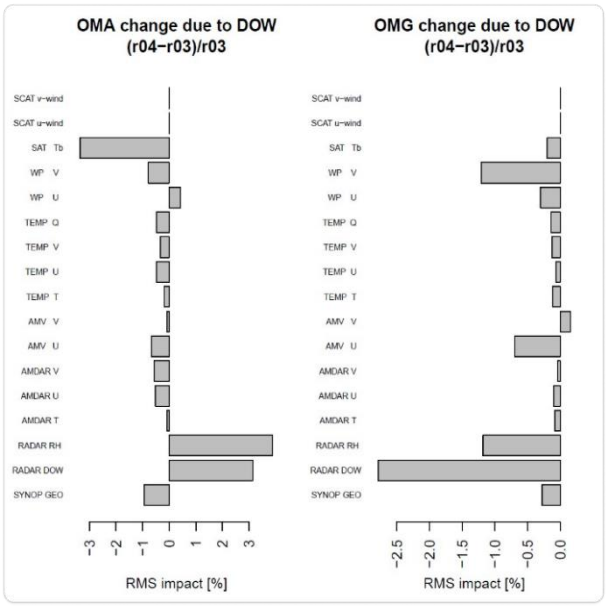


Figure 13 *Relative change in the RMS of OMA (left) and OMG (right) departures due to different network density, across various observation types between experiments r04 and r03.*

4.7.3. Rejection Limit Sensitivity (r02 vs r05)

Experiment r05 evaluates the sensitivity of the data assimilation system to an aggressive observation screening threshold by setting RZFACT = 0.16 instead of the default 1.0 used in the r02 baseline. This drastically narrows the background check acceptance window from approximately ± 20 m/s down to ± 8 m/s, effectively purging all observations with large background departures.

The relative change charts reveal a massive, isolated reduction in RMS departures for RADAR DOW, by roughly -17 % for OMA and -35 % for OMG. This drop does not indicate an improved model state; rather, it is a mechanical byproduct of the truncation, as discarding large departures artificially forces the remaining active radar innovations into a narrower statistical envelope. Crucially, the relative change bars for all other independent, conventional, and satellite observation types remain uniformly near zero.

The 54-hour forecast scores of r05 relative to r02 are shown in Figure 14. The vertical profile verification scores against TEMP observations reveal a positive impact on forecast quality relative to r02, with a measurable reduction in RMSE for geopotential and near overlapping across variables (wind speed, wind direction), particularly in the lower-to-mid troposphere (1000–500 hPa). Removing the large departures thus mitigated a significant portion of the degradation seen in r02, although the impact against the operational suite remained approximately neutral (Figure 9).

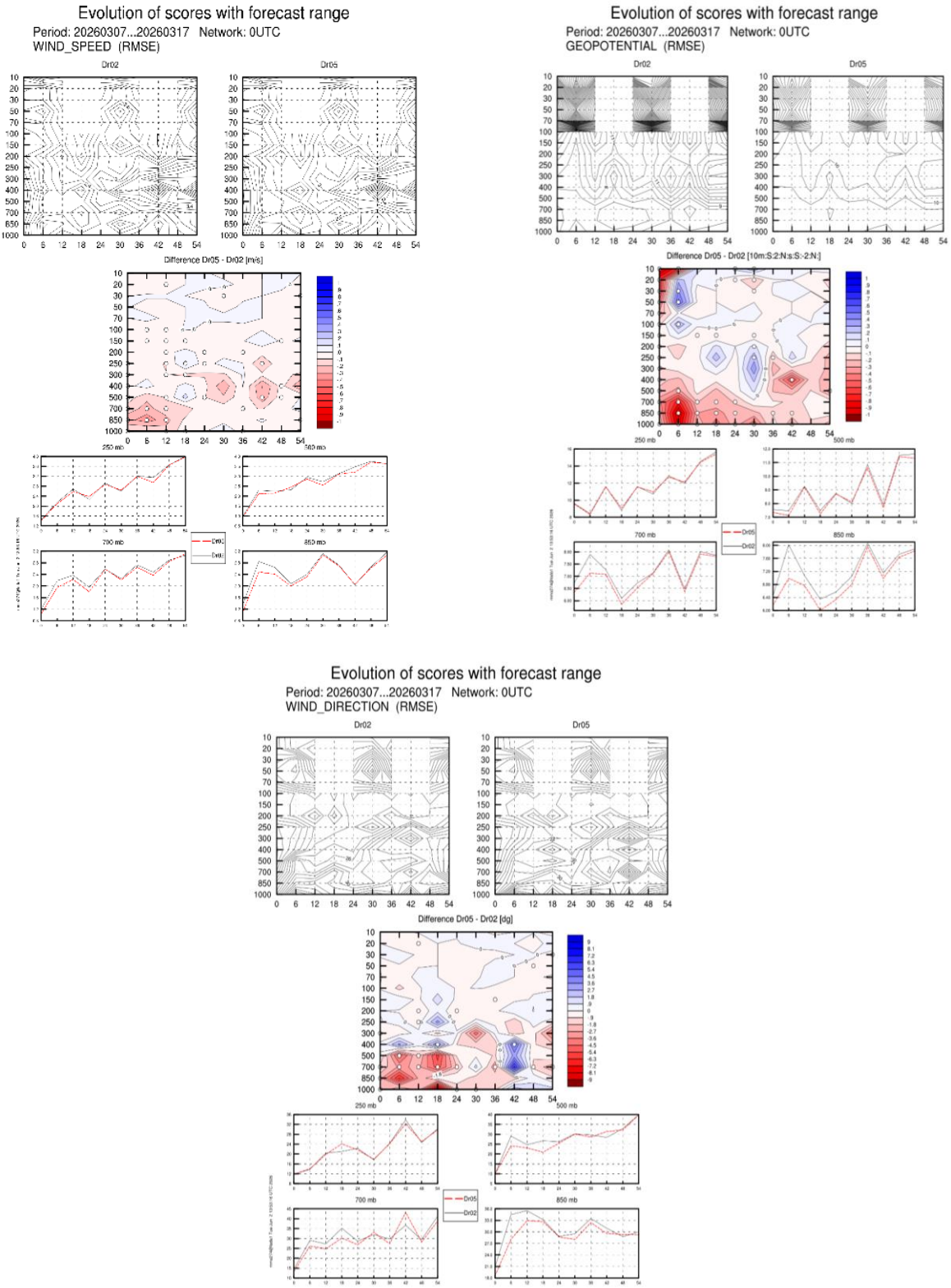


Figure 14 *Shows the verification of forecast scores between two experiments exp r05 (Dr05, red line) and r02 (Dr02, black line) for the period 20260307–17 (12 UTC). Three variables are presented: wind speed RMSE (top left), wind direction RMSE (bottom) and geopotential RMSE (top right).*

4.7.4. Superobservations vs BATOR Spatial Filtering (r02 vs r06)

The comparison of r02 (with BATOR filtering) and r06 (superobservations with BATOR filtering disabled) isolates the effect of the spatial filtering method. The relative change charts (Figure 15) display the most pronounced degradation signal among all conducted sensitivity tests. A prominent positive bar emerges specifically for surface-based observation categories (SYNOP and RADAR DOW) in both the OMA and OMG metrics. This positive relative change indicates a substantial inflation of RMS errors, demonstrating that the SO configuration significantly degrades the model's fit to low-level observations relative to the native BATOR thinning approach.

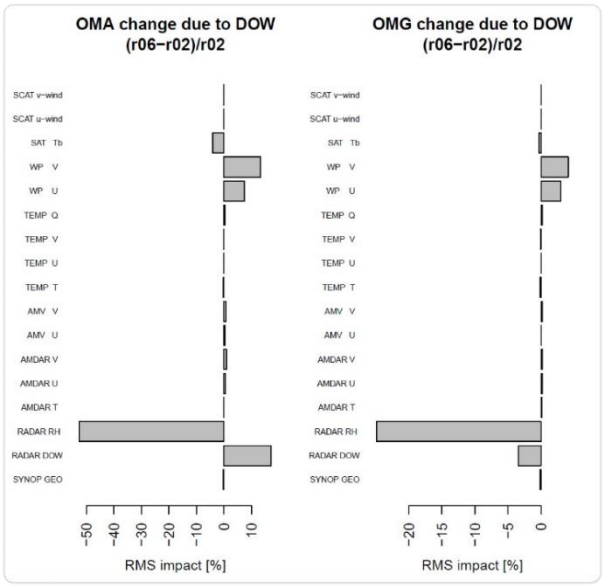


Figure 15 *Relative change in the RMS of OMA (left) and OMG (right) departures due to different spatial filtering methods, across various observation types between experiments r06 and r02.*

4.8. Summary

Taken together, the four comparisons reveal a consistent hierarchy of effects. Dealiasing (r02 vs r04) is the single largest contributor to improved data quality: it reduced the raw innovation spread by 14.8 percent, removed the aliasing bands from the innovation field and lowered both the data loss required for a Gaussian innovation distribution and the width of its core. The tighter rejection limit (r02 vs r05) removed the non-Gaussian tails by truncation but left the core of the distribution as wide as in r02, and the forecast impact against the operational suite was approximately neutral. Superobservation processing (r02 vs r06) improved gaussianity only modestly, produced the widest core of the distribution and degraded the fit to near-surface observations, indicating a representativeness mismatch between the superobservation footprint and the model grid. Finally, extending the network (r03 vs r04) left both the quality of the shared radars and the forecast scores essentially unchanged, which shows that the additional dealiased radars, including those with low NI, can be added without degrading the forecast, while their benefit remains to be demonstrated.

The persistent positive bias, ranging from +0.21 to +0.28 m/s across all experiments, points to a systematic component that is independent of preprocessing choices. This bias is most likely related to the observation error specification or to the background error statistics, which are not adapted for radial wind observations in the current ALADIN configuration.

The most important finding is that none of the tested configurations achieves clearly positive forecast impact over this two-week period. The only configuration close to neutral against the operational suite was r05, and its gain came from discarding the large departures rather than from better data. Among the configurations that keep the default rejection window, the dealiased experiments r04 and r03 were the least negative, both with improved innovation statistics compared to r02. This suggests that dealiasing is a necessary but not sufficient condition for beneficial radar wind assimilation — a conclusion consistent with the published literature, where positive impact required not only good preprocessing but also careful observation error tuning and B matrix adaptation. The r01 passive dataset now provides the foundation for Hollingsworth-Lönnberg observation error estimation (Hollingsworth and Lönnberg, 1986), which should inform the next round of active experiments.

5. Benefit for E-WFC / experiences made during STSM

The main benefit of this STSM for E-WFC is a systematic framework for diagnosing how individual preprocessing and configuration choices affect the assimilation of Doppler radar radial wind data in ALADIN. By designing the six experiments as controlled pairs that vary only one degree of freedom at a time - dealiasing, network density, rejection limit, and spatial filtering - it became possible to separate effects that would otherwise be confounded in a single combined configuration. This isolation is reusable beyond this study: any institute extending or modifying a multi-country radar wind dataset can apply the same pairwise design to identify which setting is responsible for a given change in innovation statistics or forecast scores.

During the STSM, I also worked extensively with the Block Bootstrap Anderson-Darling (BBAD) quality dashboard. Goal was applying it across all six experiments and at different levels of granularity (per radar, per radar elevation angle, per radar distance range) and locate the rejection-limit "sweet spot" and identify where aliasing artifacts were passing the background check undetected. This gave practical, hands-on understanding of the tool and showed its value as diagnostic for radar networks beyond the specific case studied here.

It should be stated clearly that the results were not uniformly positive against operational model. None of the six configurations produced a clearly positive forecast impact over the two-week evaluation period, and dealiasing, while necessary to correct the data-quality problems found in the baseline experiment, was not sufficient to deliver a beneficial forecast impact. The persistent positive bias seen across all experiments can point to a problem in observation error specification rather than in the preprocessing chain itself. I see this as a useful, if modest, outcome for E-WFC: it narrows down where the remaining problem most likely lies rather than leaving it open, and it leaves the passive r01 dataset as a ready starting point for that follow-up work.

This STSM was supported by EUMETNET and co-financed by RC LACE. I am grateful to my host, Mgr. Alena Trojáková, and the colleagues at CHMI for their support during the mission and for giving me access to the supercomputer.

6. References

ANDERSON, E., and JÄRVINEN, H., 1999. Variational quality control. Quart. J. Roy. Meteor. Soc., 125, 697-722. https://doi.org/10.1002/qj.49712555416

BROWN, R. A., and WOOD, V. T. 2007. A guide for interpreting doppler velocity patterns: Northern hemisphere edition. *NOAA/National Severe Storms Laboratory*, 2007. Available: https://www.nssl.noaa.gov/publications/dopplerguide/Doppler%20Guide%202nd%20Ed.pdf.

BROŽKOVÁ, R., BUČÁNEK, A., MAŠEK, J., SMOLÍKOVÁ, P., and TROJÁKOVÁ, A., 2019. New high-resolution operational configuration of the ALADIN model. Meteorological Bulletin, 72(5), 129–139.

BROŽKOVÁ, R., KLARIĆ, D., IVATEK-ŠAHDAN, S., GELEYN, J.-F., CASSÉ, V., ŠIROKÁ, M., RADNÓTI, G., JANOUŠEK, M., STADLBACHER, K., and SEIDL, H., 2001. DFI blending: an alternative tool for preparation of the initial conditions for LAM. PWRP Report Series No. 31 (CAS-JSC WGNE Report), WMO-TD No. 1064, p. 1.7.

EUMETNET, 2024. OPERA – weather radar network. Available: https://www.eumetnet.eu/observations/weather-radar-network/

FISCHER, C., MONTMERLE, T., BERRE, L., AUGER, L., and ȘTEFĂNESCU, S. E., 2005. An overview of the variational assimilation in the ALADIN/France numerical weather-prediction system. Quart. J. Roy. Meteor. Soc., 131(613), 3477–3492. https://doi.org/10.1256/qj.05.115

HOLLINGSWORTH, A., and LÖNNBERG, P., 1986. The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. Tellus A, 38(2), 111–136.

MONTMERLE, T., and FACCANI, C., 2009. Mesoscale Assimilation of Radial Velocities from Doppler Radars in a Preoperational Framework. Monthly Weather Review, 137(6), 1939–1953. https://doi.org/10.1175/2008MWR2725.1

RIDAL, M., et al., 2023. Optimal use of radar radial winds in the HARMONIE numerical weather prediction system. Journal of Applied Meteorology and Climatology, 62(12), 1745–1759. https://journals.ametsoc.org/view/journals/apme/62/12/JAMC-D-23-0013.1.xml

SIRECI, O. 2005. Training course on weather radar systems. Module C : Processing basics in Doppler weather radars. WMO, *Turkish state meteorological service*. 2005. Available: https://www.yumpu.com/en/document/view/5300881/module-c-processing-basics-in-doppler-weather-radars-wmo

SMERKOL, P., STRAJNAR, B., SVAGELJ, V., and ZGONC, A., 2023. HOOF – tool for homogenizing, dealiasing and superobing OPERA radar files in HDF5 format. https://doi.org/10.5281/zenodo.7816818

SMERKOL, P., et al., 2025. Validation of torus mapping dealiasing method on Doppler velocities for use in numerical weather prediction. Meteorologische Zeitschrift, 34(2), 109–123. https://doi.org/10.1127/metz/1256

TERMONIA, P., et al., 2018. The ALADIN System and its canonical model configurations AROME CY41T1 and ALARO CY40T1, Geosci. Model Dev., 11, 257–281, https://doi.org/10.5194/gmd-11-257-2018, 2018.

VÁVRA, M., 2015. Testing for normality with applications. National Bank of Slovakia. ISSN 1337-5830.